Durham E-Theses

Evaluation of Metacognitive and Self-Regulatory Programmes for Learning, Pedagogy and Policy in Tertiary EFL Contexts

WONGDAENG, MEECHAI

Evaluation of Metacognitive and Self-Regulatory Programmes

for Learning, Pedagogy and Policy in Tertiary EFL Contexts

Meechai Wongdaeng

A thesis submitted in partial fulfilment of the qualification of
Doctor of Education

School of Education
Durham University
England
2022

ABSTRACT

In the past few decades, numerous attempts have been made to promote thinking skills and improve learning standards. Among these efforts, thinking skills have been widely advocated in higher education policy and practice. For this reason, the notion of metacognitive and self-regulatory development has been extensively applied in various educational contexts, including English language education. Metacognitive and self-regulatory development is perceived to play an essential role in regulating reflective thinking, learning and agency. Emerging evidence also suggests that metacognitive and self-regulatory development is associated with improved learning outcomes.

To examine the existing evidence on the impact of metacognitive interventions in tertiary English as a Foreign Language (EFL) contexts, a systematic review was conducted. For this purpose, a priori review protocol was developed to minimise bias in identifying, screening and including studies for quality assessment and synthesis. The results from a meta-analysis indicate the potential of the metacognitive approach for EFL learning. However, the evidence remains inconclusive due to the limitations in the included studies. The findings from a thematic synthesis highlight the vital role of explicit instruction, clear pedagogical sequencing and the regulation of skills, rather than mere teaching about metacognition, for successful metacognitive development programmes.

Informed by the findings of the systematic review and a pilot study, the main study in the thesis was designed to assess the effectiveness of a metacognitive intervention on students' learning and metacognitive awareness. The study design is a cluster randomised controlled trial conducted with a target group of approximately 800 tertiary Thai EFL learners in the southernmost areas of Thailand. They were faced with a new education policy challenge requiring students to pass a standardised English test to satisfy the additional graduation criteria. The study included fourteen classes, twelve of which were randomly allocated as either an intervention or a control group. The other two non-randomised clusters were included for comparison. Standardised English tests and student questionnaires were used to collect data for the main outcomes. Classroom observations and interviews were conducted for process evaluation. Secondary data analysis and regression analysis were also included to gain more insights into the role of metacognition and self-regulation in EFL learning.

The results suggest that the intervention group made more progress in the English language than their non-intervention peers. The intervention seems to offer more benefits to lower proficiency students than the more advanced ones. However, the impact of the

intervention on metacognitive awareness had ambiguous results. The process evaluation describes both the positive features and the drawbacks of the intervention. The findings from multimethod analyses provide some valuable implications for learning, pedagogy and policy development.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

ASEAN	The Association of Southeast Asian Nations
CALLA	Cognitive Academic Language Learning Approach
EF EPI	EF English Proficiency Index
EFL	English as a Foreign Language
ELT	English language teaching
ES	Effect size
ESL	English as a Second Language
L1	First language, mother tongue
L2	Second language, a language learnt after the first language
LP	Less proficient
MALQ	Metacognitive awareness listening questionnaire
MP	More proficient
MPS	Metacognitive pedagogical sequence
NNTD	Number of counterfactual cases needed to disturb the finding
PMER	Plan, Monitor, Evaluate, Retrieve
RQ	Research question
SCT	Social Cognitive Theory

STATEMENT OF COPYRIGHT

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

ACKNOWLEDGEMENT

"Are those who know equal to those who do not know?
None will be mindful ˈof thisˋ except people of reason".

Al-Quran, 39:9

CHAPTER ONE

INTRODUCTION

This chapter begins by describing the background of the study. It discusses the status of English education today and the role of metacognition and self-regulation in language learning. Section 1.2 describes the research context of the study. Then, research objectives and research questions are proposed. Finally, the rationales, significance and academic contributions of the study are discussed, followed by a summary of the chapter and the outline of the thesis.

1.1 Background to the study

The advancement in information and communication technology and the trade and business expansion have been among the essential factors expediting global interconnections and the flow of information between countries. These enable the nations with economic, technological and cultural power to become more influential on a global scale, making the English language assume a more important role and become recognised as a global language (Crystal, 2003). This has led to the prominent role of English language education worldwide as the language has usually been associated with improving employability and increasing a country's competitiveness in the global market (Kaur et al., 2016). The English language has been part of the educational curriculum in most countries and has been linked with several indicators showing a country's promise in terms of development (Hayes, 2016). In Thailand, since the National Education Act in 1999, English has continually received high recognition in education policy and curriculum development (Darasawang & Todd, 2012).

Despite the emphasised status and support at the policy level, the results regarding English language proficiency in Thailand have remained low (Baker & Jarunthawatchai, 2017). The high recognition of English in Thai contexts means proficiency in the language creates more opportunities for education and employment (Keyuravong, 2010). Therefore, practical and effective ways to develop English competence among Thai learners are still necessary. One approach which has substantial body of evidence indicating its benefits for improving English learning is the promotion of metacognition and self-regulation (e.g., Chamot, 2008; Goh, 2008; Goh & Taib, 2006; Graham & Macaro, 2008; O'Malley & Chamot,1990; Oxford, 1990, 2011; Vandergrift & Goh, 2012). The approach has been widely applied in EFL contexts. However, robust evaluations of its effectiveness in EFL contexts are rather limited. Therefore, this study is interested in examining the role of metacognition and self-regulation in the learning of

English as a foreign language and its effectiveness for policy and practice in EFL contexts and particularly in a Thai EFL context.

1.1.1 English education in Thailand

English language education has been included in educational curriculums in Thailand at both basic education and higher education levels for a long time. English has been an important foreign language endorsed as a compulsory subject in the basic education core curriculum and available to early-year primary students since the curriculum reform in 1996 (Sukamolson, 1998). In higher education, English is a compulsory subject in most programmes across the country, not to mention the vast availability of English programmes where English is a medium of instruction.

At the policy level, English has long played an influential role in the education curriculum and the country's development agenda. There has been a consistent increase in the emphasis on English language education due to the recognition of English as an essential language for development in the globalised age (Baker & Jarunthawatchai, 2017). In light of the ASEAN (the Association of Southeast Asian Nations) community alliance which declared English as the working language of the region, English language policies in Thailand have been emphasised by the policy makers, aiming that it would facilitate international communication and increase social and economic opportunities (Kaur et al., 2016). The ability to use English for communication has been stressed for Thai graduates and the working-age population (Hayes, 2016).

Following the emphasised policies, numerous studies have been carried out to improve the effectiveness of English language education in the country. Nonetheless, such efforts do not seem to improve the achievement and proficiency of Thai learners of English in general (Baker & Jarunthawatchai, 2017). According to the Education First English Proficiency Index (EF EPI), which annually reports and ranks the English proficiency of countries where English is not the most common mother tongue, Thailand has continually performed unsatisfactorily. The EF EPI is a relevant indicator for the study because it examines the English skills of full-time students in secondary schools and universities in non-English speaking countries. In a recent report in 2019, Thailand was ranked 74th out of hundred and categorised as having 'very low' proficiency falling in the lowest band (EF EPI, 2019).

1.1.2 Problems in English language education in Thailand

Failure to boost the English proficiency of its people despite episodes of measures seems to have been an agonising problem in Thailand's English education for decades. Among the possible explanations, one intriguing factor regarding the poor performance of English is the inequality of opportunities and resources for English learning between students in urban settings and those students in remote areas (Hayes, 2010). While students in the major urban cities such as Bangkok and Phuket generally have access to better educational infrastructures and opportunities for using English in real life, students in upcountry and rural areas have less access to proper resources and opportunities to practically develop their English skills (Fry et al., 2018).

Another reason for the unsatisfactory English outcome is related to the accountability system for education policy (UNESCO Bangkok, 2017) and the quality of education research which lacks real-life impact. As suggested in Wongdaeng & Hajihama (2018) which implemented an intervention in a Thai EFL context, the limitations in the research design restricted a causal inference of the impact of the intervention. Likewise, Wongdaeng (2020) found that most task-based studies in Thai EFL contexts lack design rigour and counterfactual evidence to provide secure findings. Thus, a more rigorous design is required in future research on the effectiveness of an English teaching methodology. The quality of generation, implementation and evaluation of policies and practices and the inadequate attention to the geographical inequality are essential issues to be addressed.

English Proficiency in Thailand is associated with stable and well-paid jobs. People with a good command of English tend to have better opportunities in education and careers (Keyuravong, 2010). Unfortunately, disparities in educational quality and resources widen the gap between major cities such as Bangkok and Phuket and remote areas in the north, northeast and southernmost areas (Fry et al., 2018). As Sahlberg, a Finnish education expert, commented in a Thai newspaper, the inequality in different geographical areas, usually the urban and the rural, is a major obstacle in Thailand's education (Yokakul, 2017). For students in those areas where education quality and resources are less accessible, achieving high competency in English seems even more far-fetched. This means that English competence can widen the rural-urban inequality gap unless requisite support is arranged. The consequences of an education policy on the less privileged groups of students are an important issue to consider. Indeed, similar influences of students' disadvantaged backgrounds on their attainment occur in other countries seen as 'developing' such as Pakistan (e.g., Siddiqui, 2017) or even 'developed' countries such as the UK (e.g., Early et al., 2020; Gorard, 2018).

The educational accountability system and the education research quality are part of the country's slow progress in English language education. The governmental bodies responsible for educational affairs have launched various policies to improve the outcome of English language education, including the English-Speaking Year scheme in 2012 and budget support for English corners in schools (Kaur et al., 2016). However, most of these English improvement policies lack an appropriate evaluative mechanism to account for their effectiveness. If there are attempts of evaluation, most of them rely on asking the practitioners or the recipients of support about their satisfaction. Regarding education research, most empirical English studies are not based on a rigorous design which could warrant evidence of the effectiveness of a proposed solution or intervention. In a review by the researcher (Wongdaeng, 2020) on task-based interventions in EFL contexts during 2004-2018, 51 studies from Thai contexts were identified but only two studies had appropriate designs satisfying the inclusion criteria. Such deficiency of evidence-based effort in policy and research could be one of the factors impeding the country's progress in English language education.

1.1.3 Government actions and policies on English education

With an aim to keep the country competitive in the global economy, English proficiency has been emphasised in the government policies in various forms especially in education. One of these attempts was the curriculum reforms in 2001 and 2008. The reforms of the English curriculum especially in 2008 were to elevate the importance of English subject by increasing teaching hours and allowing more methodological freedom for teachers by minimising the grammar-laden syllabus and promoting a more learner-centred environment (Nonthaisong, 2015). Moreover, promoting thinking skills has been underscored in the secondary students' curriculum (MoE, 2008). In higher education, 'active learning' approaches which permit the learners to play an active role in learning are encouraged nationwide and are part of the teacher evaluation in some universities. Recently, a policy was launched to demand university students pass a form of standardised English test as an additional requirement for graduation (Baker & Jarunthawatchai, 2017).

From the governmental policies discussed above, methodological innovations such as the ones on active learning and thinking skills are considered important. Thinking skills have been recognised as an essential competence for learning and coping with the 21st century challenges and have been stressed in education policy across the world today (Higgins, 2014; OECD, 2005, 2019; Wegerif, Li, & Kaufman, 2015). From several studies and reviews,

frameworks for teaching thinking which are highly associated with improving learning are these pertinent to metacognition and self-reflection (e.g., Dignath & Büttner, 2008; Higgins, 2013; Moseley et al., 2005; Perry, Lundie, & Golder, 2019; Wang, Haertel & Walberg, 1990). Likewise, empirical studies on second language education have found support for the benefits of metacognitive approaches in improving English learning (e.g., Cross, 2011; Rahimirad & Shams, 2014). These demonstrate the promise of the approach for EFL learners. Nevertheless, robust assessment on the topic is still needed to provide more credible evidence on its effectiveness.

1.1.4 The role of metacognition in language learning

To find a solution to improve the learning outcomes of Thai EFL learners, especially those in rural disadvantaged areas, a review of the available evidence is crucial. Through the literature review, a report by the Education Endowment Foundation (EEF) (2018) which has collected syntheses of various kinds of interventions implemented mainly with school-aged students was found to report compelling findings. It identified metacognitive and self-regulatory interventions as highly associated with improving learning, compared to other kinds of interventions such as providing incentives, after-school tuitions or ICT facilities (see Figure 1.1). In another review by de Bruijn-Smolders et al. (2016) which targeted experimental studies in higher education, metacognition was found to positively relate with learning outcomes. This indicates the promises of this approach, instigating an enquiry as to whether the approach can be applicable to English learning in second or foreign language settings.

Figure 1.1: Promises of different interventions for improving learning (based on EEF, 2018)
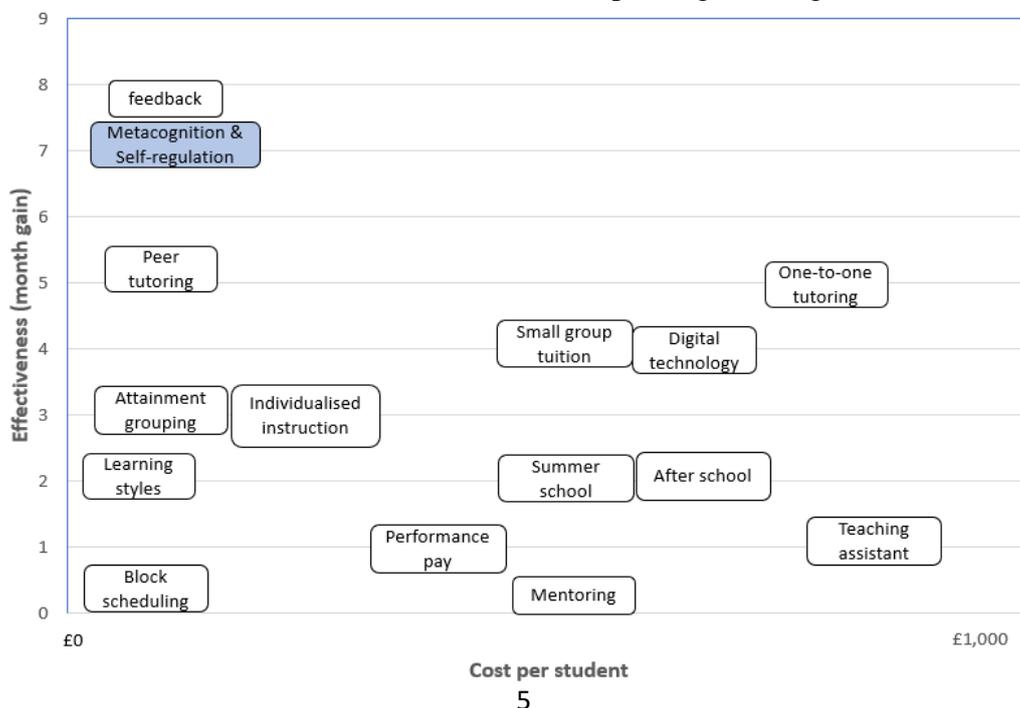
Metacognition, which is simply known as 'thinking about thinking', concerns the awareness of one's own cognitive processes (Flavell, 1976). Strategic thinking and reflective thinking which underpin metacognition are the most salient features associated with meaningful learning (Moseley et al., 2005). The motive behind the metacognitive approach is that developing students to become more aware of their own thinking and more knowledgeable about cognition in general will help them learn better (Pintrich, 2002). The approach promotes students to take active responsibility for various aspects of their learning (Higgins, 2013). With its promising potential, metacognition and self-regulation has become a widely researched area among researchers and educators interested in improving learning and has provided a basis for many educational programmes (Veenman et al., 2006).

In language learning, metacognition and self-regulation is regarded as one of the critical determinants for enhancing thinking, comprehension and second language development (Cohen, 2014; Vandergrift & Goh, 2012; Wenden, 1998). Learning a language requires not only language input and linguistic competence but also cognitive competence to internalise inputs and construct meanings (O'Malley & Chamot, 1990). Language learners hold different beliefs about ways they can use in learning and are aware of the strategies they use (Wenden, 1987). This means that, typically, language learners are able to reflect on how they think and deal with the language input in order to have a particular course of responses or reactions.

In empirical terms, the approach has been applied in a myriad of studies in EFL contexts (e.g., Bozorgian, 2015; Chou, 2017; Cross, 2011; Goh, 1997; Rahimirad & Shams, 2014; Yabukoshi, 2021). Most of these studies applied the concept of metacognition and self-regulation with English learners in higher education. They reported the benefits of the approach for their EFL participants. However, most studies are limited in study design as a comparison is missing or the sample size is very small. Despite the reported benefits, robust studies on the topic to examine its effectiveness are still required. Therefore, this research project attempted to fill this gap by conducting a systematic review of the effectiveness of metacognitive interventions in tertiary EFL contexts to provide a clearer picture of the existing evidence and by conducting a primary study implementing a randomised controlled trial (RCT) of a model of metacognitive instruction with a group of tertiary Thai EFL learners to rigorously examine its impact on the target participants.

Among the different language skills, listening seems to be the least evidentially studied in EFL contexts (Chou, 2017; Field, 2008). In the contexts where English is not used as a mother tongue, reading and writing skills often dominate over aural and oral skills. This is possibly because the aural input is not readily available in such contexts, resulting in research

on listening skills being underdeveloped (White, 2008). While language learners are often taught how to draft a composition, when it comes to listening, learners are often left to deal with listening tasks on their own and rarely taught how to approach listening, limiting the research and insight on teaching listening (Vandergrift & Goh, 2012). In line with Wongdaeng (2020)'s review of task-based interventions in EFL contexts during 2004-2018, comparative studies addressing listening skills are relatively few, compared to reading and writing. This suggests a lack of evidence drawn from empirical studies concerning listening. Therefore, more studies with or without metacognition in focus which address the listening skills of EFL learners are required. For this reason, the metacognitive intervention in the main trial of this study primarily targeted listening skills.

From what has been discussed so far, it could be stated that the policy makers tend to have high recognition of teaching thinking, especially for students at tertiary levels. In addition, the reported benefits of metacognitive and self-regulatory instruction in its own right and as compared to other interventions indicate the high potential of this approach for improving learning. The approach has also been widely applied in tertiary EFL contexts and has seen positive results (e.g., Chou, 2017; Cross, 2011; Rahimirad & Shams, 2014). Therefore, the main trial of this research project worked on a thinking-based intervention by designing and implementing a model of metacognitive instruction with a group of Thai learners of English in a disadvantaged area of Thailand and assessing the effectiveness of the approach for improving the learning of English as a foreign language.

1.2 Context of the main study: English education in a disadvantaged setting in Thailand

As discussed in section 1.1.2, one of the main hindrances of English language education in Thailand is the inequality between English learning quality in different geographical areas. The EF EPI (2019)'s report (Figure 1.2), which provides specific data relevant to Thailand, shows that the English proficiency levels are different across the country. The south and the northeast regions have the lowest English proficiency level compared to other regions. This indicates the differing quality and opportunity of English education in different parts of Thailand.

To help minimise this inequality gap, the main study of the thesis worked with a group of English learners in the southernmost area of Thailand. The area is allegedly reported to be a violence-affected zone under heightened security measures which deter economic and educational development ("Education in Thailand's deep south", 2017). The area has been

affected by social and political instability due to complex conflicts (Seiff, 2016). Such instability has caused sluggishness in educational development in these southern border provinces (Tuntivivat, 2016). The average English score of the national test in the English subject of the students in the southernmost provinces was low for three consecutive years in 2016, 2017 and 2018. Generally, institutions in this area at both secondary and tertiary levels are largely predominated by students with poor English proficiency and are disadvantaged in terms of access to resources and opportunities for quality English learning.

Figure 1.2: English proficiency of Thailand by regions (EF EPI, 2019)

Central: Low (50.10)
North: Low (48.92)
South: Very low (45.02)
Northeast: Very low (43.32)
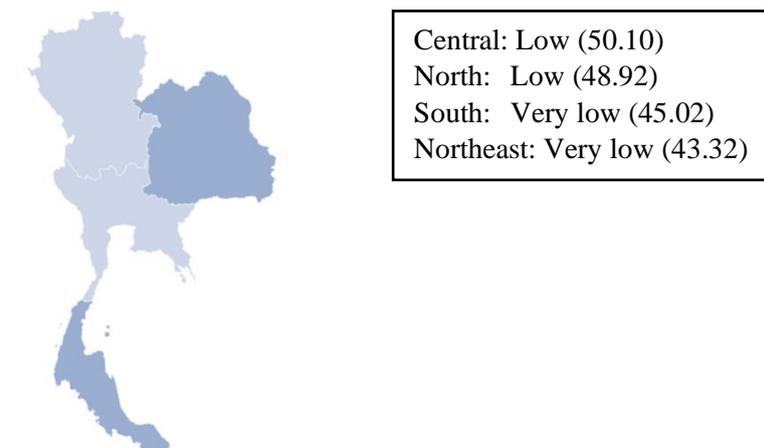
The two universities recruited for the study are located in the southernmost area of the country. The recruited universities used to be among the main universities in the south of Thailand which accommodated students from all areas of the country. However, after the security issue occurred in the area in 2004, they have seen fewer students from other regions. Currently, the university intake comes predominantly from students with similar geographical backgrounds (see further details in section 2.3 in Chapter Two). The results of the cohort 2018 students' English Placement test in one of the universities in the study (Figure 1.3) suggest that the student intake at this university is segregated by students' low English proficiency backgrounds. This clustering of low-achieving students reflects a form of academic segregation which could affect their future opportunities (Gorard & See, 2013; Siddiqui, 2017). From a language learning perspective, interactions in communicative tasks are essential for second language development as the interactions among learners can scaffold language development (Hammond, 2001). Unfortunately, the fact that the students are clustered by low-achieving learners means they have less support from more able peers, which could aid development.

Figure 1.3: Results of English Placement test, cohort 2018
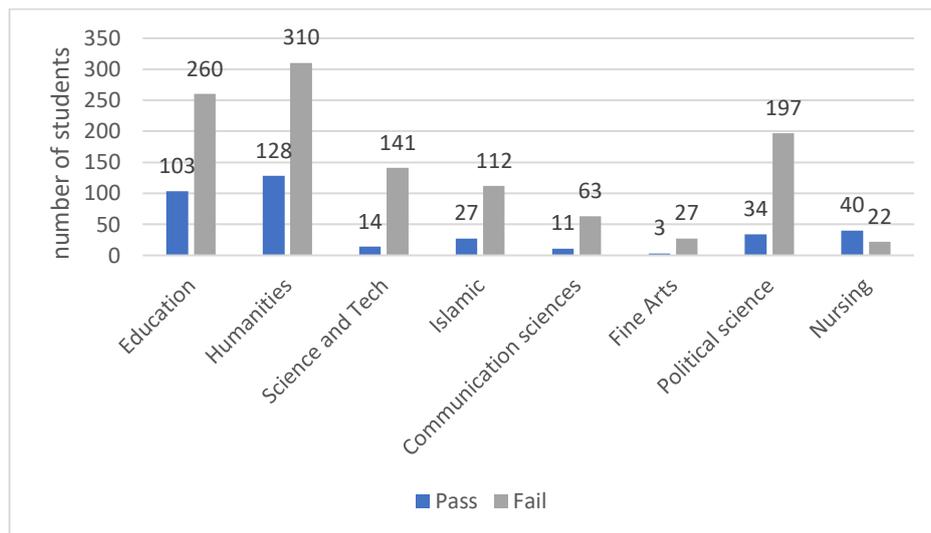
1.2.1 Exit English Exam policy and educational inequality

A recent policy announced in 2016 by the Higher Education Commission regulates that students pass an approved form of standardised English test as an additional requirement for graduation (Baker & Jarunthawatchai, 2017). The policy aims to encourage students to improve their English communicative competence. The introduction of this policy was part of the government's efforts to improve the English language proficiency among university graduates (Wudthayagorn, 2019), which is hoped to enhance the country's competitiveness.

Despite the English competency being a worthwhile quality for the graduates, the policy seems to be deficient in two aspects. Firstly, the policy fails to consider the unintentional consequence that might befall the stakeholders, especially students in the less privileged areas who might have already been disadvantaged such as those in the southernmost provinces. Under such a policy, proficiency in English is another determining factor for the success of their higher education. This means that the poor English proficiency of students in this area may become a more serious challenge or even a barrier for graduation because completing all the courses does not suffice to graduate if they fail to provide proof of accepted English proficiency. Some students may also have to pay for the cost of sitting for the tests if they have used up the free university-supported attempts but remain below the criteria.

Secondly, the policy was not accompanied with a practical strategy or mechanism to support its implementation. An appropriate support system is essential for a policy success (Trowler, 2003). Such support must not be one-size-fits-all but should be equitably distributed based on the actual needs of the stakeholders in different contexts so that such need-based

support can bridge the inequality gap (Ulterhalter, 2009). The issue has been one major threat to educational development in Thailand (Fry et al., 2018).

After the policy enactment in 2016, the universities recruited for this study started implementing the policy with their students. Passing a set requirement in a standardised test is the universities' expectation. The universities also tried to support students by providing English courses and extra English training for them. However, from the data from one of the participating universities, a sizable number of cohort 2016 students who were the first batch obligated by the policy still could not satisfy the criteria at the end of their third year.

Data of progress test results of cohort 2016 (year 3 students) at one of the participating universities indicate that toward the end of academic year 2019 which is the third year of their study, almost a quarter (24.09%) of the cohort still cannot manage to satisfy the English requirement which put them at risk of non-completion. In Table 1.1, which compares the total numbers of students who remain failing in cohort 2016 with the non-completion rates due to expulsion in the previous cohorts, it is noticeable that satisfying the new graduation requirement would be another significant threat to students' completion. The students who still cannot pass the test do need improvement in English in order to meet the qualification criteria.

Table 1.1: Comparison of the cohort 2016's rate of failing English with the expulsion rates of students in cohort 2013-2015

Faculty	Expulsion 2013 (N of student)	Expulsion 2014 (N of student)	Expulsion 2015 (N of student)	Failing English 2016 (N of student)
Education	13	10	9	29
Humanities	62	77	41	79
Science and Tech	36	47	51	81
Islamic	22	17	8	34
Communication sciences	15	14	14	65
Fine Arts	4	5	1	14
Political science	12	13	9	107

From such secondary data, it could be observed that the segregation of those with a poor English background might affect their chances of graduation because the English failing rate is higher than the expulsion rates in the preceding three years. The data from cohort 2016 above can be used to envisage the English passing/failing rates of students in the following years. Based on this data, it can be projected that the students in the following cohorts may also be put under the non-completion risks by the additional language requirement. It seems that the Exit English Exam policy, which claims to encourage students to improve their English for better opportunities upon graduation, is unintentionally putting more obstacles for the disadvantaged students' chances of success if practical support is not provided.

As discussed in section 1.1.4 about the promises of metacognitive intervention on second language learners, this approach can be applied to the participants of this study. Several studies such as Chou (2017), Cross (2011) and Rahimirad & Shams (2014) reported positive effect of their metacognitive instructions for participants in higher education. To secure a more evident estimate of the effectiveness of the approach, a systematic review was conducted, as part of the thesis, to synthesise the existing evidence of its effectiveness in tertiary EFL contexts. With the potential of the approach, the present study also implemented a model of metacognitive instruction with English learners at southernmost Thai universities to support this group of students to cope with the additional challenge. To rigorously evaluate the effectiveness of the intervention, the main study adopted a cluster randomised controlled trial (RCT) design with an adequate sample size for secure evidence.

1.3 Research objectives and research questions

The principal aims of the study are to establish more secure evidence of the impact of metacognitive interventions in tertiary EFL contexts and to investigate the role of a metacognitive intervention in the English learning and metacognitive awareness of a group of learners in Thailand's southern universities. To minimise bias in synthesising the existing evidence, a systematic review approach is an optimum design for identifying, reviewing and synthesising relevant studies (Torgerson, 2003). Among the approaches employed to evaluate the impact of metacognition and self-regulation in language learning settings, the active approach with an intervention and information of strategy use before and after the intervention can provide the most convincing and valid results (Graham, Santos & Vanderplank, 2011). To achieve the ultimate aims of the study, the following objectives were pursued.

- To identify, review and synthesise the existing evidence of metacognitive interventions in tertiary EFL contexts.

- To examine the extent to which metacognitive instruction may have an effect on the listening and overall English achievement of EFL learners in southern Thai universities.

- To investigate the extent to which metacognitive instruction may have an effect on the metacognitive awareness for the listening of English learners in southern Thai universities.

- To examine if there is an association between metacognitive instruction and the outcome differences in biographical variables such as gender, first language background, socio-economic backgrounds and pre-existing proficiency levels.

- To explore the teachers and students' perceptions towards the metacognitive instruction.

With such objectives, the study seeks to estimate the overall effect of the interventions reported in the existing studies. For the main trial, the evidence of the impact of the metacognitive intervention is indicated by the primary outcomes which are English listening and overall English achievement and the secondary outcome which is metacognitive awareness. Additionally, it hopes to provide insights into the relationship between the metacognitive instruction and characteristics of the learners such as their gender, first language, socio-economic backgrounds and pre-existing English proficiency. Moreover, it pays attention to the in-depth data about the participants' perceptions and reactions towards the intervention which can shed light on the process of the implementation.

With such purposes, research questions were formed to direct the research design and methods to obtain relevant evidence. The research questions of this study are as follows.

RQ1. What is the existing evidence of metacognitive interventions in the tertiary EFL contexts?

To answer this question, the following sub-questions are investigated.

1.1. How effective are the metacognitive interventions in the tertiary EFL contexts and what is the quality of the evidence?

1.2. What are the possible factors or characteristics which contribute to the effectiveness of metacognition-based interventions for EFL learners?

RQ2. To what extent does the metacognitive instruction have an impact on the listening and overall English achievement of EFL learners in southern Thai universities?

RQ3. To what extent does the metacognitive instruction have an impact on metacognitive awareness for the listening of English learners in southern Thai universities?

RQ4. In what manner is the impact of metacognitive instruction associated with differences in biographical variables such as gender, first language background, socio-economic backgrounds and pre-existing proficiency levels?

RQ5. What are the teachers and students' perceptions towards the metacognitive instruction?

To answer these questions, multiple methods were included to provide robust evidence. This includes a systematic review of existing studies for evidence of effectiveness and guidance for intervention design, a randomised controlled trial for impact evaluation of the main study, several methods for process evaluation and secondary data analysis for more comprehensive implications for policy and practice.

1.4 Rationale and significance of the study

While there is a large quantity of empirical EFL studies on metacognitive and self-regulatory development, the evidence of its impact on the learning outcomes remains equivocal due to the limitations of the existing studies. In addition, rigorous review studies on the topic in tertiary EFL contexts are scant. Therefore, a review study based on a systematic approach is essential for securing more robust evidence for the topic.

The primary study of the current thesis implemented a model of metacognitive instruction. The trial was conducted to examine the effectiveness of the intervention on improving the learning outcomes and metacognitive awareness of English learners in southern Thai universities. The study sought to provide evidence and insights to complement the theoretical, empirical and contextual needs, as elaborated below.

From a theoretical perspective, there has been a substantial consensus on the beneficial role of metacognition and self-regulation in learning (e.g., Flavell, 1976; Kuhn, 2000; Moseley et al., 2005; Pintrich, 2002, Quigley et al., 2018; Sternberg, 1998; Veenman et al., 2006). However, there has still been a contentious discussion on the optimal conditions for teaching thinking. Some scholars support the general approach to teach thinking separately (e.g., Marin & Halpern, 2011), another group advocates an infusion approach which embeds thinking

instruction in specific subject domains (e.g., Moore, 2011) while others see a mixed approach as the most appropriate (e.g., Abrami et al., 2008). This study which employed a mixed-method approach for delivering metacognitive instruction can offer more information on this argument. Moreover, the literature on the association between metacognition and gender and ethnic difference remains limited and inconclusive (Pintrich & Zusho, 2007; Callan et al., 2016). The present study also sought to explore how the impact of a metacognitive intervention is associated with gender, proficiency levels, first language backgrounds and other socio-economic variables.

In addition, the associative impact of metacognition on particular characteristics or subject areas needs further investigation, apart from the well-established areas such as reading, mathematics and science (Callan et al., 2016; Higgins et al., 2005). One area with expanding investigations is the role of metacognitive instruction in second language development (Goh, 2008; Lui & Li, 2015; Vandergrift & Tafaghodtari, 2010). This study can provide further explanations of the role of metacognition and self-regulation in English learning in a Thai EFL context. Furthermore, as listening comprehension is one of the primary outcomes of the study, the results can fulfil the literature about EFL listening, which is the very core skill for successful understanding and communication (Goh, 2008; Vandergrift, 2007) but is still underresearched (Graham, Santos, & Vanderplank, 2011; Yabukoshi, 2021).

As regards the evidential need, there is a considerable extent of evidence reporting the positive impact of metacognition on learning (e.g., Dignath & Büttner, 2008; EEF, 2018; Higgins et al., 2005). In the EFL contexts, the metacognition-based studies have been substantially expanding. However, most empirical studies on the topic are based on non-robust designs which can undermine the causal claim of the reported evidence, if there is any (e.g., the non-comparator design in An and Shi, 2013; Harputlu & Ceylan, 2014; Zeng & Goh, 2018). In the present study, an RCT, which is one rigorous research design, was used to assess the impact of a metacognitive intervention on the English achievement and the metacognitive awareness of tertiary EFL learners in a Thai context.

Considering the research context, the students at the target southernmost universities seem to have fewer opportunities for developing their English language capacity due to the low-proficiency segregation. They need support for developing their competence in English and thinking capacity to cope with the new challenges, one of which is the Exit English Examination required for them to graduate. To support these disadvantaged students,

metacognitive and self-regulatory instruction might be an appropriate approach because of its potential for improving English learning outcomes (Chou, 2017) and developing metacognitive awareness (Lui & Li, 2015). Furthermore, the approach is reported to be particularly helpful for less-capable learners (Cross, 2011; Goh, 2008; Vandergrift & Tafaghodtari, 2010). This makes the approach suitable for application with the target participants of the study. Moreover, the study can respond to the United Nations' Sustainable Development Goal (SDG) 4 for inclusive and equitable quality education which aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (United Nations, 2015).

1.5 Chapter summary and thesis outline

English is an important foreign language in most of the non-English speaking counties, including Thailand. Thus, policies and practices in English education would inevitably affect large groups of people. In Thailand's education, English is the most important foreign language but also one of the serious challenges for a lot of Thai learners at the same time. Despite numerous policies, improvement has not yet been seen in the learning outcomes. Some of the main problems lie in the education quality divide in major cities and rural areas and the accountability system of education policy and quality of research.

Evidence-based policy and practices are significant for the effectiveness of English education. Some evidence indicates the promises of metacognitive and self-regulatory instruction for EFL learners. For a fundamental understanding of metacognition and self-regulation, Chapter 2 reviews the literature on definition, relevant principles and theories, assessment and some ongoing arguments pertinent to metacognition and self-regulation. It, then, provides explanations on how the metacognitive approach can be applied to second language teaching and learning and specifically to listening comprehension. From such theoretical underpinnings, a specific model of metacognitive intervention for the main trial is discussed. The chapter also provides more information about the research context regarding the background of the research setting and the Exit English Exam policy.

Chapter Three attempts to examine the existing evidence of the effectiveness of the approach for EFL learners by presenting the processes and the findings of a systematic review. The chapter also proposes suggestions learnt from the reviewed studies about the appropriate and unfavourable features of metacognitive interventions. Chapter Four provides background information on essential research elements and describes the implementation and the results of

a preliminary pilot study. In addition, the design and methods of the main RCT are discussed in detail. Then, how the data was analysed is explained.

As the data is quite rich, the results and discussions cannot be condensed into one chapter. Chapter Five presents and discusses the findings of the impact of the intervention on language outcomes while Chapter Six discusses its impact on the secondary outcome which is metacognitive awareness. Chapter Seven examines the association between the primary outcomes and the characteristics of the learners. Chapter Eight describes the process evaluation and discusses its findings. Then, the findings from the secondary data analysis are discussed in Chapter Nine.

After the discussions of various aspects of the findings, Chapter Ten identifies the limitations of the study. It, then, proposes the implications for practices for teachers and learners as well as policy development. Finally, Chapter Eleven is a conclusion which provides a summary, academic contributions of the study, suggestions for future research on the metacognitive and self-regulatory instruction for EFL learners and the key takeaways from the thesis.

With the multiple methods and earnest attempts involved in this thesis, it is hoped that this study will provide beneficial insights into the role and impact of metacognitive interventions for EFL learners. The inclusion of a robust impact evaluation and an in-depth process evaluation is expected to provide credible and helpful implications for policy and practice. Most importantly, the careful attention paid to the theoretical underpinnings and the research elements, namely, design, scale, attrition, outcome measure and validity threats is hoped to provide valuable contributions to the English education literature and research and play a role in enhancing evidence-based policy and practice in the Thai EFL contexts.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

This chapter reviews the literature on fundamental concepts related to metacognition and self-regulated learning and how these concepts are applied in second language learning. The principle and design of the intervention are explained and more detail of the research context is presented.

2.1 Fundamental understandings of metacognition and self-regulation

With the expanding interest in research on teaching thinking, metacognition has become one of the most prevailing approaches applied by many researchers and practitioners today for developing thinking skills and learning achievement (e.g., Dignath & Büttner, 2008; de Bruijn-Smolders et al., 2016; Efklides, 2008; Goh, 2008; Higgins et al., 2005; Lui & Li, 2015; Perry et al., 2019, Wang, Haertel & Walberg, 1990). Metacognition is believed to play a role in enhancing cognitive development and academic learning by raising students' awareness of their own thinking and helping them to actively engage in their own performance (Paris & Winograd, 1990). In the revised version of Bloom's taxonomy of learning by Anderson and Krathwohl (2001), apart from the reordering of cognitive processes of evaluating and creating, metacognition was included as one of the key knowledge dimensions which interacts with the cognitive activities during learning.

With such a theoretical base, a metacognitive approach can be an appropriate approach to be applied in EFL contexts. However, the evidence of how promising the approach is for EFL learners needs further examination. Therefore, a deeper understanding of this concept is explored in this section. Then, its application in EFL settings is elaborated and the evidence of its effectiveness in EFL contexts is examined in section 2.2.

2.1.1 Metacognition: definition and key components

The term *metacognition* is commonly known by its simple definition as 'thinking about thinking'. Indeed, the term and the underlying concept is multifaceted and has been used and interpreted variedly in different disciplines, adding more complexity to define what exactly metacognition is. The coinage of the term was usually ascribed to John Flavell. The term was defined and illustrated in Flavell (1976) as:

> 'Metacognition' refers to one's knowledge concerning one's own cognitive processes and products or anything related to them, e.g. the learning-relevant properties of information or

data. For example, I am engaging in metacognition (metamemory, metalearning, metacognitive attention, metalanguage, or whatever) if I notice that I am having more trouble learning A than B; if it strikes me that I should double-check C before accepting it as a fact; if it occurs to me that I had better scrutinize each and every alternative in any multiple-choice type task situation before deciding which is the best one; if I sense that I had better make a note of D because I may forget it…

(Flavell, 1976, p.232)

In a similar but shorter definition, Flavell (1979) described *metacognition* as 'knowledge and cognition about cognitive phenomena' (p.906). These definitions of the concept highlight the importance of a person's knowledge or awareness of their thinking and learning. Likewise, Brown (1977) identified metacognition as an important aspect of knowledge and asserted that "what is of major interest is knowledge about one's own cognitions rather than the cognitions themselves" (p.4). These explanations clearly emphasise the role of one's own awareness of their cognitive processes. Therefore, the awareness of cognition which can be manifested in many ways such as thinking, knowledge and cognitive skills is fundamental to operationalising metacognition.

To elaborate on the concept, Nelson (1996) describes metacognition as a higher-level or meta-level of thinking which monitors and informs the object-level of thinking where cognitive processes such as decoding occur. This definition also acknowledges the role of cognition in operationalising metacognition and underscores the "meta" level which operates the monitoring and control functions. From the view of expertise development, Sternberg (1998) believed that metacognition is an important human capacity for developing expertise as many aspects of expertise development such as prediction of difficulties, selection of strategies and monitoring involve metacognition.

In a description of how people think and learn based on a review of multiple thinking-based studies, Moseley et al. (2005) made clear the distinctions between cognitive and metacognitive processes. They proposed that the cognitive processes which are represented by information-gathering, building understanding and productive thinking can be exercised in non-strategic and non-reflective ways but the metacognitive processes which are represented by strategic and reflective thinking require information-gathering and other cognitive skills. They argued that strategic and reflective thinking, which are the most salient features of metacognition, are highly associated with meaningful learning.

As a broadening view of Flavell's and Nelson's conceptualisation of metacognition, Efklides (2011) describes metacognition as "a representation of cognition that is built on information coming from the monitoring function and that informs the control function, such as strategy use, when cognition fails for any reason" (p. 6). In other words, metacognition and cognition are interconnected because metacognition is informed by cognition through the monitoring function and informs cognition through the control function. She also emphasised that metacognition is multifaceted with multiple dimensions, namely, metacognitive knowledge, metacognitive experience and metacognitive skills and they interact with cognition alongside emotion in one's attempt to reach a particular goal (Efklides, 2008).

A recent guidance report by a team at the Education Endowment Foundation (EEF) defines cognition as "the mental process involved in knowing, understanding and learning" and cognitive strategies as "fundamental to acquiring knowledge and completing learning tasks" (Quigley et al., 2018 p.9). Meanwhile, metacognition is defined as "the ways learners monitor and purposefully direct their learning" and metacognitive strategies are the ones we use to "monitor and control cognition" (p.9). This definition of metacognition supports the role of monitor and control functions of metacognition in the regulation of cognitive processes to achieve learning goals.

While the benefits of metacognition have been substantially reported, its multiple interpretations and association with various disciplines add complexity to the concept. Many terms are often associated with metacognition such as metacognitive awareness, metacognitive beliefs, metacognitive experiences, metacognitive knowledge, metacognitive skills, judgment of learning, executive skills, higher-order skills, heuristic strategies and self-regulation (Veenman et al., 2006). These reflect different perspectives held by the researchers in their understanding of metacognition. It may not be possible to conclude that one definition is better than another. However, it is important that the researchers provide a clear representation of their understanding of the concept in their study (Haukås et al., 2018).

While there are some discrepancies between definitions, most of them, if not all, emphasise the role of executive processes, i.e., the monitor and control functions in regulating the cognitive processes (Livingston, 2003). Therefore, the operational definition of metacognition or metacognitive awareness in the present study should not only involve the knowledge or cognition of one's own thinking as it is usually known. The executive role of metacognition in monitoring and controlling cognition as well as emotion should also be taken into account. With such realisation, metacognition in this study is operationally defined as *the*

learners' awareness of their own knowledge, experience and emotion and the ability to apply that awareness to regulate their thinking while learning or accomplishing a task. This metacognitive awareness, in turn, interacts with the cognitive and affective domains during the learning processes.

After different definitions of metacognition and the operational definition for this study have been presented and discussed, it is appropriate to explore the constructs of metacognition. Many researchers agree that two main components of metacognition are metacognitive knowledge and metacognitive regulation which is sometimes referred to as metacognitive strategies (Veenman et al., 2006). Metacognitive knowledge includes knowledge of oneself, knowledge of tasks and knowledge of available strategies (Pintrich, 2002). Meanwhile, metacognitive strategies are the processes the learners operate to regulate their learning and manipulate their metacognitive knowledge through reflective planning, monitoring and evaluating one's own learning (CAIE, n.d.; Tobias & Everson, 2000). The process through which learners employ their self-awareness to monitor and direct their strategies or learning behaviours to achieve learning goals are also referred to as self-regulation (Zimmerman, 2002).

Figure 2.1: Metacognition and its components (based on Pintrich, 2002 and CAIE, n.d.)
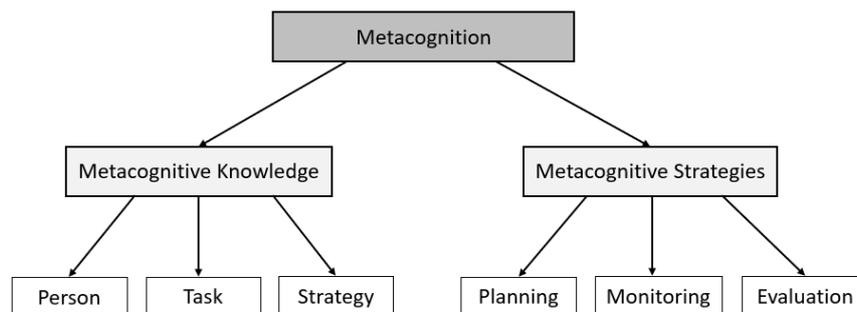
Some authors may include metacognitive experiences as another component of metacognition (e.g., Efklides, 2011). As Flavell (1979) put, metacognitive experiences are 'any conscious cognitive or affective experiences that accompany and pertain to any intellectual enterprise (p.906)'. Some researchers (e.g., Moseley et al., 2005) view thinking as a human activity which involves cognition (knowing), affect (feeling) and conation (wanting and willing). This means when dealing with thinking or cognitive activity, the affect and conation domains cannot be neglected. Therefore, this study focused on the two main components, i.e., metacognitive knowledge and metacognitive skills. Within this remit, the affective and motivational domains were not excluded from consideration.

To pinpoint the distinction between the metacognitive knowledge and metacognitive strategies or regulation, Veenman et al. (2006) pointed out that the former is declarative knowledge (know what) one has about oneself, tasks and strategies while the latter concerns the procedural knowledge (know how) required for the actual use, control and manipulation of such knowledge in a learning effort. In other words, metacognitive knowledge is the awareness of what a person knows about oneself, about the world and about the task which they are dealing and the available strategies which could be applied to accomplish the task. Meanwhile, metacognitive strategies are exercised to regulate when and how to use such knowledge, how to problem-solve in case of lacking some knowledge, how well one performed and how to do better. These occur through the processes of planning, monitoring and evaluating.

Another key term, self-regulation or self-regulated learning (SRL), which is often confused with, used interchangeably with or used together with metacognition should also be discussed here. Alongside metacognition, the notion of self-regulation is also a major construct of this study.

Zimmerman (1989) gives a broad description of self-regulated learners to be "metacognitively, motivationally, and behaviourally active participants in their own learning process" (p.4) and further defines self-regulation as "self-generated thoughts, feelings and actions that are planned and cyclically adapted to the attainment of personal goals" (Zimmerman, 2000 p.14). These definitions emphasise the agentic role of the learner and underscore multiple dimensions of self-regulation. In his view, metacognition plays an important role in self-regulation but self-beliefs and affective reactions are also crucial for attaining self-regulation. From the social cognitive perspective, self-regulation is derived from an interaction of personal, behavioural and environmental processes through the cyclical phases, namely, forethought phase, performance phase and self-reflection phase (Zimmarman, 2002). Through self-regulation, the learners take an active role to monitor and control their learning behaviours, emotions or learning strategies to achieve learning goals (Cleary & Zimmerman, 2004).

In Boekaerts (1996)'s perception, metacognitive and meta-motivational awareness are essential mechanisms for self-regulation. She maintained that self-regulated students are "(meta)cognitively and (meta)motivationally aware of what they are doing and what needs to be done to successfully attain self-defined or set goals" (p.102). These mechanisms form two interrelated regulatory systems in her six-component model of self-regulation. The content

domain, cognitive strategies and cognitive regulatory strategies constitute the system of cognitive information processing. This system interacts with the motivational-emotional system which entails motivational beliefs, motivation strategies and motivational regulatory strategies.

From the previous presentation, it is noticeable that there are various ununified models of self-regulation. Being aware of this non-uniformity, Pintrich (2000) made comprehensive notes of four assumptions which most self-regulation models share, namely, 1) active and agentic role of the learner, 2) potential for control of learning, 3) goal, criterion or standard orientation and 4) mediating role of self-regulatory activities between personal and contextual characteristics and achievement or performance. Given these features, Pintrich defined self-regulation as "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behaviour, guided and constrained by their goals and the contextual features in the environment" (p.453). He agreed that motivational constructs are an essential part of self-regulation processes and proposed a model which entails cognition, motivation and affect, behaviour and context as the areas for regulation.

Seeing metacognition and self-regulated learning as a multifaceted mechanism, Efklides (2011) defines self-regulated learning (SRL) as "the setting of one's goals in relation to learning and ensuring that the goals set are attained" (p.6). She shares the view that metacognition is an integral part of self-regulation alongside cognition, motivation and environment, similar to Pintrich (2000). While both self-regulation and metacognition involve monitoring and control functions, Efklides explained that metacognition is mainly limited to the monitoring and control of cognition while self-regulation is a broader process which monitors and control behaviour, cognition, motivation and the environment. The interactions between metacognition, motivation and affection are the basis of her Metacognitive Affective SRL (MASRL) model of self-regulation.

In the EEF's guidance report on metacognition and self-regulation which was geared towards educational settings, self-regulation was defined as "the extent to which learners are aware of their strengths and weaknesses and the strategies they use to learn" (Quigley et al., 2018 p.8). The report identifies cognition, metacognition, and motivation as the three essential components of self-regulated learning. Metacognition seems to be the main focus of the model as it influences the ways the learners monitor and direct their cognition and learning. However,

cognition and motivation are perceived as essential because it is impossible to be metacognitive without necessary relevant cognitive strategies and adequate motivation to apply those strategies. The three components interact in non-hierarchical ways during the learning processes.

In the definitions and models of self-regulation presented so far, the common emphasis is on the learners' active role in their own learning by exercising the monitor and control functions on their cognitive processes and other constructs such as motivation and environments. In all descriptions, metacognition plays a vital part in self-regulation. Some models perceive metacognition as the main executive function which monitors and controls other components such as cognition and motivation, while some others categorise it as one of the fundamental constructs which mutually interact to execute self-regulation. In both views, metacognition is an essential mechanism for self-regulation.

In fact, there are debates over whether metacognition or self-regulation is a more overarching conception (Moseley et al., 2005; Veenman et al., 2006). Some researchers consider self-regulation to be a subordinate component of metacognition (e.g., Kluwe, 1987), whereas others perceived self-regulation as superordinate to metacognition (e.g., Zimmerman, 2000). With an emphasis on the learners' active and agentic role in their learning following a social cognitive perspective, the present study holds the view that metacognition is one core component of self-regulation which interacts with other components to enable self-regulated learning (Efklides, 2011; Pintrich, 2000; Quigley et al., 2018; Zimmerman, 2000).

2.1.2 Theoretical underpinnings of metacognitive and self-regulatory instruction

The term metacognition is usually credited to John Flavell as the scholar who coined the term. However, contemporary theories of thinking and learning have been largely benefited from the legacies of early theorists of education and psychology and more recent educational theorists such as Vygotsky and Bandura. Moreover, there are other relevant principles for conceptualising metacognition and self-regulation. This section will discuss relevant theories which provide principles underpinning the idea of metacognitive and self-regulatory development.

For any concept including metacognition and self-regulation, there may be several theories or explanations put forth to explain it. Some researchers may try to identify or relate some specific theories to be the main basis of a concept. In this study, the social cognitive theory which views learning as a dynamic and reciprocal interaction of personal, behavioural,

and environmental influences (Bandura, 1986, 1999; Zimmerman 2000) essentially frames the conceptualisation and design of the metacognitive intervention for the main trial. However, in the researcher's belief, the relevant theories have different degrees of relevance and limitations and can explain or complement a concept from different perspectives. Thus, instead of attempting to single out a particular theory to explain the underpinning principle behind metacognition and self-regulation, the relevant theories to be discussed below provide helpful information for understanding the notion of metacognition and self-regulation.

Sociocultural theory

Sociocultural theory was usually accredited to Vygotsky as a founder. According to this theory, cognitive development is largely influenced by social interactions with others (Bråten, 1991). The interactions with people with expert guidance can scaffold or support novice learners in their metacognitive skills such as planning, goal setting and monitoring progress (Moseley et al., 2005). In line with the metacognitive approach to teaching, the modelling and scaffolding role of the teacher is highly beneficial for assisting learners to develop self-regulatory skills (Quigley et al., 2018).

Within the sociocultural theory, the development of language competence is a socially-mediated process which occurs through social interactions and dialogues with more competent members of society (Lantolf et al., 2015). These supportive dialogues with more-able individuals such as peers, teachers and parents *scaffold* or help the learners to build up competence by internalising knowledge co-constructed in a shared activity (Mitchell et al., 2013). Indeed, this is one of the key principles of various self-regulation models. As Zimmerman (2000) suggested, self-regulation can be developed and optimised by placing an emphasis on "the role of socialising agents in the development of self-regulation, such as parents, teachers, coaches, and peers" (p.34) to progress from an observational level to higher levels of self-control and self-regulation. Thus, social interaction can play a pivotal part in making meaning and building self-regulatory skills. When their competence is highly developed, the learners would require less mediation from more capable others (Oxford, 2017).

The role of social interaction emphasised in the social cognitive theory can partially be relevant to developing self-regulatory competence for language learning. Even though the concept of metacognition and self-regulation may superficially seem to mainly be concerned with the awareness and control of one's own mind, the input from social interactions can be one of the fundamental feedback sources for a learner's reflection which can enhance or

accelerate the development of self-regulatory skills (Bandura, 1986). Therefore, the role of interactions with more competent others such as teachers, more able peers or appropriate sources of input should not be neglected in attempts to develop metacognition and self-regulations for students, especially for those with lower proficiency.

However, the sociocultural theory is limited in several ways and cannot adequately explain the principle behind metacognitive and self-regulatory development. Primarily, the sociocultural theory largely emphasises the role of social interactions as a driving force for developing competence while the learner's personal factors such as cognitive, affective and biological variables are not adequately considered (Ameri, 2020). Moreover, the sociocultural theory may be more relevant to language development as interactions can be helpful for developing language competence but may not necessarily be the main influence for the development of other skills such as self-regulation which requires a learner's agentic role to monitor and control one's own thinking. Therefore, it was not the main theory underpinning the metacognitive intervention and principle in this study. Indeed, social cognitive theory which recognises the role of the environmental influence and also emphasises the personal and behavioural factors can provide a more comprehensive explanation for metacognitive and self-regulatory development. This will be further explored in the following section.

Social cognitive theory

Social cognitive theory (SCT), introduced by Albert Bandura, postulates that human learning essentially occurs in a social environment through a process of observation and actual performing, rather than through receiving stimulus, imitation and being rewarded for a behaviour (Schunk, 2012). From the behaviourist perspective, behaviours are substantially influenced by operant conditioning (Aubrey & Riley 2019). That is to say, some behaviours are likely to be repeated if they are accompanied with rewards or positive reinforcement but they may occur less due to punishment or negative reinforcement. In the social cognitive perspective, learning or behavioural development does not occur entirely as a response to external influences or through the stimulus-response-reinforcement process. In response to behaviourist views on human learning behaviours, Bandura (1986) explained:

> People do not behave just to suit the preferences of others. Much of their behavior is motivated and regulated by internal standards and self-evaluative reactions to their own actions. After personal standards have been adopted, discrepancies between a performance and the standard against which it is measured activate evaluative self-reactions, which serve to influence

subsequent behavior. An act, therefore, includes among its determinants self-produced influences.

(Bandura, 1986, p. 20)

According to the SCT, human behaviours are largely developed and shaped by the dynamic and reciprocal interactions among behaviour patterns, environmental variables and personal factors in the forms of cognitive, affective and biological variables. These components are viewed as motivating constructs linked to regulatory mechanisms which activate and direct functions (Bandura, 1991a). The three factors, i.e., person, behaviour and environment interact in dynamic and reciprocal ways. For instance, students who do not understand ask a question (person → behaviour), students underline what they read to help them understand better (behaviour →person/cognition), teachers' clear feedback helps students understand better (environment →person), students' incorrect answers to a question lead a teacher to explain a concept again (behaviour →environment), and students like using wooden blocks to learn about simple addition and perform well, so the teacher decides to use this technique again (person → behaviour →environment).

One distinctive feature of the SCT is self-regulatory functions (Bandura, 1986). According to the SCT, self-regulation fundamentally operates through three processes: self-observation (or self-monitoring), self-judgment, and self-reaction (Bandura, 1991b). Self-observation provides information for realistic goal setting and examining progress. As illustrated in Figure 2.2, there are several factors which affect how self-observation influences reactions. *Temporal proximity* concerns how close and frequent the observation is to the action or behaviour. *Informativeness of performance feedback* is about how perceptible evidence of progress is. This is to say, self-observation can enhance performance when evidence of progress can be perceived. *Motivational level* is an important factor as people with more desire to change tend to react more correspondingly to the observed progress. *Valence of behaviour* concerns the type of behaviours to which more attention is given. A focus on accomplishments or success is often encouraging and can raise self-satisfaction and augment change while heavy attention to failure may be discouraging and lead to no reaction or even lower performance.

Figure 2.2: Self-regulation processes in SCT (based on Bandura, 1991b pp.249-257)

Self-observation	Self-judgment	Self-reaction
Functions: provides information for realistic goal - setting and examining progress	**Function:** provides information for self-reactions	**Function:** regulates courses of action in relation to the standards

Factors
• Temporal proximity
• Informativeness of performance feedback
• Motivational level
• Valence of the behaviour
	Factors
	• Development of personal standards
	• Social referential comparison
	• Valuation of activities
	• Perceived performance determinants
		Forms
		• Evaluative self-reactions
		  -Positive (ex. satisfaction)
		  -Negative (ex. dissatisfaction)
		• Tangible self-reactions
		  -Rewarding
		  -Punishing

Self-judgement provides information for self-reaction. This can take place through several subsidiary processes. *Development of personal standards* provides one with a benchmark for success or failure of their performance or actions. Such standards can be informed by multiple sources of influence such as direct tuition from the surrounding people or observation of reactions from social environment. *Social referential comparison* may be required as some activities have no obvious measure of adequacy, thus, reference to the attainment of others is applicable. For example, the referential comparison is required when one's aim is to be in the upper 10% of a group. *Valuation of activities,* which concerns how significant the activity/task is to the learners can also affect one's reaction. *Perceived performance determinants* are about how success is ascribed. Students are more likely to feel satisfied with or proud of an accomplishment perceived to be from their own efforts, rather than from external aids.

Self-reaction refers to a mechanism by which courses of action are regulated in relation to personal standards. People usually attempt to act or perform in a way that produces positive reactions. Self-reactions can be categorised into two types. *Evaluative self-reactions* are in the form of feeling, impression and realisation towards one's performance or behaviours. This can be positive, such as self-satisfaction, or negative. Positive reactions can lead to more accomplishments. *Tangible self-reactions* are in the form of actions or activities. They can be rewarding such as recreational activities to reward an accomplishment or punishing such as not attending a social gathering due to a recent failure of an activity. Effective use of self-reactions or self-incentives is one of the key factors for success in self-regulation (Bandura, 1991b).

Advocating the social cognitive perspective of self-regulation, Zimmerman (2000) extended the triadic model of self-regulation to the three cyclical phases of self-regulation, namely, forethought phase, performance phase and self-reflection phase (see Figure 2.3). The forethought phase which precedes and sets readiness for the actual performance of a task has two main categories, i.e., *task analysis* which involves goal setting and strategic planning and *self-motivation* which includes self-efficacy, outcome expectation, intrinsic interest and goal orientation. The performance phase which occurs during performing a task has two main categories of *self-control* which includes making image, self-instruction, attention focusing and applying strategies and *self-observation* which consists of self-recording and self-experimentation. The self-reflection phase which involves processes occurring after performance has two categories, i.e., *self-judgement* which involves self-evaluation and causal attribution and *self-reaction* which encompasses self-satisfaction and adaptation and defence. The self-reflection can influence one's reaction to the ongoing or recent experience and the forethought of subsequent attempts or tasks.

Figure 2.3: Cyclical phases of self-regulation (Zimmerman, 2002 p.67)
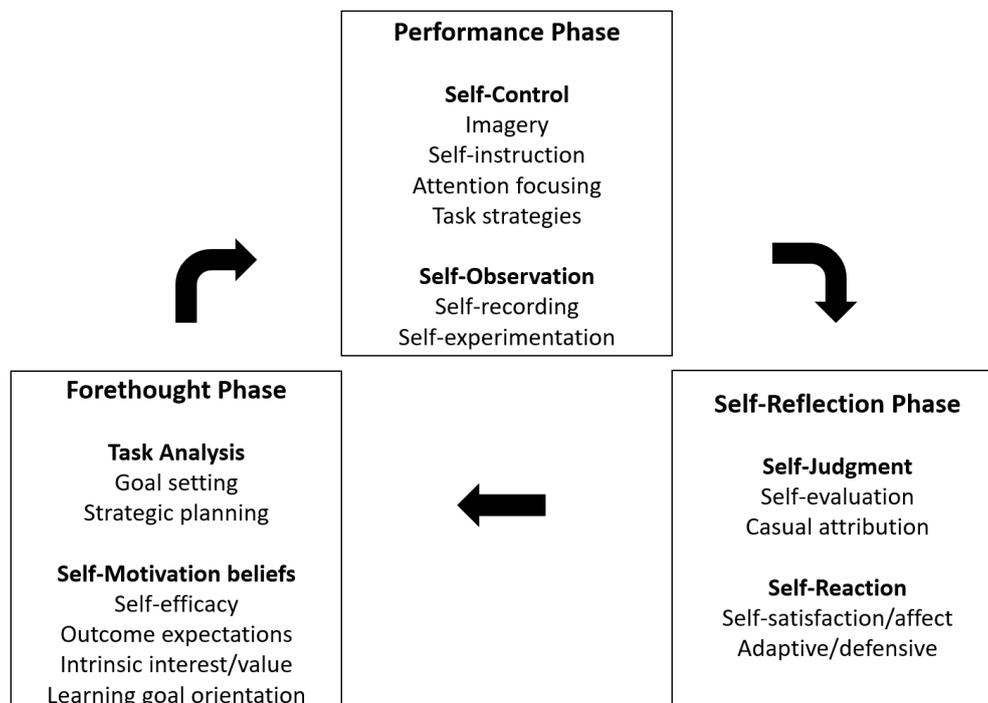
In line with the social cognitive perspective, Pintrich (2000) defined self-regulation as "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behaviour, guided and constrained by their goals and the contextual features in the environment" (p.453). He provided

further explanation of the areas of self-regulation, as illustrated in Table 2.1. The four phases, namely, forethought, monitoring, control and reaction are suggested as a heuristic for self-regulated learning. More details are given for areas of self-regulation in each phase, i.e., cognition, motivation, behaviour and context. The model provides an extended explanation of how the essential variables for learning in the social cognitive theory are manifested in self-regulation.

Table 2.1: Phases and areas for self-regulated learning (Pintrich, 2000 p.454)

Phases of regulation	Areas of regulation			
	Cognition	Motivation	Behaviour	Context
1. Forethought	-Target goal setting -Prior content knowledge activation -Metacognitive knowledge activation	-Goal orientation adoption -Efficacy judgement -Ease of learning judgements; perceptions of task difficulty -Task value activation -Interest activation	-Time and effort planning -Planning for self-observation of behaviour	-Perceptions of task -Perceptions of context
2. Monitoring	-Metacognitive awareness and monitoring of cognition	-Awareness and monitoring of motivation and affect	-Awareness and monitoring of effort, time use, need for help -Self-observation of behaviour	-Monitoring changing task and context conditions
3. Control	-Selection and adaptation of cognitive strategies for learning, thinking	-Selection and adaptation of strategies for managing motivation and affect	-Increase / decrease effort -Persist, give up -Help-seeking	-Change or renegotiate task -Change or leave context
4. Reaction & reflection	-Cognitive judgement -Attributions	-Affective reactions -Attributions	-Choice behaviour	-Evaluation of task -Evaluation of context

In the social cognitive perspective, the development of self-regulatory competence by personal discovery, although possible, is often tedious and limited but can substantially be acquired and sustained by the interaction of social, environmental and personal sources of influence (Bandura, 1986). Within these processes, observations play a significant role for the development of behaviours and skills (Zimmerman, 2000). These involve self-observation and the observations of social models from the environment such as teachers, skilled adults or more-competent peers. In attempting to self-regulate, a feedback loop which is the information used for making subsequent adaptations is utilised. Such feedback can be social such as

guidance or praise from teacher, environmental such as information from the task or context, or personal such as awareness of behavioural outcomes (Zimmerman & Moylan, 2009).

Zimmerman (2000) asserted that through the triadic reciprocal interactions, the development of self-regulatory skills emerges in a series of levels influenced by observations of models from the environment, independent self-observation and adaptive use of feedback from multiple sources in the triadic feedback loop. Differing from the behaviourist view which regards the environment as stimuli to provide conditions for learning via imitation and response, the SCT views the environment as one source of information and motivation in combination with personal and behavioural sources. Meanwhile, the role of environmental factors such as parents, peers and teachers are considered essential for learning or skills development, in line with the sociocultural theory which underlines the important role of social interactions and dialogues with more competent members of society. Not only can the more competent others give support or scaffolding but also provide models for the learners to use and adapt in their personal observation and self-regulation.

Regarding language learning, there is some evidence suggesting that the feedback students receive about their learning has positive effects on language learners' self-efficacy and learning outcomes (Graham, 2007). While the feedback from social interaction has been described as an important process in language learning (Lantolf et al., 2015), self-observation and reflection is also essential for language learners' attempt to regulate their learning (Goh, 1997; Oxford, 2017; Wenden, 1998). Thus, the social cognitive perspective of metacognition and self-regulation can appropriately be applied to the context of this study which targets second language learners. A deeper discussion of how this concept is applied in second language learning will be elaborated in section 2.2.

Self-efficacy, motivation and learner agency

Metacognitive strategies can be of little value if learners cannot motivate themselves to use those strategies. One key motivational factor identified in several self-regulation models such as those presented in the previous section is self-efficacy. It is one of the fundamental constructs of the social cognitive theory (Schunk, 2012). Self-efficacy refers to personal beliefs about one's capabilities to learn or perform effectively at a certain level (Zimmerman, 2000). It is grounded on *attribution theory* which posits that people who attribute their success to personal capabilities and failures to inadequate efforts are more likely to overcome challenges than those who ascribe success to situational factors and failure to deficiency in their ability

(Bandura, 1991a). Self-efficacy is a primary source of motivation which interacts with metacognition in the self-regulatory processes (Pintrich, 2002) and can directly affect one's behaviours such as choice of learning strategies, effort and persistence (Zimmerman & Moylan, 2009). Apparently, people with self-doubts about their capabilities can easily be discouraged by obstacles or failures. With perceived self-efficacy, one will not only be more motivated to learn or perform a task but also have a higher tendency to obtain higher attainment (Zimmaerman & Bandura, 1994). In addition, self-efficacy is an important aspect of good language learners because this self-belief is needed for the learners to make the most of their abilities (Graham, 2007).

From the social cognitive perspective, self-efficacy beliefs can be central to the mechanisms through which human agency is exercised because people have little incentive to act or persevere in overcoming obstacles if they lack beliefs that they can produce desired effects by their action (Bandura, 1999). The sense of *agency*, which is usually viewed as the capacity to act independently and to make one's own choices, is considered central to learning development (Manyukhina & Wyse, 2019). Agentic learners are those who take charge of their own learning by developing awareness of their learning, evaluating their learning needs and applying strategies to address such needs (Hacker et al., 2009). The ultimate source of human agency is self-regulatory competence (Zimmerman, 2002) and metacognitive awareness (Hacker et al., 2009).

A number of studies on language learning have emphasised the realisation and promotion of learner agency and the role of self-regulatory competence in developing agentic learners. Effective learners are aware of their role as agents and persevere in exercising that agency through various strategies to manage learning and regulate emotional responses (Bown, 2009). The sense of agency can be affected by the metacognitive beliefs and the self-efficacy a language learner has (Graham, 2007). From the sociocultural perspective, learner agency can be enhanced by the learners' selection and use of strategies in different contexts in their interactions with their environments and learning contexts (Gao, 2010). The studies on learner agency have been linked to theories about the role of both the individual self and social influence (Mercer, 2011). That is to say, self-efficacy and how the learners respond to the feedback from the environment play an integral part in developing learner agency in language development. Agentic learners tend to have more engagement with their language learning and their agency is not only beneficial for their success as language learners but also for reaching

wider opportunities and overcoming challenges beyond language learning contexts (Larson-Freeman et al., 2021).

Information processing theory

Information processing theory focuses on how information is encoded into human minds, stored in and retrieved from memory (Schunk, 2012). In Mayer (1996 p.154)'s analogy, "humans are processors of information. The mind is an information-processing system. Cognition is a series of mental processes. Learning is the acquisition of mental representations". To understand the information processing, Atkinson and Shiffrin (1968) explained that once the information is passed on to the mind via one or more senses such as seeing and hearing, sensory memory, also called sensory register, receives the input (perception) and holds it briefly before transferring it to short term memory which is a working memory. The working memory, which is limited in capacity and duration, selectively filters the information and processes it by activating related knowledge stored in the long-term memory and integrating it into the new information (see Figure 2.4). The flow of information throughout the processing systems is governed by executive processes such as rehearsal (repeating), coding, monitoring, retrieving as well as metacognitive and self-regulatory activities (Schunk, 2012).

Figure 2.4: Information processing (based on Atkinson & Shiffrin, 1968 p.92-94)
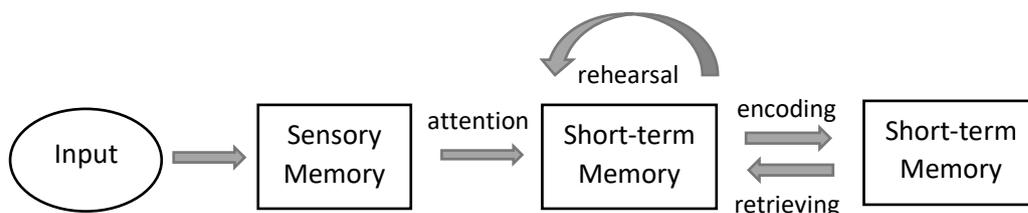
The information processing theory differs from the Piagetian development theory as it describes cognitive development as ongoing and gradual, rather than distinct stages. Based on this theory, the information stored in the long-term memory is advantageous for processing new information.

According to the information processing theory, the schema one has is an important factor for processing and understanding new information. It refers to a structure of thoughts that organizes large amounts of information into a meaningful system (Schunk, 2012). Put simply, a *schema* is a network of knowledge and experiences accumulatively stored in the long-

term memory. Schemas facilitate information processing as they elaborate new input into a meaningful entity (Anderson, 2005). Thus, the available schemas to be activated and elaborated can aid the self-regulatory processes in learning (Zimmerman, 2000), in line with the metacognition and self-regulation principles.

The information processing theory is also used to explain human listening comprehension processes. Anderson (1995, 2005) proposed a framework for listening comprehension through the interrelated three-stage processes, namely, perception, parsing and utilisation. He explained that in the perception stage, attention and recognition are given to the input to translate or decode it into recognised words in the working memory which will parse or segment them into meaningful units and pass them on to the long-term memory for utilisation. The utilisation of parsed units is enabled by relating the meaningful units to the information sources in the long-term memory for making meaning. These components of comprehension are elaborated by many language acquisition researchers such as Anderson (2002), O'Malley et al. (1989) and Vandergrift and Goh (2012). How these reception-parsing-utilisation operate in second language listening is discussed further in section 2.2.2 on metacognition and second language listening.

Cognitive load theory

As seen in the information processing theory, working memory which plays a pivotal role in the learning process has limited capacity. To elaborate on this mechanism, Sweller (1994) postulated cognitive load theory and suggested how the cognitive load can be eased or burdened. The theory posits that working memory can deal with only a limited amount of information at one time. Therefore, unnecessarily complicated information or too great a quantity of information will give extraneous load to the working memory, whereas the ability to make use of schema or the concepts which have been stored in the long-term memory can reduce this cognitive load (Ashman, 2015). In line with the theory, language learners with metacognitive awareness would make use of their prior knowledge and experience to assist learning (Oxford, 2017; Wenden, 1998). As a pedagogical implication, the instructor should realise that the learners' mental capacity to work with information at a time is limited, thus, should not overload their working memory by the unnecessary complexity of tasks or instruction (Shibli & West, 2018). Guiding the learners explicitly to be aware of strategies and know when and how to use them can be advantageous for learning. The clear and explicit

instruction can essentially facilitate students' regulation of metacognitive knowledge and skills which can activate and build up their schemas and eventually support their learning.

Cognitive load theory can be considered very useful for teachers to know (Rosenshine, 2012; Shibli & West, 2018; Wiliam, 2017). Based on evidence from substantial educational trials, the Centre for Education Statistics and Evaluation (CESE) (2017) has proposed some useful applications of the theory as the following.

- Explicit teaching: When teaching new concepts, it is more effective to provide explicit guidance followed by practice and feedback, rather than requiring students to solely discover for themselves.
- Worked examples: Novice learners who are provided with worked examples tend to perform better on a task than students who work on the equivalent task without that support.
- Expertise reversal: As students become more proficient or fluent, they should be allowed more independence in learning or solving a task and the excessive use of worked examples or other assistances for novices are less effective and may even be counter-productive.

CESE (2017)

Despite being useful, the cognitive load theory should be applied with caution. Some of the issues with the theory are, for instance, a limitation to measure cognitive loads which lead to limitations to generate evidence to prove the theory, and superfluous attempts to reduce load which can make learning processes become overly intermittent with too many small steps (Shibli & West 2018). As presented above, reducing cognitive load is not always effective, especially when the learners have become more proficient. Therefore, the awareness and understanding of this theory is apparently helpful but it needs to be optimised to fit the learning contexts and should be accompanied with appropriate teaching strategies.

The presented theories and principles are considered essential for conceptualising metacognition and self-regulation which is a primary focus of this study. The social cognitive theory and its key principles of self-efficacy and learner agency fundamentally provide a theoretical underpinning of the empirical study of this thesis. In addition, the cognitive load theory which is widely applicable in terms of pedagogy and the information processing theory which is particularly relevant to listening comprehension are meaningful for designing the

intervention. Moreover, the sociocultural theory which has largely been linked to studies on language acquisition and literacy development can provide another relevant dimension to be aware of when designing the intervention as well as in understanding learners' behaviours.

2.1.3 Metacognition: how it functions in thinking and learning processes

As seen in section 2.1.1, the definitions and interpretations of metacognition remain disparate, leading to some perplexity about the concept. The first part of this section will be devoted to explaining the mechanism of metacognitive processes in line with the operational definition proposed earlier as the *learners' awareness of their own knowledge, experience and emotion and the ability to apply that awareness to regulate their thinking while learning or accomplishing a task*.

Metacognition does not solely govern the learning processes but interacts with other components in self-regulated learning. To be metacognitive and self-regulatory, the cognitive as well as motivational and affective awareness is required. In exercising metacognitive skills, the metacognitive knowledge in interaction with those capacities catalyse strategic and reflective thinking which are the most salient features of metacognition (Moseley et al., 2005). For instance, learners can use their cognitive skills such as decoding to try to understand or accomplish a learning task. During learners' attempts to control their focus, monitor their progress and make relevant changes, metacognition, which is sometimes described as higher-order thinking or executive function, is exercised (Nelson, 1996). In fact, metacognitive processes can occur from the beginning of a learning task. Before starting a task, learners can think about what knowledge, experiences and strategies they have which might be required or useful for the task. When one has particular knowledge and realises what they know and do not know, this becomes part of their metacognitive knowledge which will be used in the metacognitive processes, i.e., in planning, monitoring and evaluating their own learning (Tobias & Everson, 2009).

For a clearer picture, the integrated model for understanding thinking and learning proposed by Moseley et al. (2005) can explicate the mechanism behind metacognitive processes. Their model is illustrated in Figure 2.5.

Figure 2.5: An integrated model for understanding thinking (Moseley et al., 2005 p.378)

Strategic and Reflective Thinking
Engagement with and management of thinking/learning, supported by value-grounded thinking (including critically reflective thinking)

⇕　　　　　　　⇕　　　　　　　⇕

Cognitive Skills		
Information-gathering	**Building understanding**	**Productive thinking**
Experiencing, recognising and recalling, Comprehending messages and recorded information	Development of meaning (e.g., by elaborating, representing or sharing ideas) Working with patterns and rules, Concept formation, Organizing ideas	Reasoning, Understanding causal relationships, Systematic enquiry, Problem-solving, Creative thinking

In Moseley et al. (2005)'s model, cognitive skills are fundamental components in thinking and learning and are necessary ingredients in strategic and reflective thinking which are the core features of metacognition. The interaction of these cognitive and metacognitive skills in conjunction with motivational and affective factors takes place during the planning, monitoring and evaluating phases of self-regulated learning. In the plan-monitor-evaluate process, reflection is a fundamental part of this process (CAIE, n.d.). This reflective competence can be promoted by encouraging learners in self-questioning throughout a learning task. Self-questioning is one of the effective strategies in improving learning outcomes (Hattie, 2009).

After exploring the concept and the mechanism behind metacognitive regulation, the next question to ask could be 'why is it important?'. First and foremost, metacognitive skills help students to take control of their own learning. In their attempts to plan, monitor and evaluate their learning, learners can become aware of their progress and the remaining difficulties which could guide them for strategic adjustment or improvement in future attempts (Tobias & Everson, 2009). Perkins (1992)'s classification of four levels of metacognitive learners illustrates the ways metacognition is reflected in students' learning behaviours. These are not a rigid pattern of learning behaviours as learners may adopt different strategies and styles to meet their learning aims.

Tacit learners are oblivious of their metacognitive knowledge. They may accept it when they know or do not know something but do not think further about strategies for more effective learning.

Aware learners realise they use certain ways of thinking in particular learning contexts. However, the thinking is not necessarily organised or strategic.

Strategic learners try to organise their thinking by applying relevant strategies they know such as classifying, evidence-seeking and problem-solving to help them learn.

Reflective learners do not only think strategically but also try to reflect on their understanding and strategic application and make an appropriate revision of the way they approach the task for better understanding or achievement.

Similarly, Zimmerman (2000) asserted that the development of self-regulatory skills evolves in sequential stages and can be acquired from and maintained by the internal self and social influence. In light of the social cognitive perspectives, he proposed four developmental levels of regulatory skills presented below.

Table 2.2: Developmental levels of regulatory skill (Zimmerman, 2000 p.29)

Level	Description
1. Observation	Vicarious induction of a skill from a proficient model
2. Emulation	Imitative performance of the general pattern or style of a model's skill with social assistance
3. Self-control	Independent display of the model's skill under structured conditions
4. Self-regulation	Adaptive use of skill across changing personal and environmental conditions

It could be noticed that both classifications of self-regulatory development above mutually hold the view that more effective learners are those who have developed their metacognitive and self-regulatory skills. In Perkins' view, more able learners are strategic and reflective and these two skills are the core characteristics of metacognition (Moseley et al., 2005). For Zimmerman, the abilities to self-control and self-regulate exhibit higher levels of development. Essentially, these two views suggest that the development of metacognition and self-regulation is ultimately important for the learner's cognitive or learning development.

Moreover, certain components of metacognitive and self-regulatory processes, namely, self-efficacy, self-reflection and strategic thinking are positive thinking dispositions which can be transferred across contexts they are taught (Perkins, 2012). These abilities tend to have a strong association with learners' academic and socioeconomic backgrounds because students with less privileged backgrounds are usually found to exercise less of these dispositions (Good et al., 2003). Therefore, research and practices concerning metacognitive and self-regulatory development are important and have been widely implemented.

2.1.4 Metacognition and related theories

- **Retrieval practice**

Retrieval practice is the use of quizzes, exams or other recall activities to help students learn and retain information. The idea is that when students pull out or 'retrieve' knowledge and information in their memory to complete a recall activity, this can exercise their brain and strengthen memory, leading to improved learning and ability to retain information (Agarwal et al., 2012). Moreover, in doing so, students can better examine what they know which leads to a more accurate metacognitive knowledge of their own understanding (Karpicke, 2009). In Dunlosky et al. (2013)'s analysis of ten learning techniques, the retrieval practice is classified as having high utility because it benefits learners of different age and ability groups. Butler et al. (2008) suggested that retrieval practices that provide corrective feedbacks can enhance retention of low-confidence learners. Moreover, from a long-term experiment with middle school students, Agarwal et al. (2017) found that the retrieval practices can be a particularly helpful learning strategy for low ability students.

The newly learned cognitive strategies are often forgotten and are not necessarily transferred to new tasks if the transfer was not done immediately (Oxford, 2017). Therefore, practice tasks are needed to accompany the instruction of cognitive strategies to support the transfer of knowledge. Retrieval practices of the learnt strategies can help learners to realise how much knowledge they can retain in their long-term memory and consolidate retrievable knowledge which will become a schema for future use.

With the presented advantages, retrieval practices were included as another key feature of the intervention in the main trial of this study. As discussed above, retrieval practices can enhance retention of what is learnt (Agarwal et al., 2012), support knowledge transfer, lead to more accurate metacognitive knowledge of their own understanding (Karpicke, 2009) and can be particularly helpful for low ability students (Agarwal et al., 2017). These potentials are

relevant to the study which attempted to introduce novel concepts and strategies to a group of learners, most of whom have low proficiency in the target outcome. The inclusion of retrieval practices is one distinctive feature of the metacognitive intervention in this study and was hoped to support students in transferring and consolidating the newly learnt knowledge and strategies.

- **Self-affirmation**

Some learners especially the underperforming ones have negative esteem or stigma about themselves, known as a stereotype threat. The influence of stereotype threat on learning is explained by self-affirmation theory, which hypothesises that stereotyped learners are aware of the negative stereotype associated with them and such awareness can undermine their learning and motivation (Steele & Aronson, 1995). The self-affirmation strategy would encourage the learners to express positive statements about themselves with an aim to alleviate the effect of stereotype threat on their learning. Good et al. (2003) found that the instruction of thinking skills can reduce the stereotype threat about math learning among female minority low-income students in the US. Similarly, Cohen et al. (2009) posits that a value-affirmation intervention could considerably reduce the achievement gap of the low-achieving African American students (0.49 average GPA rise). A similar value-affirmation intervention in an EEF (2019)'s large-scale 'writing about value' project suggested that the value-affirmation intervention has a small positive effect on improving the performance of the disadvantaged school students in the UK and is a highly cost-worthy intervention. This self-affirmation strategy can enhance self-efficacy which is a crucial component of the self-regulatory processes.

- **Resilience and growth mindset**

Metacognitive awareness helps learners to realise what they do not know or what they are weak at so that they can improve it strategically. Attempting to solve problems by multiple strategies instead of giving up is part of a person's positive dispositions, technically called 'growth mindset'. Dweck (2009) proposed that students with a growth mindset or the belief that intelligence is malleable and can be developed through efforts are likely to become resilient, i.e., being tolerant and can recover or 'bounce back' after some setbacks, disappointment and stress and are more likely to achieve important challenges. Essentially, the notion of growth mindset echoes the attribution theory which underlies one's self-efficacy in Bandura (1991a) which posits that "people who credit their success to personal capabilities and

their failures to insufficient efforts will undertake difficult tasks and persist in the face of failure" (p.71). While self-efficacy is essential for accomplishment, a resilient sense of self-efficacy, i.e., the tendency to maintain self-efficacy is also important as some people become less sure of their efficacy when they fail in a challenging task (Bandura, 1991a). Instead of losing faith in one's self-efficacy, a student with a resilient sense of self-efficacy is more likely to apply metacognition and self-regulation to attribute what one has done well and what needs improvement so as to enhance performance and accomplishment in similar subsequent tasks.

Having a growth mindset, similar to self-belief, is perceived to improve learning. The findings from Blackwell et al. (2007) indicated that growth mindsets in the students can predict an upward trajectory in their grades over the two years of junior high school. Likewise, Rattan et al. (2012) suggested feedback from teachers with a fixed or growth mindsets can affect the students' math performance. They explained that teachers with fixed mindset tend to give comfort-oriented feedback instead of strategy-oriented feedback, leading students to feel less motivated due to the perception that teachers have low expectations of them. With this in view, teaching metacognitive strategies and providing metacognitive feedback can encourage the learners to manage their own learning and cultivate the students' belief that their achievement can be improved through revisions of the way they approach the task.

2.1.5 Assessing metacognitive awareness

As metacognitive awareness concerns the process of thinking, it is unlikely to be tangibly tested. A number of techniques are employed to elicit students' reflection of metacognitive awareness such as interviews, think-aloud protocols, diaries and questionnaires (Haukås et al., 2018). From a systematic review on assessment of metacognition by Gascoine et al. (2017), it was found that self-report measures such as questionnaires and surveys are mostly used in the metacognition-related studies. The responses obtained from these assessment tools can be classified into three types, namely, *prospective judgement* when the responses are elicited before the task performance, *concurrent judgement* when the responses are elicited while performing the task and *retrospective judgement* when the responses are elicited after completing a study phase or test (Schraw, 2009).

One argument on assessment issues which should be considered is whether the assessment of metacognition is done *on-line* or *off-line*. The off-line methods are those administered either before (prospective) or after (retrospective) performance of a learning task for students to reflect on action whereas the on-line ones are assessed during (concurrent) the

task performance for students to reflect in action while performing the task (Veenman et al., 2006). Common off-line measures are self-report questionnaires and interviews, while Think Aloud Protocols (TAPs) are the most common on-line tool (Gascoine et al., 2017). Different types of assessment have different advantages which need to be considered in relation to the focus of studies, nature of tasks and metacognitive skills sought to measure.

Off-line methods such as questionnaires may be easier to administer to large samples but the responses students give on the questionnaires do not always correspond with the actual metacognitive behaviours. Meanwhile, think-aloud protocols may be more revealing about individuals' process of thinking but can be intrusive to some learners. These are the pros and cons of different assessment measures (Veenman et al., 2006). For the present study which focuses on EFL learners' metacognitive awareness in listening, asking students to stop and think while performing a listening task can interfere with the flow of listening. Therefore, an off-line questionnaire was used as the main metacognitive assessment measure in this study.

There is also a critical argument that when students are asked to stop and think in think-aloud protocols, their reflections become off-line (Mateos et al., 2008) as it would occur in retrospective questionnaires. In addition, in Saraç and KaraKelle (2012) who support the multi-methods assessments, significant positive correlations were found between data from two off-line measures, namely, teacher rating scale and a self-report questionnaire. This implies that retrospective questionnaires are, by no means, inferior to on-line assessments.

Numerous questionnaires have been trialled and tested to provide valid and reliable data about the learners' metacognitive awareness. One well-recognised questionnaire for assessing metacognition is the Metacognitive Awareness Inventory (MAI) by Schraw and Dennison (1994). The questionnaire has 52 items which assess metacognitive awareness in a general domain. The questionnaire was trialled in two experiments to examine its validity and reliability. Another widely used off-line tool is the Motivated Strategies for Learning Questionnaire Manual (MSLQ) by Pintrich et al. (1991) which focuses on motivation and learning strategy. Regrading English language learning, Oxford (1990)'s Strategy Inventory of Language Learning (SILL) is a popular questionnaire for eliciting learners' reflection of learning strategies. The questionnaire has 50 items divided into six categories, one of which is metacognitive strategies. The questionnaire items can make the respondents more aware of strategies for language learning. However, the questions seem to be general and the items focusing on metacognition may be too few to be used in this study.

There are other questionnaires which aim at specific language skills such as reading (Mokhtari & Reichard, 2002) and listening (Vandergrift et al., 2006). As discussed in section 1.1.4 in Chapter One and section 2.2.2 of this chapter about the inadequate studies on listening, the significant role of listening for developing language competence and the relationship between listening comprehension and overall language development, the main trial of this study primarily focused on listening skills. Thus, a well-established tool which has been widely applied is the Metacognitive Awareness Listening Questionnaire (MALQ) by Vandergrift et al. (2006). The questionnaire comprises 21 items covering five aspects of planning and evaluation, person knowledge, directed attention, mental translation and problem-solving. The principles underpinning the questionnaire correspond to the metacognitive awareness explored in the present study. In addition, it has been tried and undergone reliability and validity tests. It has been used in several studies such as Chou (2017), Goh & Hu (2014) and Rahimirad & Shams (2014). Thus, this questionnaire was adapted as an instrument to assess the students' metacognitive awareness in listening in this study (see Appendix 1).

As the metacognitive awareness is drawn from self-reported questionnaire responses, measurement of how accurate the students' metacognitive judgement is can provide a more apparent estimate of their awareness which can, in turn, indicate how reliable their responses are. Two main approaches to measuring the accuracy of students' metacognitive judgement are *absolute accuracy measurement* which examines how exact the precision of the match between the metacognitive judgement and the performance is, whereas a *relative accuracy* measures the relationship between a set of judgement and the performance outcomes (Schraw, 2009). As the retrospective off-line measure was used as an assessment tool in this study, a relative accuracy approach was included to measure the accuracy of students' metacognitive judgement by investigating the correlation between the student's metacognitive judgement and the relevant performance outcomes.

2.1.6 Some debates in metacognitive instruction

- **Implicit or explicit**

On whether to teach thinking skills explicitly or implicitly, the advantage of explicit instruction of thinking skills over the implicit method seems to be widely accepted (Wegerif et al., 2015). Evidence from Abrami (2008)'s review indicates that the immersion approach for teaching critical thinking has the lowest effect, providing support for explicit instruction. From EFL researchers, several studies show explicit teaching has more positive impacts on the

learners' achievement (Norris & Ortega, 2001; Taylor et al., 2006). In addition, the support for explicit instruction is in line with the cognitive load theory (Sweller, 1994) which suggests that the working memory can deal with a limited amount of information at one time and unnecessary complexity of instruction can hinder the working memory. Explicit teaching of metacognitive competence would put a lighter cognitive load on the learners' working memory. Thus, the intervention of this study's trial was based on explicit instruction of metacognitive knowledge and skills.

- **General or specific**

As explained earlier in section 2.1, metacognition is one key area in the research on thinking skills. One ongoing argument about the instruction of thinking which can be applicable inclusively to metacognitive instruction is whether thinking skills should be taught separately as a generic domain or should be incorporated in a specific subject. This implies different approaches to teaching and assessing thinking. This discussion on different instructional approaches also arises in the strategy instruction for language learners (O'Malley & Chamot, 1990). Ennis (1989) discussed a typology of four approaches for critical thinking instruction, i.e., general, infusion, immersion, and mixed model. The general approach supports teaching thinking separately from a subject content, while the infusion approach explicitly applies teaching thinking in subject-matter instruction. The immersion is similar to the infusion approach but the instruction is not explicit. The mixed-method approach combines the general approach with either the infusion or the immersion by teaching thinking in the content-specific subject along with general instruction of thinking skills.

On this argument, Abrami (2008) conducted a systematic review of interventions implementing critical thinking instruction and the evidence suggested the infusion of thinking skills in the subject area is more effective than the teaching thinking exclusively but the mixed-method approach is the most effective. Higgins (2015) voiced support for the mixed-method approach, explaining that the infusion of thinking skills in a curriculum guides the learners through how and when to employ these skills while the general instruction allows them to be aware of adapting the skills to different contexts.

In this study, the approach for instruction can be considered a mixed-method approach because explicit instruction of metacognition and self-regulation was introduced to the students first to familiarise students with the general fundamental elements of the concept before linking it to the English listening comprehension which covers most part of the intervention. Even

though the intervention was primarily geared towards English listening, the inclusion of metacognition and self-regulation in general is hoped to help students to link the strategic and reflective thinking developed from English listening tasks to other contexts such as reading, overall English learning, learning in other subjects or other situations in life.

- **Higher order thinking and cultural and religious boundaries**

Another contentious discussion is whether the promotion of thinking skills such as metacognition or critical thinking is plausible outside western cultures, especially where religion or long-standing traditions are highly valued. Some academics believe the teaching of these thinking skills is implausible in the cultures where challenging the traditional or authoritative beliefs is uncommon such as in oriental contexts (e.g., Ramanathan & Kaplan 1996). In fact, cultural and religious values are not necessarily in conflict with critical thinking or reflective thinking (Higgins, 2014). In Islam, for example, being reflective plays a significant part in developing the faith (Kraince, 2007). The abundance of interrogative verses in the Qur'an, the Islamic scripture, deliberately fosters self-inquiry and reflective thought (Norwawi et al., 2019). In Confucian traditions, reflective thinking is consistently encouraged (Li, 2015). Moreover, the results from a systematic review on the impact of critical thinking interventions on ESL learners (El Soufi & See, 2019) suggests that teaching this complex thinking is applicable in most ESL contexts. Thus, religious values or traditional beliefs may not necessarily impede the teaching of thinking skills and should not be viewed as barriers to developing metacognition. However, the practice of teaching and promoting thinking skills can differ in different settings. For instance, it may be more difficult in the contexts where civil rights or human rights are limited.

2.2 Metacognition and second language learning

With a sound theoretical base and extensive evidence of the benefits discussed in section 2.1, the metacognitive approach has been increasingly applied in second language education (e.g., Cross, 2011; Goh, 1997; Macaro & Lynn, 2008). This section will, therefore, further explore the principles and application of metacognitive instruction in second language learning which provide the basis for the intervention of this study.

2.2.1 Theoretical basis of a metacognitive approach to language learning

From the language acquisition perspective, Wenden (1987), a keen pioneer in metacognition and language learning, stated that language learners have different beliefs about ways and strategies to learn a language and are capable of reflecting on their language learning

process. O'Malley and Chamot (1990) elaborated that understanding language acquisition requires reference not only to the language but also cognition insights because cognitive processes are involved in comprehension, production and strategy application which are essential in language development. Similarly, Oxford (2002) suggested that when effective learners make use of learning strategies, they involve in self-directed thinking for developing communicative ability. Thus, learning about learning strategies can be helpful for language development. According to Anderson (2002), metacognitive awareness would help language learners to figure out what to do when facing difficulties. The application of metacognitive strategies could lead to more profound learning and improved language performance.

From the views of language acquisition scholars presented above, the role of metacognition in language learning is clearly upheld. However, from their expressions about the concept (e.g., cognition of strategy application by O'Malley & Chamot, 1990; learning about learning strategies by Oxford, 2002; thinking about strategies by Anderson, 2002), it seems more attention has been given to metacognitive strategies. This could be because the early works concerning metacognition in second learning were studies on learning strategies used by language learners (e.g., Rubin, 1975; O'Malley et al., 1985). Language learning strategies were defined by Rubin (1975), who was one of the pioneers in the field, as "the techniques or devices which a learner may use to acquire knowledge" (p.43). This research theme is aimed to guide learners, especially the less proficient ones, to be aware of and apply various strategies to achieve their learning goals (Griffiths, 2008).

More recent works tend to take a more holistic view of metacognition and self-regulated learning in explaining processes in language learning (e.g., Wenden, 1998; Goh, 2008). This group of scholars similarly views metacognitive strategies as an essential core of metacognition but distinguishes them from metacognitive knowledge which is another key component of metacognition and self-regulation. Learners do not only apply strategies but involve in several processes to regulate their own learning (Dörnyei, 2005). Both metacognitive knowledge and strategies interact with each other in the metacognitive processes when the learners engage in the learning activities.

Apart from the metacognitive strategies such as planning and monitoring via different knowledge sources and strategies, knowledge about how one learns and what to be learned is fundamental to effective decision-making and application of strategies (Flavell, 1979; Vandergrift & Goh, 2012; Veenman et al., 2006). Learners need to be aware of their own knowledge and should be equipped with different types of strategies including linguistic and non-linguistic in order to effectively regulate their learning in the metacognitive processes of

planning, monitoring, problem-solving or evaluating (Vandergrift & Goh, 2012). Therefore, both metacognitive knowledge and metacognitive strategies form the basis of the metacognitive approach in this study. With this interpretation, the term 'metacognitive awareness' would be an umbrella term which encompasses the metacognitive knowledge and metacognitive strategies the learners should be aware of and be able to apply to effectively regulate their own learning, in line with the operational definition presented in section 2.1.1.

Regarding metacognitive knowledge, Wenden (1998) identified *the learner, the learning task and the learning process* as the three categories of metacognitive knowledge in accordance with Flavell (1979). Knowledge about these factors would interact and have an effect on cognitive processes. According to Wenden (1998), person knowledge is the knowledge a learner has about factors which promote or hinder their learning; task knowledge is related to its purpose, its nature and its demands; strategic knowledge is the knowledge of the strategies, why they are useful and how and when to use them. Such metacognitive knowledge is required for the regulation of metacognitive strategies. Metacognitive knowledge about strategies, i.e., knowing when to use and how to coordinate different strategies is essential for developing expertise rather than merely knowing the strategies (Sternberg, 1998).

On metacognitive strategies, Wenden (1998) designates planning, monitoring and evaluation as the three components of metacognitive regulation. These strategies are sometimes referred to as *self-directed learning* in the literature on EFL learners' autonomy. Focusing on learners' strategies, Oxford (1990) identified *centring learning, arranging and planning learning and evaluating learning* as the three components of metacognitive strategies (see Figure 2.6). In Oxford's classification of language learning strategies, metacognitive strategies in combination with affective and social strategies are the indirect strategies for language learning.

Figure 2.6: Language learning strategies (adapted from Oxford, 1990 p. 17)
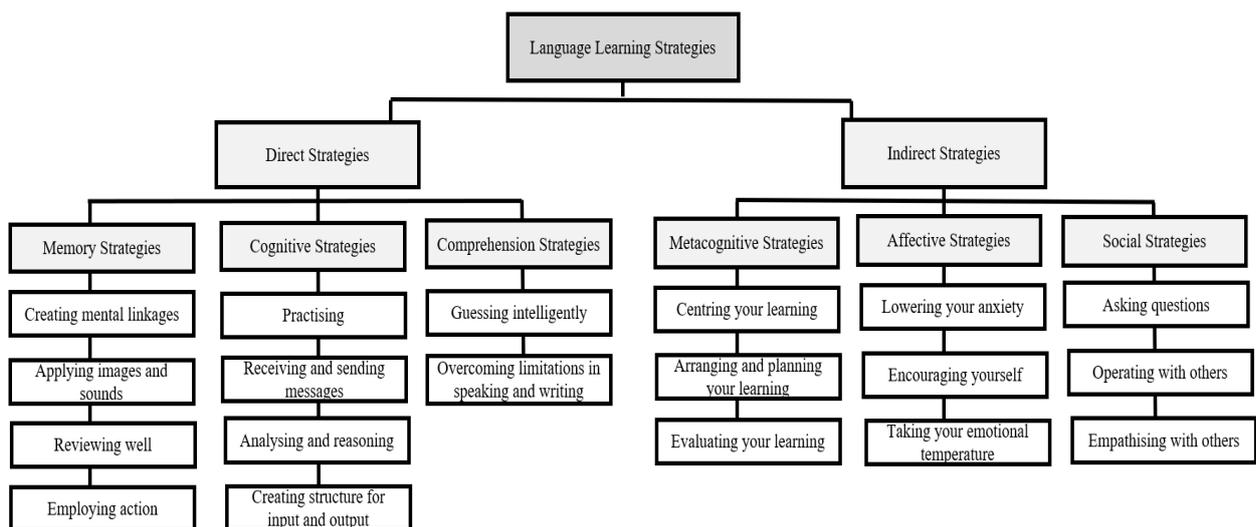
Similarly, O'Malley and Chamot (1990) classified metacognitive strategies as one of three categories of language learning strategies. In their classification, they provided more details of the meta-strategic processes and offered seven components of metacognitive strategies, namely, planning, directed attention, selective attention, self-management, self-monitoring, problem identification and self-evaluation (See Figure 2.7).

Figure 2.7: Language learning strategies (summarised from O'Malley & Chamot, 1990, p.137-139)
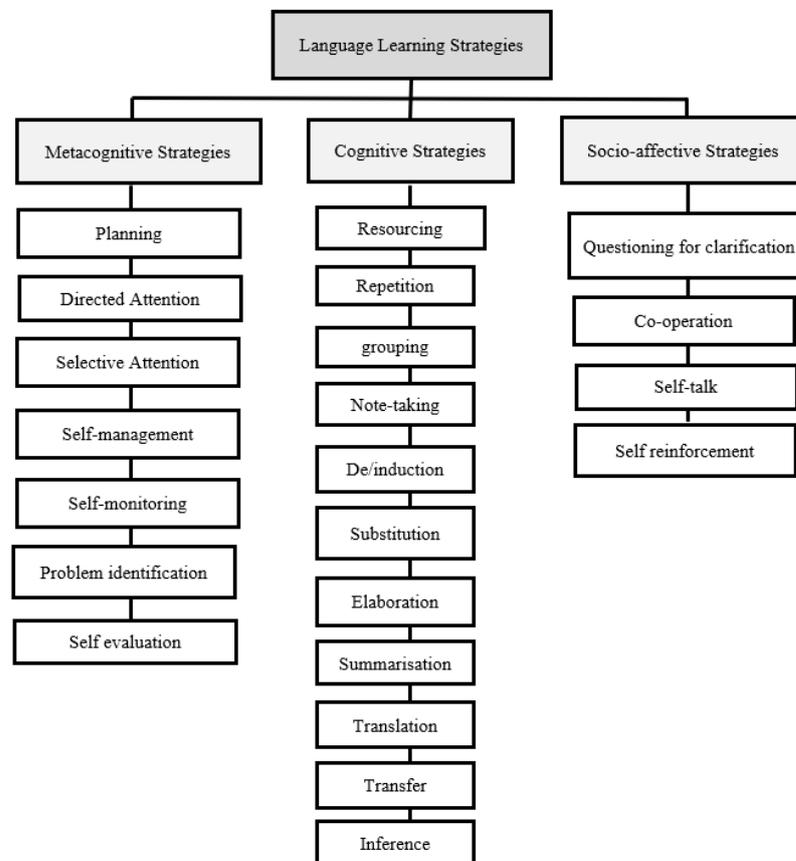
From the classifications of language learning strategies recently presented, it is clear that metacognitive strategies are not the sole component governing language learning but they interact with other variables. For example, the cognitive strategy of *elaboration*, which tries to relate prior knowledge to the knowledge gained from the task, can be used in the metacognitive process of planning. The affective strategy of *reinforcement* can be used when facing difficulties during the task. This emphasises the intertwining role of different components in self-regulated learning.

Less influenced by the learner strategy tradition and more focused on the role of metacognition in the language learning process, Anderson (2002, 2008) divided metacognitive strategies into five components:(1) preparing and planning for learning, (2) selecting and using

learning strategies, (3) monitoring strategy use, (4) orchestrating various strategies, and (5) evaluating strategy use and learning (see Figure 2.8).

Figure 2.8: Metacognitive strategies in language learning (Anderson, 2008 p.100)
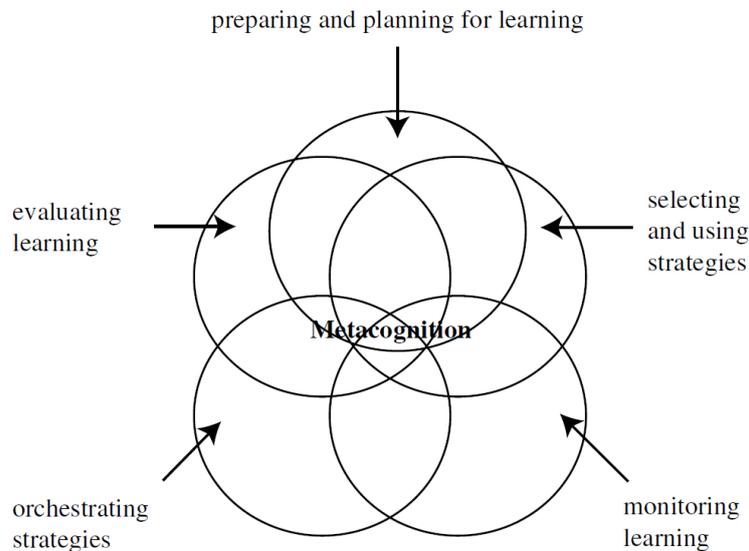
Anderson (2008) emphasised that the metacognitive processes do not occur in a linear fashion, so it is important for teachers to model these strategies for the learners. He added that metacognitive awareness can provide guidance on how language acquisition can be accelerated. The self-awareness of one's own progress and difficulties guide relevant courses of action for improvement.

What has been discussed so far in this section shows how metacognition can play a role in second language learning. Several scholars designate metacognitive strategies as one of the main components of language learning strategies (e.g., O'Malley and Chamot, 1990; Oxford, 1990) and most researchers agree that the metacognitive strategies have a crucial role for effective learning (Anderson, 2002; Haukås et al., 2018). Metacognitive knowledge about person, task, and strategy can be utilised in the learner's attempt to apply metacognitive strategies to achieve their learning goals (Vandergrift & Goh, 2012; Wenden, 1998). Such metacognitive awareness interacts with other factors such as cognition, motivation, and behavioural and environmental variables in the self-regulatory processes to improve learning (Dörnyei, 2005).

From the different classifications of metacognitive strategies presented above, it could be noticed that the roles of monitoring and reflection were included consistently. This is in accordance with the operational definition of metacognition presented in section 2.1.1. Thus, they can be essential components of the metacognitive intervention in this study. The reflective functions can be applied for planning purposes when reflecting on prior knowledge or

experiences. They can also be employed for evaluative purposes when reflecting on the past performance.

2.2.2 Metacognition and second language listening

Language acquisition is an implicit process which is triggered by exposure to language input (Krashen, 1981). Receptive skills such as listening are essential for language learning because they provide input for the learners (Nunan, 2002). Moreover, there is some evidence indicating that the learners' proficiency in listening has a high association with the progress in overall language development (Feyten, 1991). Despite such an integral role, the listening comprehension of second language learners seems to be the least investigated among other language skills (Graham, Santos, & Vanderplank, 2011; Vandergrift, 2007; Yabukoshi, 2021). The results from a systematic review (Wongdaeng, 2020) on task-based teaching in EFL contexts found that robust empirical studies on listening comprehension are minimal. Moreover, the typical approach for comprehension classrooms seems to be some forms of testing comprehension, rather than teaching ways to improve comprehension (Wilson & Bai, 2010) which keeps the listening pedagogy weakly principled (Graham & Santos, 2015). These reasons provide the basis for the main trial of the study to particularly aim at examining the metacognitive approach to EFL listening and its impact on listening outcome while the overall English outcome will also be considered. Therefore, this section can provide an elaboration of the principles of second language listening and principles for applying metacognitive instruction in EFL classrooms.

2.2.2.1 Principles of second language listening

Listening is a complex process with multiple overlapping types of processing and is one of the most crucial components of spoken language processing which is interconnected with numerous areas of development (Rost, 2011). In the first language, listening is the first skill which learners develop. In the learning of languages other than the first, however, the reading and writing skills are usually emphasised before the aural and oral skills, probably because the aural input is not readily available in the L2 environments (White, 2008).

Listening comprehension involves different types of knowledge, both linguistics and non-linguistics. The most fundamental linguistic knowledge are phonology, lexis and syntax as well as semantics and discourse structure while the non-linguistic knowledge involves

knowledge about the topic, context and general knowledge about the world (Buck, 2001). These sources of knowledge are utilised in building comprehension.

Despite being described as a receptive language skill, listening is an active skill which involves multiple processes to operate the received input. At least, two main tasks which the listener need to involve synchronously in recognising the input are identifying words and lexical phrases and activating relevant knowledge associated with those words and phrases (Rost, 2011). In line with the information processing model by J. R. Anderson presented earlier in section 2.1.2, O'Malley, Chamot and Küpper (1989) explained that listening comprehension involves three interrelated and recursive processes of perceptual processing (perception), parsing and utilisation.

Perception is the linking of input to recognisable sounds.

Parsing is establishing or identifying meaningful mental representation of the perceived input by matching them with the linguistic knowledge one has.

Utilisation is relating the parsed unit which is a meaningful mental representation to the existing knowledge or schemata about it.

These processes do not occur in a linear manner but as overlapping processes and require active engagement in listening which is distinct from simply hearing (Graham & Santos, 2015). With such underlying complexity, the teaching of listening should guide learners to various strategies such as predicting, self-monitoring, elaboration and inferencing to help learners cope with challenges in the listening processes (O'Malley et al., 1989).

As regards the listening process, Flowerdew and Miller (2005) discussed three models of processing in listening which reflect learners' strategies in listening as the following.

The bottom-up processing refers to the learners' attempt to work with the smallest unit of the acoustic message, i.e., the phonemes to build understanding.

The top-down processing relates to the use of prior or contextual knowledge, usually referred to as *schema,* to make meaning.

The interactive processing acknowledges the interaction between the bottom-up and the top-down processing which varies by individual learning approaches.

The bottom-up model corresponds to Field (2008)'s *decoding processes* which deal with word-level input such as identifying sounds and syllables which is frequently used by less

experienced learners. Meanwhile, the top-down model could be compatible with Field (2008)'s *meaning-building processes* which draw understanding from contextual information such as using contexts, making inferences, and checking understanding. Indeed, both bottom-up and top-down processes are necessary for building comprehension (Nunan, 2002; Graham & Santos, 2015).

Graham and Macaro (2008) explained that the bottom-up process mainly requires linguistic knowledge such as word and sound recognition, while the top-down process involves non-linguistic sources such as contextual knowledge, background knowledge about the topic and general world knowledge. These two processes interact in building comprehension in a *compensatory* manner when one type of knowledge such as background knowledge complements the lack of the other source of knowledge. They can also occur in a *confirmatory* manner when listening does not fail or when information from one source of knowledge supports what is known from the other source.

Apart from the sources of knowledge and the information processing, another important factor to consider in listening comprehension is the degree of interaction between the listener and speaker (Buck, 2001). In *non-collaborative* listening, the listeners' role may be mainly to listen and interpret the utterances. However, in some other situations such as in a group discussion or conversations, the listeners' role may involve asking for clarifications, turn-taking and making responses. The increased roles the listeners have to take could lead to greater difficulty in the listening tasks.

The above discussions on listening comprehension processes by Field (2008), Flowerdew & Miller (2005), Graham and Macaro (2008) and O'Malley, Chamot and Küpper (1989) have exhibited the complex procedure in which the learners need to engage to achieve comprehension. This led to the interest in exploring learner strategies in listening comprehension. Effective listeners rely on all multiple types of cognitive, metacognitive, and affective strategies (O'Malley, Chamot & Küpper, 1989; White, 2008) such as elaboration, inferencing and self-monitoring. These investigations provide a better understanding of how less and more skilled listeners behave in their learning attempts and insights into the ways in which listeners regulate their listening processes (Vandergrift, 2007). The role of metacognitive and self-regulatory processes in second language listening have increasingly been explored which provide theoretical and empirical support of the approach. The principles behind the

metacognitive and self-regulatory approach to the teaching and learning of second language listening will be explored further in the next section.

2.2.2.2 Metacognitive approaches to second language listening

Early structures of listening lessons which remain in practice by many teachers today are primarily constructed on the three-stage procedure. The stages include a pre-listening stage to present new or critical vocabulary to be heard, a listening stage to let students involved in a listening task accompanied by general and detailed comprehension questions about the listening text, and a post-listening stage to check the answers to the questions and sometimes explore the listening scripts (Field, 2008).

Table 2.3: The traditional three-stage procedure in listening classrooms (Field, 2008)

Pre-listening	Pre-teach vocabulary to facilitate understanding
Listening	-Extensive listening, often followed by general questions -Intensive listening, followed by detailed comprehension questions
Post-listening	-Play and pause for comprehension check. -Teach any new vocabulary -Analyse language. Sometimes go through the listening scripts

The procedure presented in Table 2.3 seems to attend only to the comprehension of a specific teacher-given input without considering the ways the learners process the input. Field (2008) describes such procedure as a comprehension approach in which the teachers evaluate the student's understanding via the correct responses to the comprehension questions. Field viewed this approach as inadequately addressing the needs for listening development as the approach involves more testing comprehension rather than guiding the learners through the process of listening. As listening is not a passive activity, teaching or improving listening needs to take into account the multiple variables involved in the listening processes (Vandergrift, 2004).

When engaging in listening tasks, listeners have to encounter active and complex processes for constructing meaning (Field, 2008; O'Malley, Chamot and Küpper, 1989; Vandergrift, 2007). From a study of the processes the learners apply while listening, O'Malley, Chamot and Küpper (1989) pointed out that the learners reported using different types of

strategies to comprehend the listening text, including self-monitoring which was found to be among the most effective strategies used by effective listeners. Such findings affirm Wenden (1987)'s argument that foreign or second language listeners have different beliefs about ways to learn a language and are able to become aware of their own learning process. Goh (1997), from the listening diaries of 40 EFL listeners, reiterated that the learners can reflect on their own thinking and their involvement in the metacognitive awareness is essential for improving learning. Metacognitive awareness can facilitate the way learners approach a listening task and enable them to effectively plan, monitor and evaluate their listening (Goh, 2008).

As discussed earlier, the traditional three-stage instruction is insufficient to develop a good and independent listener because such a product-based approach seems merely testing or practicing listening instead of teaching listening (Vandergrift & Goh, 2012). In writing or speaking lessons, language learners are often taught to plan or draft a piece of writing or oral presentation. In contrast, students are rarely guided on how to approach listening tasks. Indeed, awareness of listening strategies and realisation of how and when to apply them have been reported to be beneficial for learning, as shown in part 2.2.1. Therefore, teaching listening needs to realise the importance of metacognitive awareness which regulates students' learning and application of strategies if the teaching is to empower the learners in their own listening process (Vandergrift, 2007).

Studying the role of metacognition on reading comprehension which has a larger multitude of studies can provide a useful application for the like of listening comprehension research. Based on the review of more than 40 studies of strategies employment by young readers, Pressley (2002) divided skilled comprehension into two levels, i.e., letter-and-word level and above-word level. At the *letter-and-word level*, Pressley suggests that the less-fluent learners spend much time on word recognition, thus, have less time for working with comprehension because both decoding and comprehension take place in the short-term memory which has limited capacity. This has two important implications; word recognition skills matter in comprehension and focusing attention on every single word can negatively affect comprehension.

For *above-word-level* comprehension, Pressley (2002) asserted that good readers employ various strategies before, during and after reading such as setting clear goals, preliminary skimming, activating prior knowledge, selective attention, predicting, making inferences, monitoring comprehension, selective re-rereading and checking understanding.

With such findings, Pressley emphasised that metacognitive strategies should be taught together with word recognition and vocabulary in order to enhance comprehension.

Focusing on listening, Graham and Macaro (2008) provided a concise explanation of how metacognitive processes play an important role in listening. They explained that prediction, which is one useful pre-listening strategy, stimulates schema and lightens the cognitive load. During listening, the bottom-up processing which draws on linguistic knowledge and the top-down processing which involves non-linguistic knowledge interact with each other in a compensatory or confirmatory manner. The monitoring and evaluation strategies help to ensure the applied strategies are working properly. These metacognitive processes support learners to regulate their listening to plan, maintain or adjust their strategy use to solve problems.

Vandergrift and Goh, other prominent scholars on second language listening (e.g., Goh, 1997; 2008; Vandergrift, 2004; 2007) further elaborated on the metacognitive regulation in second language listening. In their book (Vandergrift and Goh, 2012), they explained that the cognitive processes during second language listening comprehension involve (1) top-down and bottom-up processing; (2) controlled and automatic processing; (3) perception, parsing, and utilisation; and (4) metacognition (p.17). They elaborated that the learners manipulate these processes by drawing on different knowledge sources such as linguistic knowledge, prior knowledge and discourse knowledge (see Figure 2.9). In these interactions, metacognition regulates the overall comprehension process through planning, monitoring and problem-solving and evaluating listening.

In the processes in Figure 2.9, learners try to perceive the input by bottom-up processing relying on linguistic knowledge. Then, the listeners parse or segment the sound input into meaningful units by using linguistic knowledge and word candidates from long-term memory through top-down processing. Consequently, the listeners further use top-down processing to relate the parsed units to their prior knowledge to interpret the meaning for utilisation. Throughout these processes, it is the metacognition which regulates the overall listening comprehension (Vandergrift & Goh, 2012).

Figure 2.9: Interactions of cognitive processes and knowledge sources during listening comprehension (Vandergrift & Goh, 2012 p.27)
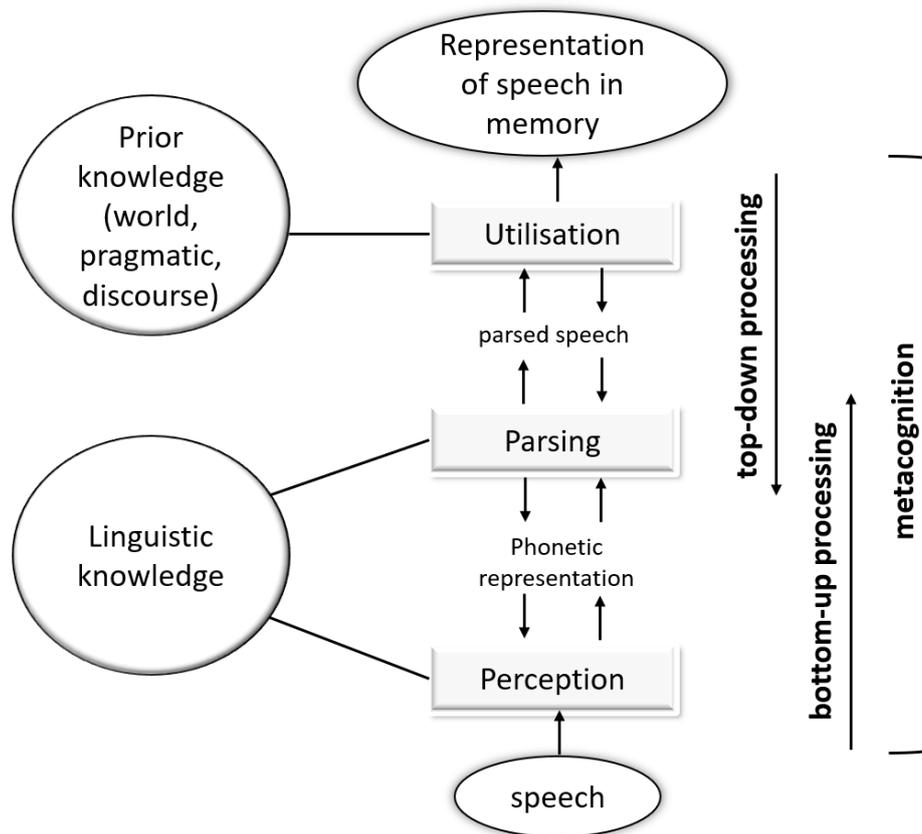
For second language listeners, Goh (2008) described metacognition as the listener awareness of knowledge and processes involved in comprehension and the capacity to regulate these processes. Similar to Flavell (1979) and Wenden (1998), Goh describes metacognitive knowledge to involve personal knowledge, knowledge about the task and knowledge of relevant strategies to be used. This metacognitive knowledge is exercised via the metacognitive strategies of planning, monitoring and evaluating to regulate listening comprehension (see Figure 2.10).

Figure 2.10: Metacognitive instruction in second language listening (from Goh, 2008 p.198)

Metacognitive Approach to L2 Listening	
Metacognitive Knowledge	**Metacognitive Strategies**

Metacognitive Knowledge

Person Knowledge
• Examines personal beliefs
about self-efficacy and self-concepts with
regard to listening in a second language
• Identifies listening problems, causes and
possible solutions

Task Knowledge
• Experiences mental, affective and social
processes involved in listening
• Differentiates different types of listening
skills (e.g., listening for details, listening for
gist, listening to infer information)
• Analyses factors that influence listening
performance (e.g., speaker, text, interlocutor,
strategy)
• Compares and evaluates ways to improve
listening abilities outside formal instruction

Strategy Knowledge
• Identifies strategies that are appropriate for
specific types of listening task and problem
• Demonstrates the use of strategies
• Identifies strategies that may not be
appropriate for their learning style or culture

Metacognitive Strategies

Planning
-Previews main ideas
-Rehearses language needed for the task
-Identifies important parts of input to attend
to
-Sets personal goals for listening
development
-Seeks appropriate opportunities for listening
practices

Monitoring
-Checks understanding of the message by
drawing on appropriate sources of
knowledge (e.g., context, factual, linguistic)
-Checks the appropriateness and accuracy of
one's understanding against old and new
information
-Considers progress of listening development
in light of what has been planned
- Assesses chances of achieving learning
goals

Evaluating
-Determines the overall acceptability
of one's understanding and interpretation of
the message/information
-Checks the appropriateness and accuracy of
one's understanding against old and new
information
-Assesses the effectiveness of strategies for
learning and practice
-Assesses the effectiveness of one's overall
plan to improve listening
-Assesses the appropriateness of learning
goals set

Knowing this information, one may ask how metacognitive awareness about listening can affect the learner's listening comprehension outcome? According to Goh (2008), metacognitive awareness influences the way the listeners approach the listening task. For instance, the learners with metacognitive awareness will plan, monitor and evaluate what they do more effectively, compared to those who take a random and incidental approach to listening. Such awareness helps learners to be more orchestrating and reflective in their listening

(Vandergrift, 2007) and improve learners' confidence in listening (Graham & Santos, 2015). The strategic thinking, self-belief and reflection promoted by exercising metacognition can help learners to better self-regulate their listening and build comprehension (Vandergrift & Goh, 2012).

2.2.2.3 Application of metacognitive approach for teaching and learning

It might be apparent now how metacognition can play different significant roles in second language listening. The next question might be what strategies to include and how these metacognitive principles could be applied in teaching and learning.

Graham and Macaro (2008), from a review of previous literature, identified strategies repeatedly regarded as essential for listening comprehension. The four essential strategies are the following.

-Making predictions about likely content

-Selective attention to certain aspects of the passage

-Monitoring and evaluating comprehension

-Using a variety of clues (linguistic, contextual, background) to infer meaning.

These strategies have been reported to be helpful for EFL listeners. They were also included in one of the most widely applied metacognitive listening models proposed by Vandergrift (2004, 2007). To guide the second language learners in their listening processes, Vandergrift introduced a seven-stage procedure of metacognitive instruction and its related metacognitive strategies (Figure 2.11).

Figure 2.11: Stages of metacognitive instruction and related strategies (Vandergrift, 2004 p.11)

Pedagogical stages	Metacognitive strategies
Planning/predicting stage 1. Once students know the topic and text type, they predict the types of information and possible words they may hear.	1. Planning and directed attention
First verification stage 2. Students listen to verify their initial hypotheses, correct as required and note additional information understood.	2. Monitoring

3. Students compare what they have written with peers, modify as required, establish what still needs resolution, and decide on the important details that still need special attention.	3. Monitoring, planning and selective attention
Second verification stage 4. Students verify points of disagreement, make corrections, and write down additional details understood.	4. Monitoring and problem solving
5. Class discussion in which all contribute to the reconstruction of the text's main points and most pertinent details, interspersed with reflections on how students arrive at the meaning of certain words or parts of the text.	5. Monitoring and Evaluation
Final verification stage 6. Students listen for the information which they could not decipher in the class discussion.	6. Selective attention and monitoring
Reflection stage 7. Based on the discussion of strategies used to compensate for what was not understood, students write goals for the next listening activity.	7. Evaluation

The procedure in Figure 2.11 provides a coherent sequence of pedagogical stages and identify metacognitive strategies relevant at each stage. There are three verification stages for students to monitor and check their performances and understanding. These verification stages occur both individually while working with the listening task and in pair/group discussions after a listening attempt. However, the procedure may not overtly pinpoint the sub-strategies in each metacognitive process which will illuminate what strategies should be used at which point. Thus, it is the teachers or programme developers' decision to incorporate appropriate strategies for each metacognitive process to suit their course and learners.

Apart from the pedagogical model through which the metacognitive instruction would be applied, implementational conditions should also be considered. Three principles for successful metacognitive implementation proposed by Veenman et al. (2006) can be applicable to language learning. The suggested conditions are as follows:

"1. Embedding metacognitive instruction in the content matter to ensure connectivity

2. Informing learners about the usefulness of metacognitive activities to make them exert the initial extra effort, and

3. Prolonged training to guarantee the smooth and maintained application of metacognitive activity" (Veenman et al., 2006 p.9).

The section has presented a practical procedure for metacognitive instruction to be applied in listening classrooms. A clear instructional sequence can minimise the students' cognitive load (Ashman, 2015). The conditions for which the approach will be applied should be well-thought-out. Applying it with specific content related to the students' background and interest can help students make the connection more easily. How metacognitive awareness can be useful to learning should be emphasised to the learners to encourage involvement. Moreover, the training should span over an adequate period of time to maintain self-regulated learning behaviours.

2.2.3 Models of metacognitive interventions for EFL learners

The idea of metacognitive and self-regulatory development has received enormous attention in EFL research and practices. According to a systematic review of metacognitive interventions among EFL learners (to be presented in detail in Chapter Three), several metacognitive instruction models have been applied in EFL contexts. This section will give an overview of the models commonly applied in EFL studies. Exploration of the different models is aimed to provide insight into the essential and helpful design elements for effective metacognitive instruction.

2.2.3.1 Cognitive Academic Language Learning Approach (CALLA)

One of the most popular models is the Cognitive Academic Language Learning Approach (CALLA) model introduced by Chamot and O'malley (1994). The model was originally designed to prepare students with limited English proficiency to participate in the classrooms where English is used for all content-area instruction. Later, it was modified to cover wider cogitative strategies and to be applicable to learners of English as a second or foreign language. This is presumably one of the reasons for its wide application by EFL educators. The model is featured by five recursive phases, namely, preparation, presentation, practice, self-evaluation and expansion. It could be seen that the five phases resonate with metacognitive processes in the metacognition and self-regulation literature. Each phase encourages the learners to apply different strategies inclusive of cognitive, metacognitive and socio-affective dimensions. The CALLA model can be summarised in Figure 2.12. The model

has been applied in various language skills such as reading (e.g., Al-Ghazo, 2016; Tavakoli & Koosha 2016), listening (e.g., Abdelhafez, 2006; Khonamri, & Ahmadi, 2015), writing (Panahandeh & Asl, 2014), vocabulary (e.g., Younsi, 2017).

Figure 2.12: CALLA model based on Chamot and O' Malley (1994)
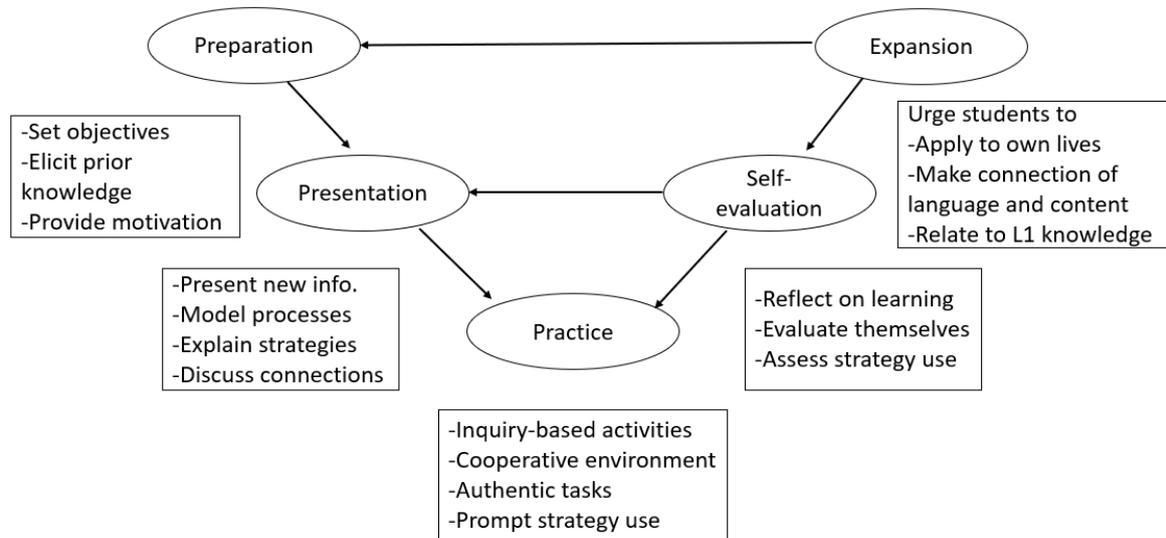
2.2.3.2 Metacognitive Pedagogical Sequence (MPS)

Another widely used model is the Metacognitive Pedagogical Sequence (MPS) proposed by Vandergrift (2004) and adapted in Vandergrift and Goh (2012). The model is specifically aimed at providing a more holistic approach to teaching listening to second language learners. It has widely been applied in EFL studies (e.g., Bozorgian & Alamdari, 2017; Chou, 2017; Rahimirad & Shams, 2014; Tanewong, 2019). Influenced by Flavell (1979) and Wenden (1998), the model is composed of and sequenced by the key metacognitive processes, i.e., planning, monitoring and problem-solving and evaluating (see Figure 2.13). In this model, students are prompted to plan before the first listening and monitor themselves while listening. After each listening attempt, verification stages are included to allow students to monitor their progress and make an adjustment plan to better cope with future listening. After all the listening attempts, the final verification stage is set to allow students to reflect on and evaluate themselves and make adaptive plans for future tasks. This pedagogical sequence is highly supportive of cultivating self-regulated listeners (Wang, 2016).

Figure 2.13: Metacognitive Pedagogical Sequence (MPS) (from Vandergrift & Goh, 2012 p.109)
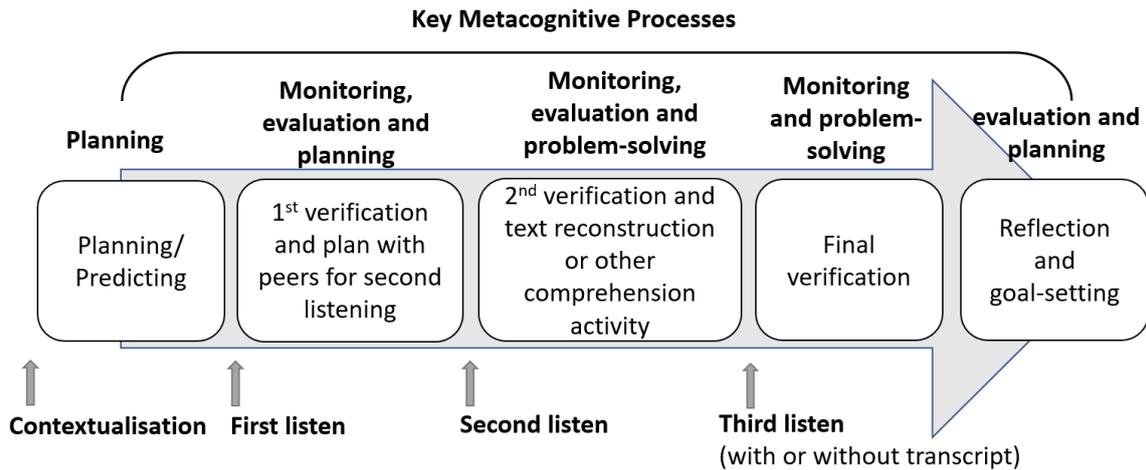
2.2.3.3 Think before, think While, think After (TWA)

The Think before, think While, think After (TWA) model is usually credited to Harris and Graham (1999) which is largely influenced by the Know, Want to know and Learn (KWL) strategy proposed by Ogle (1989). While Harris and Graham initially used the model with learners with learning difficulties, in the second language learning the model is mostly applied in reading research as well as writing. The model emphasises supporting students to activate their prior knowledge, set goals for learning and reflect on what they eventually learn (Mason, 2013). These processes occur during the six-stage procedure, i.e., developing pre-skills, discussing strategies, modelling strategies, memorising strategies, guided practice, and independent practice (see Figure 2.14).

Figure 2.14: TWA model, based on Mason (2013, p.126)
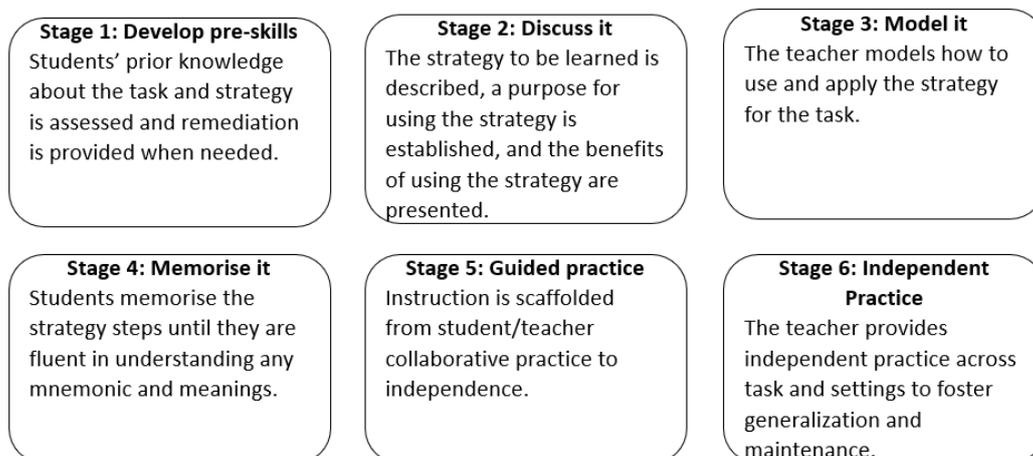
From the recently presented models, it could be seen that there are some shared components among them and some features unique in each one of them. Through reviewing instruction models, Graham and Macaro (2008) encapsulate fundamental features of metacognitive and self-regulatory instruction as follows:

-consciousness raising

-modelling strategies used by teachers and more able peers

-guided and structured practice

-action planning, goal setting and evaluation.

The presented existing models have provided clearer examples of what to be considered to include in a metacognitive and self-regulatory intervention and some reasons behind them. However, simply adopting or adapting a model from elsewhere might not be enough because the specific needs of the learners whom the intervention is designed for must be kept in mind (Richards, 2001). Therefore, the reflections from the models presented in this section, in combination with the lessons learnt from the systematic review in Wongdaeng (2022) which will be presented in Chapter Three as well as the reflections from the pilot study mutually inform the design of the intervention of the main trial which will be discussed next.

2.2.3.4 PMER model for metacognitive instruction for EFL listening

The principles for metacognitive instruction as well as the pedagogical models presented earlier provide beneficial guidelines for teachers to consider in organising lessons to promote metacognition and self-regulated learning. As a product of exhaustive review and discussion of theoretical principles for metacognitive development and pedagogical application as well as the lessons from a pilot study (to be presented in Chapter Four), a model of metacognitive instruction was developed by the researcher. This model formed the basis of the intervention in the main trial of this study.

To optimise the benefits of metacognitive and self-regulatory development, the instruction should be organised in a way that could guide the learners on what strategies to use (declarative knowledge), how to use them (procedural knowledge) as well as when and why (conditional knowledge) to use them (Veenman et al., 2006). Therefore, the PMER model (see Figure 2.15) used in the intervention is composed of the processes of *planning* which involves self-analysis, task analysis and strategic planning, *monitoring* which encompasses self-control

and self-observation and *evaluation* via self-assessment, self-instruction and self-reaction. Another key element which provides a more accurate evaluation of one's own learning is *retrieval practice*.

Figure 2.15: The PMER model for metacognitive instruction in second language listening
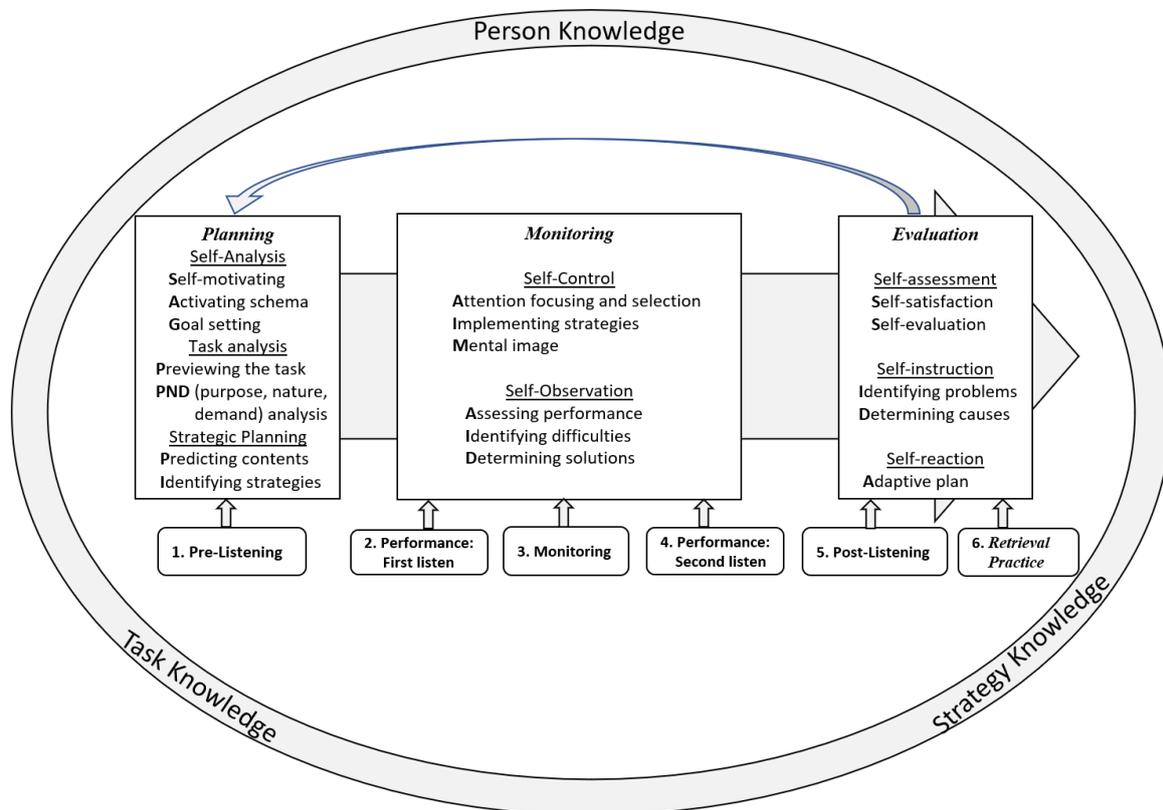
The design of the Plan, Monitor, Evaluate, Retrieve (PMER) model is informed by the principles of metacognition and self-regulation proposed by both cognitive development and language acquisition scholars. The first three processes in the model are the most common metacognitive strategies. The strategies under each process are largely influenced by Zimmerman (2000)'s model for self-regulated learning. With second language listening in view, the model differs substantially from Zimmerman's as it encompasses different strategy sets and is expressed in a mnemonic manner to aid memory. The model also includes a pedagogical sequence, in line with Vandergrift and Goh (2012) to resonate with the actual classroom practices. The role of retrieval practice is the key distinctive feature of the model which is believed to consolidate the learning of strategies. Moreover, the model realises the significant role of metacognitive knowledge as the fundamental element in each metacognitive process (Pintrich, 2002; Quigley et al., 2018). The model recognises and represents the operational definition of metacognitive awareness for this study which considers both

metacognitive knowledge and metacognitive strategies as the core components of the approach, in accordance with Flavell (1979) and Wenden (1998).

As the teachers who would attempt to use the PMER model may not be familiar with a metacognitive approach, clear guidance on the procedure at each pedagogical stage is helpful. To facilitate teachers' application of the PMER model in the listening lessons, the researcher also developed a teacher framework to guide them on what should be done in each pedagogical stage (see Figure 2.16).

Figure 2.16: Teachers' framework of the PMER model for metacognitive instruction
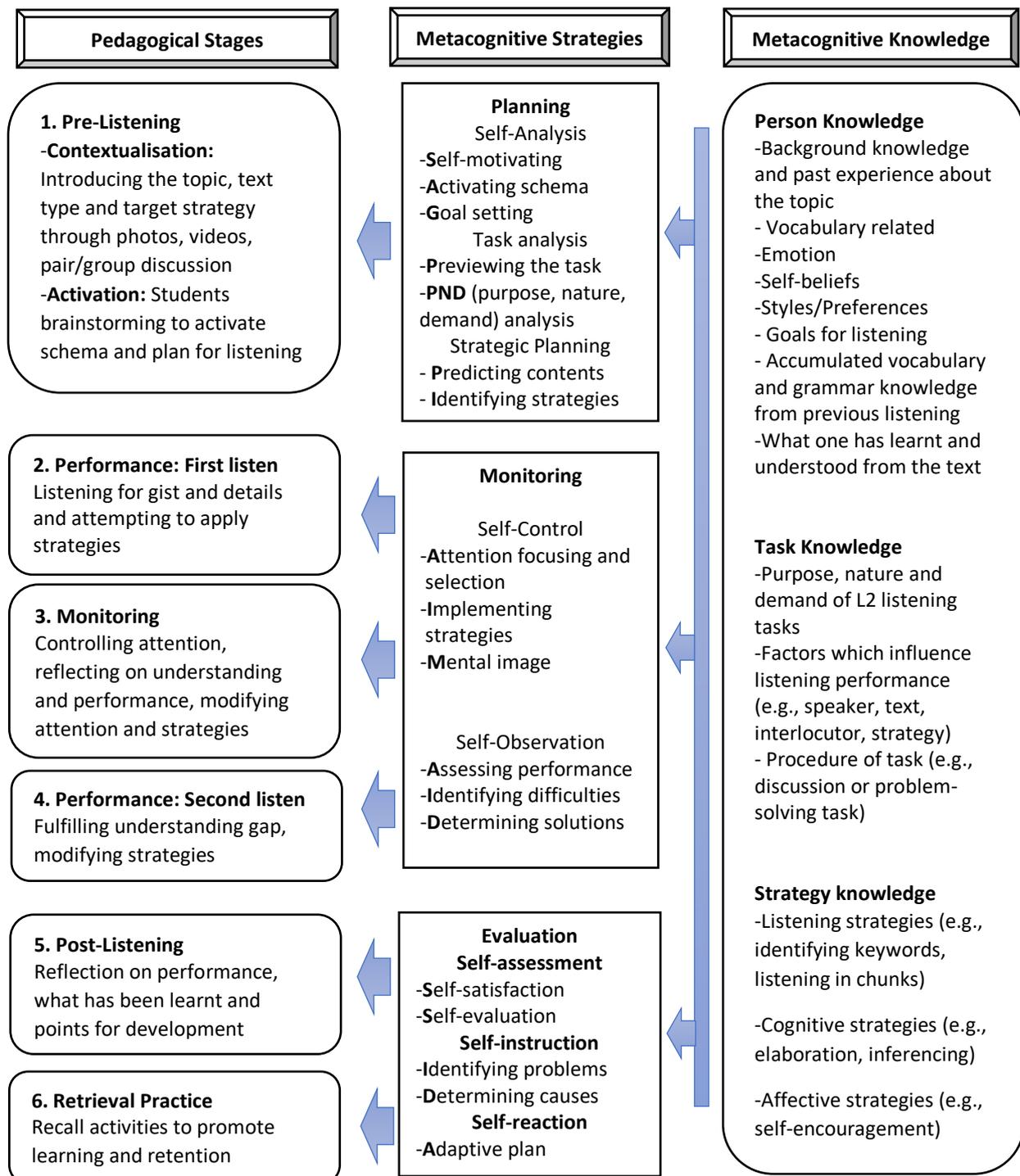
In line with the PMER model, the teacher framework in figure 2.16 provides a clearer picture for the teachers of what to do in each pedagogical stage and what metacognitive strategies to focus on at each stage. The metacognitive knowledge about person, task and strategies which is required in metacognitive regulation is also elaborated. The PMER model and the detailed teacher framework are aimed to provide learners and teachers with adequate explicit support to engage in metacognitive processes in learning and teaching.

Despite detailed information, it is important to emphasise that the ultimate aim of metacognitive instruction is not about the specific terms or skills used in the intervention. It is the positive dispositions such as self-reflection, strategic thinking to overcome difficulties and self-belief in one's own capacity which should be highlighted (Wongdaeng, 2021). Not only the knowledge about strategies but also the ability to apply appropriate strategies for different tasks is required to develop mastery in language learning (Sternberg, 1998). Such positive dispositions are useful for not only EFL listening but can be transferable across contexts (Perkins & Salomon, 2012).

2.2.4 Evidence of the effectiveness of metacognitive instruction in EFL contexts

A lot of studies have reported the positive results of metacognitive approaches for improving learning (e.g., de Bruijn-Smolders et al., 2016; Dignath & Buttner, 2008; Higgins et al., 2005; Perry et al., 2019). For instance, in a systematic review by de Bruijn-Smolders et al. (2016) which only included experimental studies in higher education with a comparison group and random assignment, 12 from 13 effects about metacognitive strategies found positive effect sizes. However, these studies are not directly inclusive of EFL studies, not to mention how trustworthy those results are. Therefore, the evidence of the effectiveness of the metacognition and self-regulation programmes in EFL contexts needs further thorough investigation.

In EFL contexts, the approach has also been widely applied (e.g., Chou, 2017; Cross, 2011; Rahimirad & Shams, 2014). During the review of literature, some studies in Thai EFL contexts on the topic were also found (e.g., Sodachan & Chayanuvat, 2018; Tanewong, 2019). Nonetheless, this arbitrary selection of studies cannot represent the overall effectiveness of the approach in EFL settings. Synthesised evidence from a systematic approach to literature review would provide a more comprehensive indication of its effectiveness in EFL contexts (Plonsky, 2011).

Some synthesis studies relevant to the topic were also identified. A meta-analysis by Raoofi et al. (2013) focuses on the effectiveness of metacognition on second/foreign language learning. The findings from the 33 included studies suggested that metacognitive interventions have the potential to promote language performance, but the effectiveness for enhancing metacognitive awareness is unclear. However, the review includes non-comparative studies and does not provide sufficient data for estimating effect sizes. Therefore, the evidence of the effectiveness remains obscure.

Another meta-analysis by Plonsky (2011) focuses on strategy-based instruction in second language learning and one main variable included for analysis is metacognitive instruction. The review applied more robust methods by including only studies with a comparison group, having an inter-rater assessment for the identification and extraction process and reporting the effect sizes of the impact. The provision of inter-rater reliability can minimise selection bias in the review (Torgerson, 2003). From the 26 studies on metacognitive interventions, a small effect size was found (ES 0.24). However, in the presentation of results related to metacognitive strategies, insufficient information is provided about characteristics such as outcome variables and age group of participants. This leaves the evidence of impact on learning outcomes inconclusive.

Despite numerous EFL studies applying the metacognitive approach in their work, studies synthesising the effectiveness of metacognitive interventions in EFL contexts remain limited. Thus, more synthesis studies with systematic and rigorous methodology on the topic would be required because evidence synthesis can provide a clearer and more accurate understanding about the area to be studied (Gorard, 2018). The systematic review is the most rigorous approach to searching for, identifying, quality appraising and synthesising all relevant studies in a field (Torgerson, 2003) and can provide a basis for planning and interpreting new primary research (Gough et al., 2012). Therefore, a systematic review on the effectiveness of metacognitive interventions in tertiary EFL contexts was conducted which will be presented in detail in Chapter Three.

2.3 Research context of the main study

To investigate the role of metacognitive and self-regulatory development in the Thai EFL context, a cluster randomised control trial was implemented in light of the systematic review's findings. The trial was conducted with a group of EFL learners in universities in the

southernmost area of Thailand. This section presents the background of the research context for the trials and the reasons for why the context and certain agenda are focused.

2.3.1 English language education in Thailand

English is the most important foreign language in Thailand and is taught informally from kindergarten and is currently a compulsory subject since grade 1 in the basic education curriculum (Ministry of Education, 2008). The compulsory education includes six years in primary education and three years in lower secondary education. After that, students can choose to continue in upper secondary education for three years or switch to a vocational route (OEC, 2017). Then, students who complete the upper secondary education or equivalent can continue to tertiary education on meeting the admission criteria, one of which is English proficiency. A bachelor's degree usually takes four years except those in medicine and education which may take longer. By the time students graduate with a bachelor's degree, they will have studied English for at least 15 years.

The teaching of the English language has been present in Thailand for more than a century (Sukamolson, 1998). English language policy in Thailand has gone through multiple changes. Since the education reform through the National Education Act 1999, English has been given important recognition in the national core curriculum (Darasawang & Watson Tod, 2012). However, during the early period, classroom teaching was still content-based and form-focused instruction still seemed to be the norm (Mackenzie, 2002). This led to a revision in the 2008 curriculum which provided more explicit goals and standards and allowed more freedom for teachers to design their own syllabus and pedagogic methodology (Nonthaisong, 2015). The increasing role of ASEAN (the Association of Southeast Asian Nations) further reinforced the important status of English and led to more policies emphasising English such as declaring 2012 the *English speaking year* (Baker & Jarunthawatchai, 2017).

Since the post-industrial world, the status of English education in Thailand, presumably similar to many countries in East Asia contexts, has primarily been viewed as a way to improve human capital while the purpose of internationalisation is referred to but not obviously catered for in practice (Byram, 2008). For several reasons such as the historical backgrounds and the influential role of English in economic and educational arenas, English has been considered a global language (Crytal, 2003). This has potentially driven English language education globally. Byram (2008) argues that foreign language learning in the globalised era does not only involve developing linguistic competences but inevitably entails *tertiary socialisation*

which influences the learners' focus beyond their own society into experiences of other cultural beliefs. He maintained that foreign language education in the globalised era should involve educating and socialising people into *intercultural citizenship* or "competences in order to be able to act sensibly in and across political entities, at whatever level" (Byram, 2008 p.157).

Proficiency in English of Thai citizens is considered one of the key measures to keep the country competitive in the global economy and this belief led to a massive investment by the government in English education (Hayes, 2016). After the official establishment of the ASEAN as one economic and cultural community in 2015, English, which was declared an official lingua franca of the region, was deemed even more important to Thais, especially the young and the working-age generations. However, the resources devoted to English education do not seem to have improved the achievement and proficiency of Thai learners of English (Kaur et al., 2016). The general English proficiency of Thai students has been ranked consistently lower for three years in a row since 2018, according to the EF English Proficiency Index ("English skills drop again", 2020). This shows a huge gap between the expectation from graduates and the actual English proficiency that they have.

Despite extensive investment and a growing number of research studies to improve the effectiveness of English language education in Thailand, noticeable improvement has not yet been seen. Such failures may be explained by several reasons. One of the reasons relates to the teacher quality and teacher development for using the English language in their teaching to an appropriate level. In an online assessment of civil servant English teachers by the Ministry of Education in 2015, more than 90 per cent of the teachers failed to achieve the target level of B2 of the CEFR framework (Franz & Teo, 2018). Likewise, a survey with a group of high school English teachers by Noom-Ura (2013) suggest that proficiency development and strategies for English communication are the top development needs expressed by the teachers. Teacher quality has inevitably affected students' learning (Kaur et al., 2016). Their insufficient proficiency can cause teachers to rely heavily on grammar-translation and repetition drill methods which do not adequately promote the target communicative competence (Nonthaisong, 2015).

Another factor which may hinder the English language education of the country is the disparities in achievements by urban schools which are usually attended by more affluent students and rural or regional schools which accommodate less well-to-do students (Hayes, 2010; Fry et al., 2018). This is reflected in the EF English Proficiency Index (2019) which

examines the English skills of full-time students in secondary schools and universities in non-English speaking countries around the world. The EF EPI (2019) reports unparallel proficiency levels between the central and more regional areas in Thailand (see Figure 1.2 in Chapter One). The south and the northeast regions have the lowest English proficiency level compared to other regions.

The rural-urban divide remains one major threat to the development of the overall English proficiency of Thais (Baker & Jarunthawatchai, 2017). This exhibits a form of inequality in education caused by geographical location and socio-economic status. The clustering of students with similar disadvantaged backgrounds can negatively affect their life chances (Gorard & See, 2013; Siddiqui, 2017). This clustering is more clearly observed in higher education institutions in the up-country areas. According to a Finnish education expert, Pasi Sahlberg, the inequality of educational opportunities in urban and rural areas is one of the main problems of Thai education (Yokakul, 2017).

In addition, the quality of English language teaching (ELT) research has a significant influence on the stagnation of English proficiency development. From one of the researcher's systematic reviews (Wongdaeng, 2020) synthesising the overall effectiveness of EFL task-based interventions on language learning and another review of metacognitive interventions for EFL learners (see Chapter Three), studies from the Thai context hardly satisfy the inclusion criteria in both reviews. That is because most studies which investigate the effectiveness of a pedagogical innovation or intervention do not include a comparison group, disqualifying the inclusion criteria. Without counterfactual evidence from the comparison group, the reported impact of any intervention cannot be secured (Gorard, 2013). Similarly, in other systematic syntheses on strategy-based instruction by Plonsky (2011) and Shirvan et al. (2016), studies from Thai contexts were hardly included. This echoes the insufficient quantity and quality of EFL studies in the Thai context and the overlooked role of synthesising existing evidence. Indeed, well-planned synthesis studies on language education can provide compelling information for theory and future research, practice and policy (Oswald & Plonsky, 2010). Therefore, a systematic review was conducted as a synthesis of the existing evidence which will be presented in Chapter Three.

2.3.2 Southernmost provinces of Thailand and opportunities for quality English education

The southernmost provinces of Thailand usually refer to the three border provinces, namely, Pattani, Yala and Narathiwat which share their borders with the north of Malaysia (Brooks & Sungtong, 2015; Pherali, 2021). The population is composed of people with different faiths and cultures but is predominantly Muslims whose mother tongue is a dialect of Malay while Thai is still an official language (Huebner, 2019). Public schools in the areas use the same core curriculum as in other parts of the country but there is also an abundance of private schools which provide religion-integrated curriculum (Liow, 2010). Each province is fortunate to have one public university which provides opportunities to the locals for higher education.

In terms of academic achievement, the students in this area generally perform quite unsatisfactorily on the national standardised tests. Based on the data from the National Institute of Education Testing Service (NIETS), which administers high-stake tests in English and other subjects to assess the outcome of schooling which are used for university admission, the average English score of the test takers in all of the three provinces rank in the bottom group of all provinces nationwide for three continuous years in 2016, 2017 and 2018. This low achievement has been supposedly linked to the reported unrest situations in the area (Tuntivivat, 2016). The area has been affected by social and political instability due to complex conflicts ranging from aggressive resistance of some alleged insurgency groups to human right abuses by the people in authority (Seiff, 2016), cultural repression (Pherali, 2021) and socioeconomic marginalisation (Tuntivivat, 2016). The area has been under heightened security measures which have deterred economic and educational development ("Education in Thailand's deep south", 2017).

The higher education institutions (HEIs) in the area have provided a wide range of education programmes and have followed the regulations from the national Higher Education Commission. Generally, two or more English courses are mandatory for students to complete a degree in line with the national policy to produce graduates with a satisfactory level of English proficiency. English is one of the subjects which can cause concerns to university students in the southernmost provinces because their English proficiency is generally low.

In these HEIs, the student intakes are mainly from these disadvantaged provinces (see Table 2.4). This seems to show a form of segregation by disadvantaged geographical

backgrounds. Moreover, from the admission English score of one cohort of the students in this area (see Table 2.5), most of the students had low proficiency in English. This demonstrates another kind of segregation by academic performance. These segregations can affect their learning experiences and opportunities during study and opportunities in life (Gorard, 2018).

Table 2.4: Percentage of students by geographical backgrounds in the southernmost HEIs

Cohorts	Geographical Backgrounds	University A Students (%)	University B Students (%)
2018	Deep South	62.72	80.95
	Others	37.28	19.05
2019	Deep South	62.48	75.31
	Others	37.52	24.69
2020	Deep South	61.69	80.43
	Others	38.31	19.57

Table 2.5: Admission English scores of one cohort of the student participants

Score range	N of students
25 and lower	762
26-39	720
40 and higher	246

Segregation refers to the extent to which students with similar socio-economic backgrounds, usually the negative ones, are clustered within the same school or institution in each intake (Gorard & See, 2013). The segregation issue matters because ample evidence suggests that this clustering phenomenon has a clear link with educational participation and outcomes (Gorard, 2018). Attending school with low-achieving peers has a significantly detrimental effect on the achievements of an average student (Mendolia et al., 2018). The segregation by academic performance and income can hinder disadvantaged students from equal access to opportunities in life (Siddiqui, 2017). Likewise, geographical inequalities have been one of the significant problems in English language education in Thailand. (Hayes, 2010).

The data from Table 2.4 and Table 2.5 suggest that the English learners in the southernmost HEIs seem to be disadvantaged by academic and socio-economic segregations. Students' socio-economic backgrounds have a high association with their academic attainment

(Early et at., 2020). Students in the target contexts have relatively lower opportunities for practical learning resources and interactions in English with more-capable others inside and outside the classroom. Therefore, they may need more support for improvement in their English proficiency. In addition, the Higher Education Commission recently launched a new regulation in 2016, requiring students to achieve a certain level of a standardised English proficiency test designed by the university or other equivalent assessments for graduation. More details of the policy will be discussed in the next section. This causes more concern for the English learners in the southernmost areas, suggesting a need for support or intervention to help these students cope with the new challenge.

2.3.3 Fundamentals of education policy

Before discussing the Exit English Exam policy, a basic understanding of educational policy is beneficial. In a UNESCO framework for education policy planning by Haddad and Demsky (1995), education policy is defined as *an explicit or implicit single decision or group of decisions which may set out directives for guiding future decisions, initiate or retard action, or guide implementation of previous decisions* (p.18). In this definition, education policy is viewed as a set of principles or actions to be followed by stakeholders to achieve a particular educational goal. Trowler (2003) argued that education policy should not be seen as just a statement of actions to bring about desired goals but the ideology, power relation, conflicting interests and value systems between policy makers and those responsible for taking action need to be considered in the policy development process.

Cerych and Sabatier (1986 cited in Trowler, 2003 p.125) proposed seven strategies for successful policy implementation as the following:

- creating and sustaining the commitment of the people involved

- having clear and stable policy objectives

- ensuring that the policy innovation has priority over competing demands

- creating a real expectation of solid outcomes, not just a symbolic one

- assuring the correct and adequate causal theory underlying the policy

- allocating sufficient financial resources

- creating a stable environment within which policy is being implemented.

These factors are seen by Trowler (2003) as necessary but insufficient for appropriately putting policy into practice. He reiterated that the ideology and the power relation, conflicting interests and value systems between the authority in charge and the people responsible for taking action need to be considered in the policy development. Key stakeholders such as institutional leaders or teachers should not be treated as the passive implementers of policies but should be involved in the policy development at an early stage via representation from professional associations or interest groups, government policy forums and think-tanks (Bell & Stevenson, 2006).

Policy development involves multiple processes and stakeholders to be considered. Haddad and Demsky (1995 p.24) recommended a seven-process cycle for policy development. It includes:

1. Analysis of the existing situation,

2. Generation of policy options,

3. Evaluation of policy options,

4. Making the policy decision,

5. Planning of policy implementation,

6. Policy impact assessment,

7. Subsequent policy cycles.

As seen in the cycle, policy development is a dynamic process which involves consistent observations and evaluations throughout the process. Gorard (2018) emphasises that public policy such as education must be evidence-informed. New policies or interventions must be based on good existing evidence and be accompanied with rigorous evaluations for its impact to assess its cost-worthiness and provide implications for the policy cycle. Voices and participation from practitioners and stakeholders should be included early in the processes as well as during and after the implementation (Bell & Stevenson, 2006).

Considering policy in regard to language education, Kirkpatrick and Liddicoat (2019 p.3) described it as "a form of human resource development planning that operates to develop language abilities that a society identifies as important for social, economic, or other objectives". Spolsky (2004) asserted that one main feature of language policy is the interrelation of language practices, language belief and ideology and the language management or planning activities which cater for the practices and ideologies. According to Crandall and Bailey (2018), language education policies can involve the following:

• promotion of languages/varieties as mediums of instruction, including in bilingual/multilingual contexts;

• acquisition of additional (often official or international) languages;

• choice of languages for initial literacy and for teaching specific subjects such as science or technology;

• support for minority or indigenous/community languages;

• recognition of the linguistic resources of plurilingual individuals

• decisions for which languages and skills are assessed.

(Crandall & Bailey, 2018 p.1)

The language planning process can be top-down, also called macro, which involves managing, allocating and promoting languages for education; bottom-up, also known as micro, which involves decisions about language use in the classroom and meso level which relates to such decisions at the institution or programme level (Crandall & Bailey, 2018). The explicit language policies in policy documents, curricula and other educational texts essentially influence practices and arrangements in educational contexts at different levels (Kirkpatrick & Liddicoat, 2019). At times, the actual practices by the participants receiving the policy are neglected in the policy planning which can lead to a mismatch between the policy statement and the actual practices (Spolsky, 2004; Trowler, 2003). Thus, language education policy development should involve views from multiple stakeholders and be flexible enough to allow practical applications by different agents.

2.3.4 Exit English Exam Policy

The Exit English Exam policy was launched in 2016 by Thailand's Office of Higher Education Commission with an aim to elevate English language standards among higher education graduates (Wudthayagorn, 2019). The policy was initiated as a measure to prepare the graduates for the ASEAN community establishment and improve the country's competitiveness. In the policy statement which was translated into English by Baker and Jarunthawatchai (2017), Thailand's higher education institutions are demanded to:

- set a benchmark of English proficiency which students need to achieve

- arrange for students to take English tests to assess their proficiency levels. Available standardised English tests can be used or universities can develop their own tests corresponding to the Common European Framework of Reference (CEFR) or other standards.

- revise English teaching practices to facilitate students' learning.

From the policy aim and statement, it could be implied that the policy is seen as a set of actions to be followed. This top-down policy allows room for some operational decisions to be made by participants at the meso and micro levels. For instance, universities can use the available standardised tests or develop their own equivalent test. However, what seems to be missing is the supporting resources and moderating agencies which facilitate the implementation in the actual classrooms (Trowler, 2003). Following the view of Shohamy (2006), language assessment in this policy was used as a mechanism of language manipulation and an instrument for authoritarian power because passing tests can be perceived as compliance with the rule, irrespective of students' contextual needs and limitations.

While this policy is one of the attempts to improve the country's readiness and competitiveness for the establishments of the ASEAN economic community (AEC), the Eurocentric framework such as CEFR, TOEFL, IELTS, TOEIC was prioritised as a standard norm of assessment. This norm does not adequately realise the nature of the English used in the ASEAN region where English is mostly spoken among non-native speakers (Baker & Jarunthawatchai, 2017). Other university-proven tests are also accepted but such tests are usually based on the Eurocentric test norms. This demonstrates that despite the increasing role of English as a Lingua Franca (ELF) in this region, the ELF realm is not reflected in the testing and assessment where the standard norm is considered more appropriate. Moreover, the administration of standardised tests produces extra costs for universities. Students who have used up the free support from universities but still need to take more tests in order to meet the criteria may eventually have to pay for the test. This is another challenge for the stakeholders, especially those in a disadvantaged setting.

Another issue to be considered is the top-down policy imposed on a nationwide scale which may cause extra challenges to the practitioners in some regional areas. There are usually conflicts among policy makers and those who put it into practice about the important issues and problems of the relevant policy (Trowler, 2003). The policy as a set of actions like this results in a change or an addition of responsibilities for the participants especially those in the

less privileged areas who might have already been disadvantaged. For instance, those in marginal areas of the country might find it more difficult or problematic to put the policy into practice than the practitioners who have more access to resources for the policy implementation. It is reasonable that the policy allows meso-level interpretations about a practical course of strategies in their own contexts. However, support and access to relevant resources is one necessary factor for sensible policy management (Haddad & Demsky, 1995). Such top-down policy without adequate support especially for those in less privileged circumstances may even lead to a wider inequality gap in the system.

If English proficiency does increase one's life opportunities, the need-based support for different stakeholders may be more suitable. The supports given on the basis of the different needs of the stakeholders can provide equitable opportunities which can bridge the inequality gap by their socio-economic and geographical backgrounds (Ulterhalter, 2009). Educational equity matters as some evidence shows the countries which are performing well educationally are those with high quality and equity (OECD, 2012). Equity in education means fairness and inclusion to ensure that personal and social circumstances such as ethnic origin or socio-economic status are not an obstacle to achieving education goals and a certain standard of opportunities is accessible for all (Field, Kuczera & Pont, 2007). In an attempt to widen participation in higher education among the disadvantaged groups, Boliver et al. (2017) recommend reasonable contextualised concession for these groups to admit them into highly-recognised HEIs. Such concession should not be set arbitrarily but with an adequate level of required qualities for them to succeed to avoid putting students with concessions at unnecessary risk of failing to cope with the awaiting challenges (Boliver et al., 2021).

In this study, the PMER model of metacognitive intervention was introduced to a group of students in a disadvantaged area with an aim to support them with their learning. The positive dispositions, namely, self-efficacy, strategic thinking and self-reflection promoted in the intervention are hoped to equip students with skills helpful for coping with the new graduation requirement. This can also respond to the United Nations' Sustainable Development Goal (SDG) 4, which aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (UN, 2015).

2.4 Chapter Summary

The metacognitive and self-regulatory approach to teaching and learning has been widely applied in various disciplines, including the EFL contexts. Metacognition can be

understood as an awareness of one's own knowledge, experience and emotion and the ability to apply such awareness to regulate learning. The fundamental features promoted in the approach, namely, self-efficacy, strategic thinking and self-reflection are linked to enhancing meaningful learning. For a clearer picture of its effectiveness in tertiary EFL contexts, a systematic review was conducted, which will be presented in Chapter Three.

The PMER model of metacognitive instruction was designed in the main trial of this study for use with a group of tertiary EFL learners in a disadvantaged area of Thailand to assess its effectiveness for this group of learners. The intervention focuses on listening comprehension because studies on listening comprehension pedagogies remain minimal. With some evidence of the association between listening comprehension competence and overall language proficiency, the impact of the intervention on overall English proficiency will also be investigated.

The recent introduction of the Exit English Exam policy requires students to qualify for certain English proficiency criteria for graduation. Within this policy, the students in the southernmost areas of the country who have already been disadvantaged due to less access to resources and opportunities for communicative English use and the segregation by some characteristics such as low English proficiency may find it more challenging to cope with the new graduation criteria. The introduction of this intervention with guidance on learning strategies can be a helpful support for this group of students. The attempt can respond to the United Nations' SDG 4 to promote inclusive and equitable education for all.

CHAPTER THREE
SYNTHESIS OF EXISTING EVIDENCE

A systematic review synthesising the existing evidence is presented in this chapter. It begins with the rationale and scope of the review, followed by the design and methods of the review. The results which derived from both meta-analysis and thematic analysis are presented and discussed to provide practical implications for the main trial of the study. Finally, limitations are discussed, followed by a chapter summary.

3.1 Introduction

As presented in the previous chapter, metacognitive and self-regulatory instruction has been widely applied in EFL contexts and numerous studies have reported the benefits of the approach. However, studies attempting to synthesise the overall effectiveness of the approach are scarce. Existing reviews which include metacognitive interventions as one variable in the analysis (i.e., Plonsky, 2011; Raoofi et al., 2013; Shirvan et al., 2016) were drawn from both second and foreign language contexts and did not particularly look at the higher education sector. To capture a clearer picture of the existing evidence on the topic, as enquired in research question 1 of the thesis, a systematic review was conducted as part of this thesis to identify, examine and synthesise overall evidence of the effectiveness of metacognitive interventions on learning achievement in tertiary EFL contexts.

3.2 Rationales and scope of this review

In education, metacognition has been recognised as an area of development for learning. In the revised version of Bloom's taxonomy of learning by Anderson & Krathwohl (2001), metacognition is one of the key knowledge dimensions which interacts with cognitive activities during learning. For such establishment, research on metacognition and learning has been increasingly prevalent in various fields, including in EFL contexts. Despite an extensive quantity of empirical EFL studies on the topic, evidence of the impacts the approach has on EFL learners is still unclear because only a few studies attempt to rigorously synthesise the reported results. Reviews on the topic which focus on tertiary education level are even more scant. Therefore, the present review seeks to fill this gap by applying a systematic review approach to identify, analyse and synthesise all identified studies on the effectiveness of metacognitive interventions in tertiary EFL contexts.

The distinction between the term EFL and ESL is important in this review. The term EFL refers to the context where the use of English is commonly seen in classroom settings but rarely found in other real-life situations, while the term ESL is applied in the context where English is used as another common language apart from the mother tongue (Williams & Williams, 2007). With this definition, the status of English in countries such as China and Thailand, which can be considered EFL contexts, would be different from those in Singapore and South Africa, where English is more prevalent in daily use. The two contexts are distinct in several aspects. For instance, different sets of motivations and degrees of exposure to the target language give learners in ESL settings greater opportunities to practise their strategies and skills (Dornyei 2005; Yabukoshi 2021). Moreover, the findings from Plonsky (2011)'s review suggest that ESL learners seem to have better learning outcomes, compared to their peers in EFL settings. Therefore, this review focused on studies with EFL learners who seem to need more support to see if the metacognitive interventions are beneficial to them.

This review only included studies conducted in tertiary education. As EFL learners are generally less proficient compared to their ESL counterparts, learning a foreign language with a technique featuring thinking may sound difficult to students, especially younger learners. Even though some evidence from interdisciplinary synthesis suggests that the metacognitive approach may not be more difficult for younger learners (EEF, 2018), learners of a foreign language who still have limited competence in that language may find this kind of technique an overwhelming challenge. Moreover, the inclusion of studies with a similar context and participants' levels regarding the target outcome would enhance the homogeneity of the effects to be included in the synthesis (Plonsky, 2011). Furthermore, the findings will be helpful for the main trial of the thesis with participants in higher education.

3.3 Design and methods

This review was conducted with a systematic review design to investigate the potential effectiveness of metacognitive interventions on tertiary EFL learners. The systematic review is an appropriate approach to searching for, identifying, quality appraising and synthesising all relevant studies in a field (Torgerson, 2003) and can provide a basis for planning and interpreting new primary research (Gough et al., 2012). The approach can secure reliable findings through exhaustive and unbiased search and inclusion of studies as well as cautious attention to the quality of the included studies (Torgerson et al., 2019). This review addressed the following questions:

1. How effective are the metacognitive interventions in the tertiary EFL contexts and what is the quality of the evidence?

2. What are the possible factors or characteristics which contribute to the effectiveness of metacognition-based interventions for EFL learners?

3.3.1 Inclusion and exclusion criteria

Priori criteria for determining whether to include or exclude a study were established before the searching processes. These criteria are based on the *Population, Intervention, Comparison and Outcome* (PICO) categories (Torgerson et al., 2012) shown in Table 3.1.

Table 3.1: PICO inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
a. Population	- Learners of English as a foreign language (EFL) -EFL learners in higher education level	-Learners of English in the English-speaking countries or in countries where English is an official language such as in Singapore and South Africa (source: https://projects.ncsu.edu / grad/handbook/docs/official _language_english.htm) -Learners of English in bilingual settings -EFL learners in other educational levels
b. Intervention	-Use of metacognitive interventions to improve learning outcomes in English - Systematic reviews or meta-analyses of experimental studies complying with the criteria of this review	-No intervention being studied -Use of interventions other than metacognition - Focus only on strategy instruction without looking at the metacognitive processes
c. Comparison	-Include a comparison group -business as usual or alternative treatment given	-Pre/post, non-comparison group
d. Outcomes	-Language competences such as reading, listening, vocabulary as primary or secondary outcomes	- Language competence is neither primary nor secondary outcomes but other outcomes such as perceptions, satisfaction, motivation, awareness are instead targeted.

3.3.2 Identification of studies

The search for studies for review attempted to include all identified studies published from 1990 to 2019. 1990 was specified as the start date because studies of metacognition in second or foreign language learning was uncommon until the early-late 1980s (Oxford, 2011) and started to become more prevalent only afterwards. Relevant studies were identified from seven databases. Five of these databases were the most common databases among EFL authors. Another was a Thai journal database to allow more relevant studies in another language, in this instance Thai, to be identified. The last source was Google Scholar which can identify grey and unpublished literature to help minimise publication bias (Torgerson, 2003). The search took place in July 2020. The search string "metacogniti*" AND "English" AND "effect" OR "experimental" was used with some adaptations in some databases to ensure that the least possible relevant studies were missed. For instance, in some databases such as ThaiJo, which does not recognise the search term with an asterisk as in "metacogniti*", complete words, i.e. "metacognition" and "metacognitive", were used instead.

3.3.3 First and second stage screening

After searching through each database, the identified studies underwent deduplication and two screening stages. In the first stage, the titles and the abstracts were screened for relevance. Pre-established criteria were used to ensure efficiency and minimal bias in the review process (Torgerson, 2003). Irrelevant studies which were not congruent with the inclusion criteria were removed.

In the second screening, the full texts of the remaining studies were examined for further details based on the same criteria. Studies which did not meet the inclusion criteria were further excluded. The remaining studies were kept for data extraction and synthesis of the findings.

3.3.4 Data extraction and quality appraisal

The remaining studies were data-extracted based on a template adapted from the EPPI-Centre (2003). The extracted data included the number and characteristics of participants, the nature of included intervention and control conditions, outcome variables, outcome measures and results.

Another significant stage of this review is the quality appraisal of the included studies. This stage is important for a systematic review to obtain trustworthy evidence. Reporting findings without prudent attention to the quality of the obtained results could lead to having

weak or unreliable evidence, undermining or subverting the proposed conclusion (Gorard, 2014). The quality appraisal criteria in this review were adapted from Gorard (2014)'s 'Sieve' for research findings and the EEF (2019)'s classification of security of findings which shares a number of characteristics with Gorard's work. Such criteria were chosen because Gorard's sieve was specifically designed for causal research in education and public policy and the adapted elements by EEF (2019) on the 'design' render the criteria more harmonious for non-randomised studies which are highly prevalent in EFL studies and the guidance on 'validity' is explicit for a transparent judgement. Summaries of the extracted data and the quality rating of included studies are presented in Appendix 2.

3.3.5 Data Synthesis

There are several methods for synthesising the findings in a systematic review such as narrative synthesis, vote-counting and meta-analysis (Torgerson, 2003). Factors such as review questions and the homogeneity of the studies should be considered for the selection of an appropriate approach for synthesis (Thomas et al., 2012). For this review, a mixed-method approach combining meta-analysis and thematic synthesis was used for the synthesis. Meta-analysis can be used to identify the overall impact of studies in order to illuminate the enquiries 'does it work?' and 'how well does it work?' at the same time (Higgins, 2018). Thus, meta-analysis was conducted by calculating an average effect size of the included studies to provide an estimate of the overall impact of metacognitive interventions in tertiary EFL contexts inquired in the first review question. What effect size is and how it is calculated and interpreted are presented in section 4.6 of Chapter Four. Thematic synthesis can explain the mechanism and key variables identified in each study for the purpose of answering the second question on the possible factors contributing to the effectiveness of metacognition-based teaching and learning.

Some of the included studies report more than one outcome measure of treatment effect from the same participants. Treating multiple outcomes in one study as a separate study can cause more weight to be assigned to studies with multiple outcomes and lead to an inaccurate estimate of the summary effect (Borenstein et al., 2009). One way to avoid this issue is to calculate an average effect size for each study (Lipsey & Wilson, 2001). Thus, an average effect size was calculated for each included study with multiple outcome measures drawn from the same participants.

3.4 Results

3.4.1 Search results

The searching of electronic databases took place in July 2020. The total records of 2,942 were identified from the seven sources. The search results are presented in Table 3.2 below.

Table 3.2: The databases searched and the number of identified studies

Databases	N of Hits
Scopus	180
JSTOR	920
ScienceDirect (Elsvier)	1,315
Education Resource Information Centre (ERIC)	207
Web of Science	167
Google Scholar	146
ThaiJO	7
TOTAL	2,942

3.4.2 Results of screening

The overall results from each of the review processes are presented in the PRISMA diagram in Figure 3.1.

Figure 3.1: PRISMA flow diagram of the review processes (based on Moher et al., 2009)
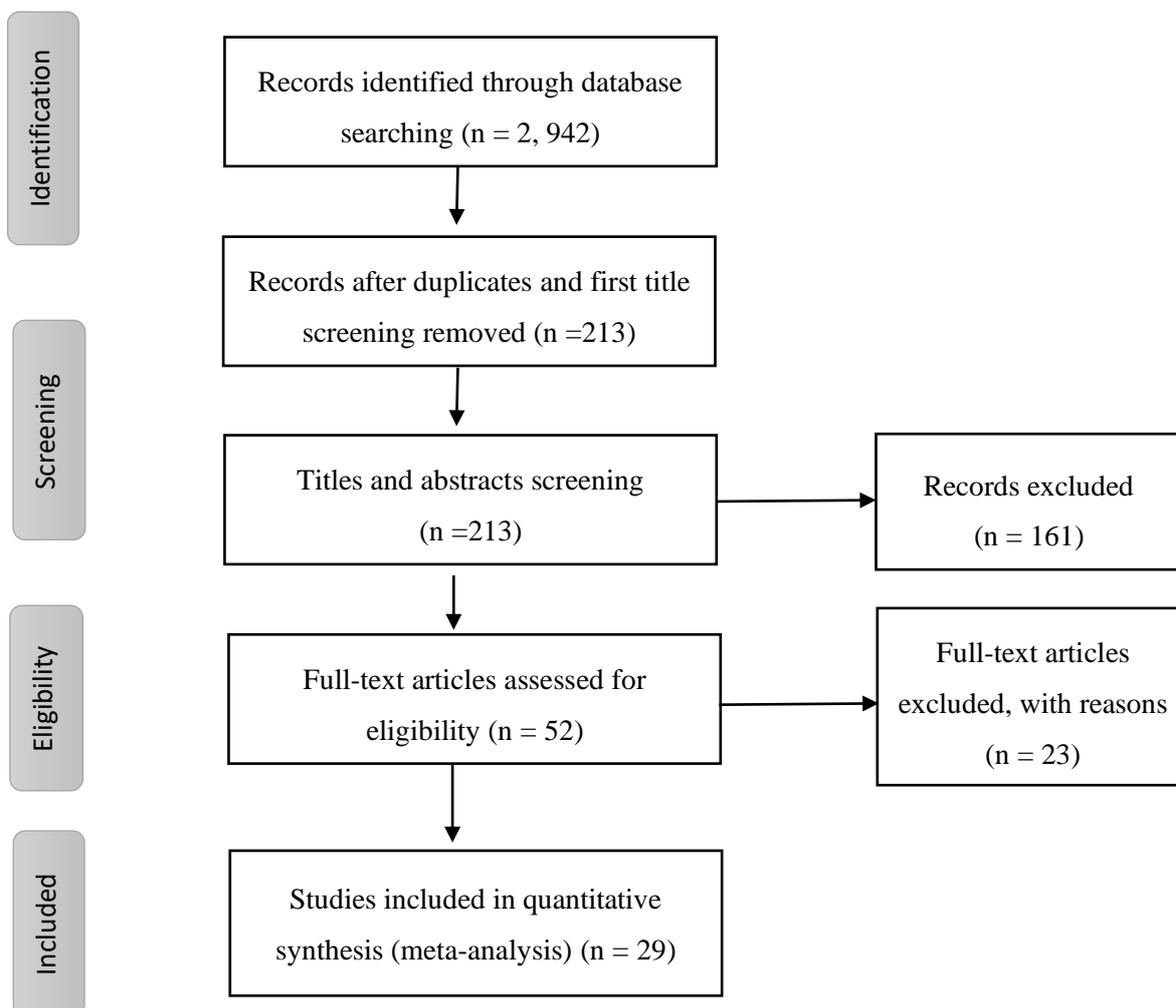
The 2,942 identified studies underwent deduplication and the first screening by titles and abstracts. Studies that did not comply with the inclusion criteria were removed, leaving 52 for further screening. In the full-text screening, an additional 23 studies were removed. Among these, 12 studies were based on strategy instruction without metacognitive strategies as the main components in the intervention constructs, six did not report adequate information to convert the results to Hedges' g for a compatible synthesis, four were non-randomised studies with an overly trivial sample size which may be at risk of chance finding (Torgerson & Torgerson, 2008) and could exaggerate the results (Coe, 2002) and one study happened to be the same as another included study despite different titles. Finally, 29 studies were included for data extraction and synthesis.

3.4.3 Results of data extraction and quality assurance

The extracted data with key characteristics of the included studies are tabulated and shown in Appendix 2. The key details are design, focus and length of intervention, scale, outcome and measures, validity assurance and results. The purpose of the quality appraisal was to illustrate how strong the reported evidence is, rather than to exclude any further studies.

Regarding the design, all of the studies are based on control trials satisfying the inclusion criteria. Without counterfactual evidence from the control group, the causal inference drawn from the results cannot be warranted (Shadish et al., 2002). Among these, four studies increased their design strengths by randomly allocating the participants to groups. However, none of them made clear whether the group assignment was concealed from the participants. Regarding scale, almost all of the studies have a small sample size which puts them at risk of having unsecured results (Gorard et al., 2017). None of them estimated the number of students needed to treat for the effect size they project to detect, as suggested by the EEF. In terms of outcomes, all studies had pre-specified outcomes and more than one third relied on standardised or independent outcome measures. Most studies addressed certain aspects of validity assurance, but almost a quarter carried the possibility of the experimenter effect. Dropouts and fidelity are hardly mentioned in any study. However, the pre and post response rates are similar in most studies. Overall, the studies are fairly good on design and outcomes, but had lower ratings on scale, fidelity and validity. Therefore, the results may be able to provide an indication of the promise of metacognitive interventions, but due to limited rigour, neither firm conclusions nor a definitive interpretation can be drawn.

3.4.4 Results of meta-analysis

The overall effect of metacognitive interventions synthesised by a meta-analysis is presented in Figure 3.2 below. The evidence strength of each study is represented by stars. A maximum of four stars indicates the strongest evidence while zero is the weakest evidence.

Figure 3.2: Forest plot of the overall effect of metacognitive interventions

Authors, year, outcome	Forest plot of effect size	Effect size	Weight	Evidence strength
Abdelhafez (2006), Listening & Reading		1.62	3.17	★★
Al-Ghazo (2016), Reading		3.77	2.89	★
Altay & Altay (2017), Reading		0.22	3.14	★﹂
Bozorgian & Alamdari (2017), Listening		1.17	3.24	★﹂
Cabrera-Solano et al. (2019), Vocabulary		0.79	3.08	★
Chang (2010), English & Reading		0.28	3.22	★﹂
Chou (2017), Listening		0.51	3.22	★﹂
Farahain & Avarzamani (2018), Writing		2.71	3.06	★
Habibian (2015), Reading		1.32	3.08	★
Khonamri & Ahmadi (2015), Listening		2.28	2.93	★
Kobayashi (2018), Listening		0.62	3.11	★
Msaddek (2016), Reading		1.77	3.21	★﹂
Nakatani (2005), Speaking		0.64	3.16	★﹂
Panahandeh & Asl (2014), Writing		0.66	3.16	★﹂
Rahimirad & Shams (2014), Listening		1.16	3.10	★﹂
Rasekh & Ranjbary (2003), Vocabulary		0.96	3.12	★
Razi & Çubukçu (2014), Reading		0.46	3.23	★﹂
Roohani & Asiabani (2015), Reading		1.42	3.15	★﹂
Seifoori (2016), Speaking		0.48	3.25	★★
Shirvan et al. (2016), English		0.82	3.29	★﹂
Tanewong (2019), Listening		0.52	3.17	★﹂
Tavakoli & Koosha (2016), Reading		1.50	3.21	★★
Teng (2016), Writing		2.32	3.12	★﹂
Teng & Reynolds (2019), Reading 1		6.15	2.72	★﹂
Teng & Reynolds (2019), Reading 2		1.35	3.19	N/A
Tsai & Talley (2014), Reading		2.14	3.20	★﹂
Wang (2017), English		0.19	3.18	★﹂
Yang (2013), Listening 1		0.36	2.96	★
Yang (2013), Listening 2		0.85	3.20	N/A
Yang (2013), Listening 3		0.85	3.05	N/A
Younsi (2017), Vocabulary		2.61	2.93	★
Zenots (2012), Reading		1.17	3.25	★﹂
Overall		1.33		

Heterogeneity: $I^2 = 94.85\%$

From Figure 3.2, it is apparent that most of the reported results from metacognitive interventions are effective. Most of the effects (28/32) are larger than 0.4, which are considered substantial effects in education (Hattie, 2009). Overall, the average effect size of the included studies is 1.33 which is considered a large effect size. This seems to affirm the strong potential of metacognitive promotion for improving EFL learners' achievement. However, since a large majority of the studies obtained their results from very small sample sizes, the chances of having inflated results are highly plausible (Coe, 2002). Therefore, a firm conclusion about the effectiveness from the results cannot be made. Such results can only indicate the promise of metacognitive intervention.

To investigate the threat to the validity of those results, a heterogeneity test and an estimate of publication bias can provide further information. As discussed in the scope of the study, this review applied a number of criteria to screen the studies. For example, only studies which are controlled trials, in EFL contexts and in tertiary education would be included. These attempts were aimed to minimise the heterogeneity of the included studies. Despite such attempts, it is unlikely that heterogeneity would be completely eliminated and an investigation of heterogeneity is still essential.

Heterogeneity refers to the variation of effect size in each included study, so that the smaller the variation, the more precise the estimate of the overall effect would be (Borenstein et al., 2009). One way to examine heterogeneity is to conduct an eyeball test by looking at a forest plot to see how close the effect sizes are to each other. From the forest plot above, it is noticeable that a fairly high number of the ranges of effects are disconnected from the others. This could indicate a considerable degree of heterogeneity.

For a more thorough investigation of heterogeneity, statistical testing of heterogeneity such as I^2 could be performed (Borenstein et al., 2009). At the bottom of the forest plot in Figure 3.2, the I^2 value (94.85%) suggests that the heterogeneity of the included studies is very high, in line with what was observed from the eyeball test. This means that the effect of each study differs quite hugely. Such high heterogeneity undermines the precision and trustworthiness of the synthesised effect, since the estimate may not represent any distinct group of studies and may lead to misleading interpretation (Lipsey, 1992). This is possibly because most of the studies have very small sample sizes which can exaggerate the actual size of the effect (Coe, 2002) and increase the lack of precision. This implies a serious need for EFL

studies on the topic to work with larger sample sizes if they are to provide more reliable evidence.

Another issue to consider is that systematic reviews can occasionally fail to retrieve studies with negative or null effects, since studies with positive effects are usually preferred for publication (Torgerson, 2003). This can create a publication bias. The issue is a threat to validity in meta-analyses as it can lead to a spurious estimate of the treatment effect (Sutton, 2009). Even though this review attempted to minimise the publication bias by trying to retrieve unpublished records from Google Scholar, it was still found to be present, as seen in the funnel plot in Figure 3.3.

Figure 3.3: Funnel plot for examining publication bias
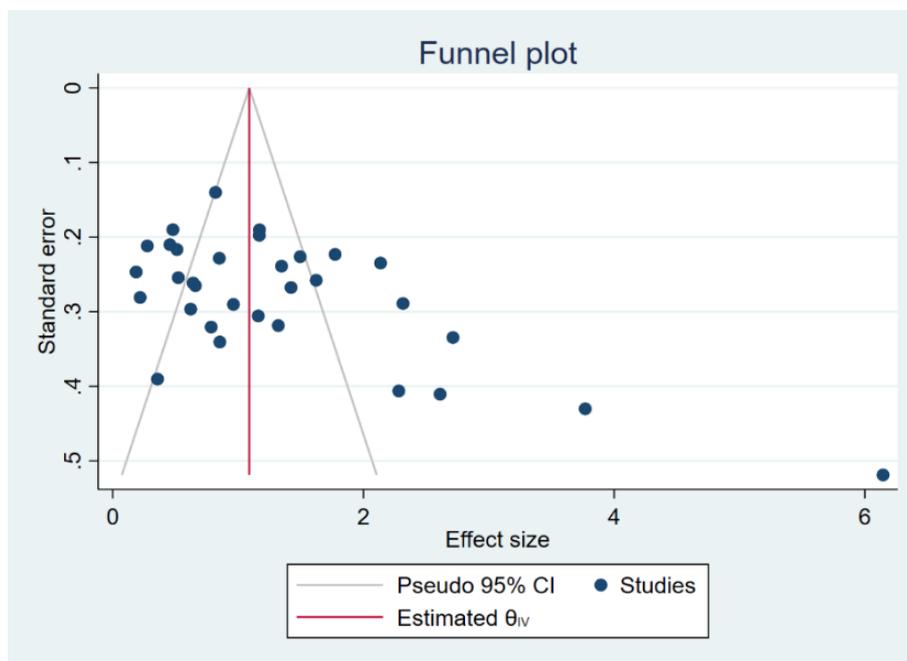
From the funnel plot, it is noticeable that the distribution of the effects does not form an inverted funnel curve and that numerous effects fall outside of the estimated funnel curve. This suggests a considerable degree of publication bias in the obtained result. This could be due to the fact that studies with negative effects were not published or the relevant studies did not comply with the eligibility criteria and were, therefore, excluded. This finding again signifies a need for studies on the topic with more rigorous designs to warrant clearer interpretation.

With all the results presented, it could be concluded that the metacognitive approach has the potential to improve the learning outcomes of EFL learners. However, the overall effect size of 1.33 cannot be a reliable estimate of the effect for several reasons. The limited quality of the included studies, the very high heterogeneity and the publication bias cause such an estimate to be far from accurate. Essentially, the quality of the included studies is the most critical issue. If design, sample size, attrition, outcome measure and validity assurance are given due attention, the other two issues are of less concern as they would consequentially be enhanced.

3.4.5 Results from thematic synthesis

The results from the thematic analysis will be presented by the evidence strength based on Gorard (2014)'s criteria with 0-4 stars, where 4 indicates the strongest evidence. Then, the results in each group will be discussed by language outcomes. A summary of the included studies based on their evidence strength and characteristics are presented in Table 3.3.

Table 3.3: Summary of the characteristics of the included studies by their evidence strength

Evidence strength	Design	Sample size	Attrition	Outcome measures	Validity assurance	Results
Group 1: 3 studies (2 stars)	Controlled trial with random allocation	40 or higher per arm	All did not discuss attrition but number of students taking pre-tests and post-tests are the same.	Pre-specified outcome and use of standardised measurement	Average attempts of validity assurance	All reported positive results
Group 2: 17 studies (1.5 stars)	Controlled trial without random allocation	Below 40 but more than 20 per arm		Pre-specified outcome and use of standardised measurement	Average attempts of validity assurance	13 reported positive results. 4 reported no effect or mixed results.
Group 3: 9 studies (1 stars)	Controlled trial without random allocation	Below 20 per arm		Pre-specified outcome but some has concerns on reliability of assessment	Some (5) has risk of researcher effect/ weak validity check	All reported positive results.

Among the 29 included studies, three studies received the highest rating of 2 stars. They are controlled trials with standardised and independent tests for measuring the outcomes and average attention to validity assurance. What makes these studies get the moderate rating (2 stars) which is the highest compared to other studies is the inclusion of random allocation and the sample sizes which are rather small but relatively larger than in most other studies. All

studies reported a positive effect of the interventions. Seifoori (2016)' study focused on the speaking skills of 114 Iranian undergraduates. The intervention was designed to include pre-task planners (PTP), on-line task planners (OLP), and pre and on-line task planners (POLP) and was run over 15 sessions. From the results with the intervention group outperforming the control group (ES 0.47), it was suggested that using both pre and on-line task planners has the highest benefits to the students. Tavakoli and Koosha (2016)'s intervention focused on reading comprehension among 100 Iranian students and was based on the CALLA model by Chamot and O'Malley (1994) which features planning, monitoring and evaluating reading. The results clearly favoured the intervention group with an effect size of 1.50. The researchers emphasised guiding students when and how to employ strategies over the knowledge of strategies. Abdelhafez (2006)'s thesis aimed at reading and listening comprehension of 80 Egyptian students. It also followed the procedure in the CALLA model and highlighted opportunities for guided and free practices. The results indicate the invention group improved significantly higher with an average effect size of 1.62.

17 studies are classified as 1.5 stars. These studies are based on controlled trials with a smaller sample size compared to the previous group. 13 of them reported positive impact of the intervention while four studies reported no effect or mixed results. Regarding language outcomes, eight studies targeted reading comprehension with six studies reporting positive effects and two no-effect studies. The key factors linking to improvement were reported. Developing metacognitive strategies helped learners to self-regulate (Msaddek, 2016; Razı & Çubukçu, 2014; Roohani & Asiabani, 2015). Strategic behaviours made reading easier (Roohani & Asiabani, 2015; Zenots, 2012). Reflection skills aided the identification of contextual clues for comprehension and assisted students in monitoring the reading processes (Teng & Reynolds, 2019; Tsai & Talley, 2014). A possible reason for a no-effect outcome was attributed to the online mode of learning, which involves fewer student interactions (Altay & Altay, 2017).

Within this group, four studies targeted listening comprehension, three of which reported effectiveness and one showed no effect. The Metacognitive Pedagogical Sequence (MPS) proposed by Vandergrift (2004) was applied in all studies. Reasons for the effectiveness were given as follows. Awareness of metacognitive strategies makes the students perceive to be more self-assured and feel less tension (Rahimirad & Shams, 2014). Conscious planning helps the students to apply strategies more appropriately (Bozorgian & Alamdari, 2017; Chou, 2017). Moreover, appropriate use of metacognitive strategies and top-down processing is

essential for successful listening (Chou, 2017). In the no-effect study, the control group's lessons which were also process-based in nature, similar to the intervention group may cause the similar improvement rate (Tanewong, 2019).

Nine studies were classified as attaining one star. These were controlled trials with even smaller sample sizes compared to the previous groups. A number of studies in this group were also at risk of having experimenter's effect with the researcher being the sole teacher of the trial classes. In addition, some studies had an issue regarding reliability of assessment, such as in the assessment of productive skills without a clear rubric or an independent examiner. Positive effects were reported in all studies and the CALLA model was applied in many of them. Similar reasons were given for students' improvement. Meanwhile, overly lengthy and repetitive instruction of metacognitive knowledge were perceived as causing boredom in some cases, leading to less focused attention (Yang, 2013). Despite the fact that positive effects were reported in all studies, limitations in the studies undermine the evidence of effectiveness of the interventions.

It was noticeable that only three studies of all 29 studied received a moderate rating of evidence strength. Even though all three studies in this group reported positive results, this cannot be taken as an indicator of effectiveness of metacognitive interventions in EFL contexts, since the number of studies is very few and they are rated as two-star, both of which factors cannot warrant trustworthy results. Moreover, the sample sizes in all three studies are small. The fact that the large majority of the included studies received a below-average rating suggests that the evidence of effectiveness of the metacognitive approach in EFL contexts is still weak. Moreover, a sole study conducted in a Thai context cannot provide strong evidence of the intervention. Therefore, it is reasonable that more rigorous studies on the topic should be conducted, including studies in Thai EFL contexts.

The results from the thematic analysis correspond to the meta-analysis as the potential of the metacognitive interventions for language development among EFL learners is observed. However, the effectiveness of metacognitive interventions remains inconclusive owing to the limited quality of the included studies, since only three studies among all 29 studies have moderate evidence strength. From the thematic analysis, useful suggestions on key features for successful metacognitive instruction were extracted. These are summarised below.

Explicit training. The issue of whether explicit or implicit instruction is more appropriate for teaching thinking skills has been a subject of debate for decades. Today, however, more and stronger evidence seems to favour explicit instruction (Wegerif et al., 2015). The findings from this review support this view, since all those interventions with positive results were based on explicit instruction. Likewise, a review which particularly focuses on instruction types in L2 settings by Norris and Ortega (2001) advocates explicit instruction. Therefore, explicit training of skills is fundamental for metacognitive interventions.

Clear instructional procedure. Most successful studies show a clear sequence for a smooth embedding of metacognitive components in the language classroom. The most common models are the Cognitive Academic Language Learning Approach (CALLA) based on Chamot and O'Malley (1994) for reading and other skills, the metacognitive pedagogical sequences (MPS) based on Vandergrift (2004) for listening and the KWL model by Ogle (1989) for reading.

Focus on regulation of skills. Despite knowledge and understanding of metacognition being important, what is more vital is the regulation of the skills related to such knowledge via planning, monitoring and reflection. Being able to remember the terms and processes is less important than the ability and opportunities to strategically apply such knowledge and skills in the attempt to regulate one's own learning.

Providing materials. The majority of studies provided materials for students not only to study from but also to work on. For concepts related to mind and thinking, providing supporting materials help prevent EFL learners from becoming overwhelmed. The materials can be in different forms, such as handouts, portfolios or diaries for note-keeping. The essence of providing materials is to encourage students' practice of metacognitive regulation.

Useful supplementary considerations. Apart from the features above, considerations of the following elements can be helpful. First, a collaborative learning environment should be promoted. This can be done by giving students opportunities to reflect on their own learning and to exchange ideas with peers in pairs or small groups. Secondly, support from teachers is crucial. The support can include demonstrations of how each step of the task works, giving feedback on students' performance and providing guidance for further development. In addition, resources for self-paced independent practice should be provided, since each

individual may require a different length of time to internalise new concepts. Finally, consideration should be given to the use of the mother tongue, depending on learners' proficiency levels and classroom norms. In this instance, teachers would need to observe and make a decision whether or not to allow the mother tongue for the optimal learning experience.

3.5 Discussion

The review has revealed that a considerable number of EFL researchers are interested in promoting metacognition or self-regulation in their classrooms. One reason may be due to the widespread endorsement of teaching thinking in higher education policy (Wegerif et al., 2015). Another reason may be the metacognitive process itself, which can be an integral part of language learning processes (Wenden, 1987), and its high potential for improving learning. In addition, the relatively low cost of implementation can make the approach worth attempting.

The overall results from the review support the potential of metacognitive intervention for EFL learners. In line with (Teng, 2020), metacognitive regulatory skills are highly associated with EFL learners' performance. Most included studies report positive results from their controlled trials. However, such results can merely show the promise of the interventions but cannot securely affirm their effectiveness due to limitations in most studies in terms of design, scale and validity precautions. For more trustworthiness and credibility of the results, more robust investigations are required. To elaborate, studies need to be based on robust designs for causal inference such as controlled trials with random allocation, regression discontinuity design or matched comparison (EEF, 2019). Larger sample sizes are also of paramount importance. Over a hundred participants per arm would increase the precision of the effect size estimate, leading to more reliable conclusions (Gorard et al., 2017). Considerations of the validity threats such as the deliverer of the intervention and examiner of student performance, the fidelity of intervention delivery and standardised outcome measures would increase the trustworthiness of the findings.

Both meta-analysis and thematic analysis were conducted to provide review results. Results from technical methods in the meta-analysis, such as heterogeneity analysis and publication bias, and results from the narrative criterion-based analysis similarly demonstrate that the results in most of the included studies, as well as the estimate of the overall effect of the included studies, were not secure and, therefore, not reliable. Even though the technical methods were included to show the obtained results from the meta-analysis were not secure, what matters most is the quality and evidence strength of the included studies determining how

trustworthy such results are. Elements such as research design, sample size, attrition, outcome measures and validity assurance are crucial factors for determining the trustworthiness of research results (EEF, 2019; Gorard, 2014).

For thinking-based approaches, such as metacognition and critical thinking, one argument often raised is whether these are appropriate for contexts outside Western cultures where questioning superiors seems uncommon (Ramanathan & Kaplan, 1996). Results from this review suggest that metacognitive instruction is feasible in tertiary EFL contexts. Another review by El Soufia and See (2019) focusing on critical thinking in EFL context similarly advocates thinking-based instruction in an explicit manner. From many included studies that looked at perception, awareness or motivation as other outcome variables, most found positive impressions regarding introduced thinking-based interventions (e.g., Habibian, 2015; Farahian & Avarzamani, 2018).

Another major finding is the obsession that most EFL researchers have with significance testing. In fact, most experimental studies in education, not only in EFL research, have relied mainly on significance testing to show the effect of their studies (Cohen, 1994). Coe (2002) argues that the main issue with significance testing is that the p-value depends essentially on the size of the sample and the size of the effect, so there is a chance of obtaining a significant result if the sample is huge despite the actual effect being tiny, or if the effect is great while the sample may be very small. Thus, EFL researchers who are concerned about evidence-based investigations should be more sceptical of the p-value (Plonsky, 2014). Indeed, the effect size derived from the standardised mean differences of the outcomes or other alternative measures should be reported instead or, at the very least, should be included (Higgins, 2018). The report of effect sizes has been encouraged by the American Psychological Association since 1994 (Wilkinson et al., 1999). By reporting effect sizes as well as giving careful consideration to the quality of the study, the reported result is more meaningful and can, in turn, facilitate and enhance the synthesis of evidence.

As part of the present thesis, a primary empirical trial will also be conducted in a Thai EFL context to assess the effectiveness of a metacognitive intervention. In this systematic review, only one study from Thai EFL contexts satisfied the inclusion criteria and this single study also had a small sample size. This underscores the lack of credible evidence on this topic and deficiency in quality of evidence-based research in Thai EFL contexts. Therefore, more studies with higher attention to research quality and evidence strength are required.

The results from this synthesis have provided useful information for the main trial of this thesis. First, the evidence of the effectiveness, which remains inconclusive, indicates a need for more studies with more cautious attention to key research elements. Secondly, the reported high potential provides a justification for the application of the metacognitive approach with the trial participants. Thirdly, suggestions provided from the thematic analysis can serve as guidelines and precautions for the implementation of the main trial.

3.6 Limitations

A few major and minor defects were acknowledged as being limitations of this review. The first and major issue is that of eligibility criteria. The review employed a priori criteria for records screening. However, the criteria were not recorded anywhere, such as in a review protocol which would have enhanced the validity of the method. Secondly, the search strings used in the identification process were not sufficiently varied to allow higher chances of identifying more studies. Indeed, if terms such as trial, evaluat* or experiment* had been included in the search, more studies might have been identified.

In addition, despite the PICO criteria, evidence of the impact was greatly undermined by the limited quality of the included studies. Thus, stricter review criteria should be considered in future to filter studies with more credible evidence. Moreover, the search for relevant studies from seven data sources, despite seeming to be exhaustive compared to other reviews on the topic, may not have included all of the relevant studies and other studies may have been missed. More comprehensive data sources should be attempted in future to allow more robust studies to be included. A hand search of reference lists in the relevant systematic reviews or studies would be another option to increase the chances of including more relevant robust studies. However, this strategy may also increase the risk of bias.

3.7 Chapter summary

This chapter presents and discusses the results from a systematic review, which exhaustively identified, screened, examined and synthesised the relevant studies on metacognitive interventions in tertiary EFL contexts. The synthesis found promising potential in the approach to improving EFL learners' language performance. However, limitations in the included studies, high heterogeneity and risk of publication bias inhibit a firm conclusion being reached. Despite a considerable quantity of studies on the topic, EFL research seeking to provide robust evidence should show greater caution in the designs, scale, attrition and validity assurance to warrant the reported results. Indeed, impact evaluation studies in EFL contexts on

any topic should pay greater attention to these elements in order to provide more valid and valuable information to EFL literature and thus make tangible impacts.

The synthesis has answered the research question one and provided useful information for the design of the intervention for the main trial which will be discussed later in Chapter Four. Suggested features for successful interventions, namely, explicit instruction, clear pedagogical sequence, materials provision for learners and focus on the regulation of strategies were informative for the intervention design. Moreover, the principles of metacognitive instruction discussed in Chapter Two and the lessons learnt from the pilot study to be presented in Chapter Four were considered in the design and implementation of the intervention.

CHAPTER FOUR

DESIGN AND METHODS OF EMPIRICAL STUDY

This chapter devotes seven sections to explicate the empirical investigations and methodology for the trial of metacognitive instruction in the target EFL context. It begins with methodological backgrounds discussing key research elements, research design, details of the intervention and the instruments for data collection. Section 4.2 elaborates on the ethical considerations upheld throughout the study. The next section describes the pilot study and reflections on the overall preliminary phase. Section 4.4 presents the main study by discussing the design and all the key elements in the trial, namely, participants, randomised allocation, implementation of the intervention and data collection. A discussion of secondary data and how it complements the main study ensues, followed by procedures for data analysis in section 4.6. Finally, the chapter is briefly summarised in section 4.7.

4.1 Backgrounds to research methodology

This section provides a background understanding of the fundamental elements of research. This will provide a basis for the pilot study, which is an efficacy trial, and the main study, which is an effective trial, of the metacognitive intervention for EFL learners. The information to be presented in this part is useful not only for these trials, but also for any studies concerned with using or providing robust evidence to inform policy and practices in education and second language teaching.

4.1.1 Important research elements for studies on effectiveness

The design and required elements of a study are essentially determined by the aims and the research questions (de Vaus, 2001). The design for descriptive questions which aim to describe a phenomenon such as *what is going on* and *what exists* would be different from explanatory questions which seek to explain if a factor or factors affect the other (White, 2009). With such dissimilar aims, the required elements for studies with explanatory questions are certainly different from the necessary elements in studies with descriptive questions.

Gorard (2013) provides an insightful explanation about the key elements of research, namely, observation, time and sequence, comparator, intervention and allocation. Obviously, the first two elements, i.e. observation and time and sequence are fundamental in any study. Observations, in this case, refer to an attempt of data collection ranging from a real observation at a site, any kind of interviews, use of questionnaires, a standardised test or other kinds of

measurement that can elicit relevant data for the research questions. The number of times data collection takes place would make a study cross-sectional with a single snapshot of data collection, or longitudinal with a series of data collections over time (de Vaus, 2001). Most studies addressing descriptive questions are based on these two elements. If a study seeks to make a comparison between variables or factors in relation to another variable, a comparator would be required. For instance, in a study to investigate if a particular factor such as an intervention or a programme has an effect on another variable, a comparator is indispensable, as it can provide counterfactual evidence from participants who have not received the intervention (Shadish et al., 2002). For causal research questions asking if variable A affects or leads to changes on variable B, randomised allocation is an additional crucial element for enhancing comparability and minimising threats to internal validity (Hedges, 2012).

For the main empirical study of this thesis, at least two main research questions (see section 4.4) interrogate if a metacognitive intervention is the X factor affecting the learning and other outcomes of participating EFL learners. Therefore, all or at least four of the aforementioned research elements would be required if the study is to obtain valid evidence. Hence, a cluster Randomised Controlled Trial (RCT) was chosen for this study. Further details of the research design will be discussed next.

4.1.2 Research design

Research design provides a structure that allows an appropriate type of evidence to be collected in order to answer the research question. It is the point at which research questions are translated into research projects (Hakim, 2000). It is an essential aspect of research, as White (2009) proposes, "it marks the stage in the research process when the researcher moves from thinking about asking questions to thinking in detail about how these questions might be answered" (p.98). Developing a research design essentially concerns decision-making about the kinds of evidence required to address the research question and the appropriate methods to secure such evidence (de Vaus, 2001). Therefore, the research design must be well aligned with the enquiry being investigated (Hedges, 2012).

To investigate the impact of the metacognition-based intervention on the learners' achievement of English and their metacognitive awareness, a cluster RCT was employed as the study's design. The study addresses some causal questions. Thus, a research design with a comparator and random allocation is optimal to allow causal inference between the outcome and the intervention. The RCT is one of the most robust research designs for establishing a

causal link between an intervention and any effect, signifying the impact of the intervention (Gorard, 2013; Shadish et al., 2002). RCTs are, therefore, sometimes referred to as the 'gold standard' design for evidence-based educational practice (Hutchison & Styles, 2010).

Several factors make the RCT an optimal design for causal research questions. RCTs can secure a causal inference between an intervention and its outcomes because they consist of the necessary elements for establishing a causal inference. Shadish et al. (2002) state that a causal relationship holds if

"1) the cause precedes the effect, 2) the cause was related to the effect, 3) we can find no plausible alternative explanation for the effect" (p.6).

With these conditions in mind, the RCTs have all of the five key research elements which can provide the most relevant evidence for making a causal link between the intervention and its outcomes (Gorard, 2013). The two fundamental features of RCTs, namely, comparator and random allocation, make this kind of trial superior to other quasi-experiment designs for impact evaluation. The comparison group can provide counterfactual information from participants who have not received the intervention, while random allocation can ensure the likelihood that groups are similar on both known and unknown features and are, therefore, comparable (Shadish et al., 2002). Random allocation also makes RCTs the most effective for eliminating selection bias (Torgerson & Torgerson, 2003). With all the five elements, RCTs, if administered properly, can minimise threats to internal validity such as temporal change, maturation (Hedges, 2012) and regression to the mean (Torgerson & Torgerson, 2008), which may be alternative explanations for the outcomes.

When discussing randomisation, one confusion which usually occurs is the misunderstanding between random allocation and random sampling (Torgerson & Torgerson, 2008). Random sampling is a way of probability sampling to allow chances for every member of the population to be included in the sample through random selection (Mertens, 2010). Random sampling enhances external validity and is helpful if researchers want to make generalisations of the results for wider samples of the same population (Cohen et al., 2018). Meanwhile, random allocation is a way of assigning the participants of a trial or an experimental study to either an intervention or a control group to ensure they are more likely to be equivalent in key variables (Hutchison & Styles, 2010). Random allocation is considered an essential feature of RCT because it helpful for controlling both known and unknown variables that can affect the outcomes, selection bias, regression to the mean and temporal

changes (Torgerson & Torgerson, 2008). In this study, random allocation was included to enhance internal validity.

Randomised controlled trials can be of two main types: efficacy trials and effectiveness trials (Torgerson et al., 2015). While both types of trials seek to assess whether an intervention has an effect on the target outcome, they differ primarily in the conditions under which they are implemented. *Efficacy* trials are studies conducted under optimum conditions, such as trials carried out in laboratories or trials implemented by the researcher who designed the intervention. They are usually implemented before the intervention can be delivered to larger samples in an authentic setting. In comparison, *effectiveness* trials implement and assess the effectiveness of an intervention in authentic settings where fidelity may be less optimal, such as in the actual classroom with real teachers implementing the intervention on real target participants.

As a cluster-randomised trial, the randomisation in this study occurred at a group level rather than on an individual basis. Cluster randomisation seems to be common and appropriate in educational settings, since group-level allocation can minimise contamination caused by a possible spill-over between the intervention and control conditions (Hutchison & Styles, 2010). However, the cluster RCT may risk having a clustering effect on the outcomes because the results drawn from learners in the same cluster could be influenced by the dependence among cluster members (Campbell et al., 2000). Thus, cluster RCTs generally require a larger sample size if the intra-cluster correlation coefficient (ICC), an estimate showing the degree to which individuals are dependent, is higher than zero (Torgerson & Torgerson, 2008). In addition, a cluster RCT can be prone to recruitment bias if the cluster members are recruited after randomisation. Therefore, participants in this study were identified before the randomisation of clusters.

As interventions in educational research are usually applied as a classroom practice, a process evaluation should also be conducted. The process evaluation can enhance the contextual understanding of the outcomes and the participants' fidelity to the intervention guidelines (Siddiqui et al., 2018). For this reason, a process evaluation was another important element of this trial. A protocol was designed for classroom observation in each of the teacher's classes. After implementation of the intervention, open-ended questionnaires and semi-structured interviews with teachers and a number of voluntary intervention students were conducted.

4.1.3 The intervention

The intervention used in the study was the PMER model discussed in section 2.2.3 in Chapter Two. The design of the model was based on the metacognition and self-regulation principles from cognitive development perspectives in Flavell (1979), Pintrich (2002) and Zimmerman (2002) and language learning perspectives in Vandergrift & Goh (2012) and Wenden (1998). These studies were consulted because they align with and complement each other well, as discussed in Chapter Two. Moreover, these studies have been consulted by many EFL studies concerning learners' active roles in their own learning. However, the PMER intervention did not adopt any model in those studies verbatim, but rather adapted the strengths in each of them to suit the target setting. Moreover, suggestions for metacognitive interventions drawn from a systematic review proposed in Chapter Three were taken into consideration. In addition, lessons learnt from the pilot study (to be presented in section 4.3) were deliberately considered in order to make the model more relevant to participants. The deliberate amalgamation of essential features and principles of metacognition and self-regulation together with lessons learnt from empirical studies make the PMER intervention a thorough and appropriate tool for the EFL learners in the southern Thai contexts.

The intervention consists of three phases, namely, *orientation, regulation* and *facilitation*. In the orientation phase, the concept of metacognition was introduced to the students. Both general concepts about metacognition and specific applications to English listening were presented. To raise the awareness of their metacognitive knowledge, students were presented with useful strategies for improving listening, i.e. listening strategies, cognitive strategies and affective strategies, before the concept of metacognition. This was to warm students up with narratives that they were more familiar with in order not to overwhelm their cognitive load. This could also improve their strategic repertoire before introducing a new or more complex concept. Indeed, cognitive, affective and strategic competences are needed in exercising metacognition (Quigley et al., 2018).

As illustrated in Figure 4.1, knowledge of various strategies is one of the three elements of metacognitive knowledge. For holistic development, strategic knowledge does not involve only listening strategies, but also cognitive and affective strategies, in line with O'Malley and Chamot (1990) and Vandergrift and Goh (2012). The other two elements of metacognitive knowledge, i.e., person and task knowledge (Flavell, 1979; Pintrich, 2002) were constantly

raised each time learners were involved in a learning task. Subsequently, the three-phase metacognitive strategies, namely, planning, monitoring and evaluating, were introduced.

Figure 4.1: Orientation to metacognitive awareness
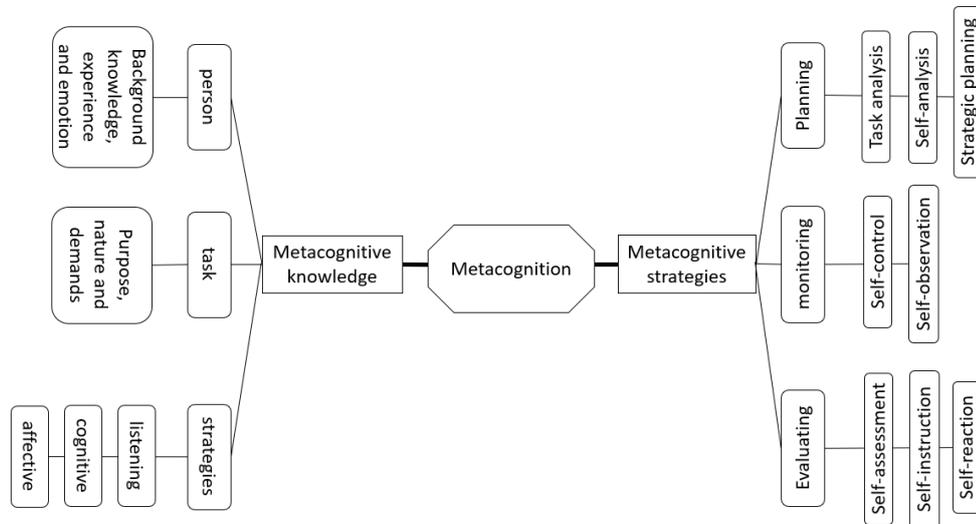
As one of the successful characteristics of metacognitive programmes suggested in Chapter Three, metacognitive regulation should be emphasised over knowledge about metacognition. This was the second core of the PMER intervention. After the orientation phase, the practice of metacognitive regulation was implemented through the metacognitive pedagogical sequence adapted from Vandergrift and Goh (2012) presented in Figure 4.2.

Figure 4.2: Pedagogical sequence for metacognitive regulation in listening

1. Pre Listening	2. First Listening	3. Monitoring	4. Second listening	5. Post Listening	6. Retreival Practice
•Purpose: Contextualising and activating schema via photo, videos, pair/ groups works •Activity: Introducing task, recalling schema, prediction, goal-setting	•Purpose: Listening for gist and details and attempting to apply strategies •Activity: Focused listening and applying strategies, monitoring performance	•Purpose: Reflecting on understanding and strategic performance, modifying attention and strategies in next attempt •Activity: Pair/ group discussion on their understanding, performance, experienced difficulties, adjustment for next attempt	•Purpose: Fulfilling understanding gap, modifying strategies •Activity: Practicing selective attention, modifying strategies	•Purpose: Reflection on performance, what has been learnt and points for development •Activity: Whole class discussion on the listening content, evaluation of performance and make notes of points for improvement	•Purpose: Recall activities to promote learning and retention •Activity: Doing a recall task such as a quiz to pull out students' knowledge and discussing answers together

Each metacognitive listening session allowed two listening attempts for students to practise, apply and modify skills in planning in the pre-listening, monitoring during listening and evaluating in the post-listening stages. In Vandergrift and Goh (2012)'s model, three listening attempts are allowed so that students have three verification stages. In the present trial, the listening attempts in each session were kept to two, because the researcher believes that dual attempts are more compatible with students' real-life practices and future standardised tests. In real-life situations, asking for repetition or clarification can be a practical strategy to gain comprehension, but too many repetitions can be interruptive. For some students who will take a standardised test to meet graduation criteria or other purposes, most tests only allow one listening attempt or twice as a maximum. Another major distinction between the PMER model and that of Vandergrift (2004) and Vandergrift and Goh (2012) was the addition of retrieval practices as a final stage of the pedagogical sequence. Retrieval practices could exercise students' brains and strengthen memory (Agarwal et al., 2017) and lead to a more accurate metacognitive knowledge of their own understanding (Karpicke, 2009). The six-stage sequence portrays the procedures through which the metacognitive regulation was exercised.

Apart from awareness and opportunities to apply metacognitive regulation, support and facilitation are crucial for the learning experience. From one of the suggestions mentioned in the systematic review in Chapter Three, providing materials can be helpful for students not only to study from, but also to work on. The pilot study teacher also suggested having practice materials outside the classroom. To support and encourage metacognitive regulation, students were provided with learning resources, which included a practice book for metacognitive sessions and online Independent Practices for each metacognitive session. The practice book encompassed background knowledge of listening strategies, metacognitive knowledge and metacognitive strategies and practice tasks to encourage strategic and reflective thinking.

The metacognitive tasks asked students to reflect on their metacognitive knowledge before starting to listen, after the first listening and after both listening attempts. The tasks were primarily composed of self-questioning prompts to urge students to strategically and reflectively think about how to approach the listening tasks, difficulties they experienced, accomplishments they made and adaptations for problem-solving and development. Reflective tasks are the basis for the intervention, since reflective and strategic thinking are the most salient features of metacognition (Moseley et al., 2005). Reflection is an essential component of self-regulation (Zimmerman, 2002). As maintained by the Cambridge Assessment International Education (n.d.), "reflection is a fundamental part of the plan-monitor-evaluate

process. Encouraging learners to self-question throughout the process will support this reflection". Importantly, self-questioning is found to be a helpful technique employed by good language learners (Anderson 2008).

For each of the practice tasks, a worked example was provided to help students understand the task more clearly. This would also decrease their cognitive load and learn more successfully (Sweller, 2006). Clear and relevant examples guide the learners in making a connection between the learnt strategies and the task (Ashman, 2015). Examples of the metacognitive tasks and worked examples can be found in Appendix 3.

According to the EEF guidelines by Quigley et al. (2018), teacher support plays a vital role in a classroom that is concerned about promoting metacognition and self-regulation. Since the teachers who delivered the intervention in this trial were also new to the concept, they were provided with tools and supports to ensure they could appropriately train and facilitate the students. Before starting the trial, the intervention teachers were given a two-hour training on the backgrounds of the intervention and how to use it. A teacher's kit including a teachers' manual for metacognitive instruction, lesson plans, class materials and multimedia tools for each metacognitive session was prepared for them (see samples in Appendices 4 and 5). Such supports were supplied in the hope of helping the teachers to deliver the intervention smoothly and to increase fidelity to the intervention.

4.1.4 Outcome measures and instruments

Apart from research design and scale of the study, outcome measure is another critical element in securing trustworthy findings (Gorard et al., 2017). To obtain evidence to answer all the research questions, both impact and process evaluations were conducted. The tools used for these purposes are as follows.

4.1.4.1 Impact evaluation

To evaluate the impact of the intervention on students' learning outcomes and metacognitive awareness in English listening, the following instruments were employed.

- **English proficiency test**

To assess students' learning outcomes, test items from the Tell Me More (TMM) software were adapted as a pre-test and a post-test for all trial participants. The software offers a standardised testing system and a training programme for English and other languages. The

software was operated by one of the participating universities to assess its students' English proficiency. The researcher obtained the sample test items with permission from authorised staff. Apart from the pre/post-tests, the results from the actual TMM Placement and Progress Tests were used as baseline data and delayed outcome measures respectively for all trial participants who were entitled to take the tests.

The pre-test and post-test consisted of 35 items and were divided into two sections, i.e., Listening and Language Use. The listening section contains 20 items covering word-level completion, picture description and comprehension of short monologues and conversations. The Language Use contains 15 items assessing vocabulary, grammar and sentence-level reading comprehension.

The full TMM test comprises 64 items with similar features to the pre/post-tests, such as grammar, vocabulary, listening and reading competence. However, there was no obvious division of the items into skill-based sections. The software reports test results to test-takers in band scores ranging from 0-10. As regulated in the university's qualification criteria, students need to gain a TMM score of five or more or an equivalent of an alternative proficiency test to satisfy the criteria.

- **Metacognitive awareness assessment**

The most common tool for assessing metacognitive awareness is self-report questionnaires (Gascoine, Higgins & Wall, 2017). In this study, a questionnaire adapted from the Metacognitive Awareness Listening Questionnaire (MALQ) by Vandergrift et al. (2006) was used for assessing learners' metacognitive awareness in listening (see Appendix 1). The questionnaire comprises 21 items and asks respondents to express how much they agree with each statement using a 5-point opinion scale. The questionnaire items are divided into five categories which reflect learners' metacognitive awareness in their listening. The five categories are: *person knowledge, planning & evaluation, directed attention, problem-solving and mental translation*. The questions ask learners to reflect *on actions* of positive and negative behaviours in their listening.

Each category has a different number of questions. Person knowledge has three items, planning and evaluation has five items, directed attention has four items, problem-solving has six items and mental translation has three items. According to the authors of the MALQ, all mental translation items, two person-knowledge items and one directed attention item are indicators of negative behaviours or attitudes that demonstrate low metacognitive awareness in

listening. The remaining questionnaire items reflect respondents' desirable metacognitive awareness.

4.1.4.2 Process evaluation

Process evaluation is another meaningful element of an RCT in education, as it enables an exploration of the implementation fidelity and explains the mechanism and contextual understanding of the results (Siddiqui et al., 2018). Furthermore, it can investigate any difficulty encountered in the implementation itself and explore participants' perceptions about the intervention. The methods for process evaluation were as follows.

- **Class observations**

An observation protocol (see Appendix 6) was designed and used for at least one classroom observation of each teacher in both intervention and control groups. Classroom observations can capture pedagogical practices, use of materials, language use, classroom interactions and students' reactions in the classroom (Spada, 2019). The main aim of classroom observations was to investigate the fidelity of the manner in which teachers delivered the intervention, observe the teaching styles and the content and material use of each teacher, explore how the students participated in the metacognitive tasks and make notes of the possible barriers or difficulties encountered by both teachers and students.

- **Semi-structured interviews**

Interviews were conducted with both teachers and volunteering students in the intervention group to discover their perceptions and experiences with the metacognitive instruction. The interview is one of the common ways of obtaining qualitative and quantifiable data from different stakeholders for further clarification (Tashakkori & Teddlie, 1998). The interview can also illuminate any difficulties or barriers in delivering the intervention and illustrate practical links between the intervention and its impact (HM Treasury, 2011). For the present study, interviews were conducted with all the teachers and some volunteering intervention students. The aim was to explore participants' perceptions about the intervention, the difficulties they may experience during its implementation, the benefits of the intervention and whether the metacognitive awareness may be helpful in their attempts to satisfy the university's graduation criteria or for general English development.

4.1.5 Overall procedures of the study: multi-method design

One of the main strengths of the study is the multi-method design to derive well-rounded findings. As seen in Figure 4.3, a systematic review on the topic was conducted to explore the existing evidence. This has already been presented in Chapter Three. Before the main trial, a pilot study was administered. This is presented in section 4.3 below. The main study consisted of a cluster randomised controlled trial (RCT) with the inclusion of non-randomised comparison groups (see section 4.4). Data were collected both before and after the main trial through various methods, namely, pre-tests and post-tests, questionnaires, classroom observations, interviews and secondary data for both impact and process evaluations.

Figure 4.3: Overall procedures of the study
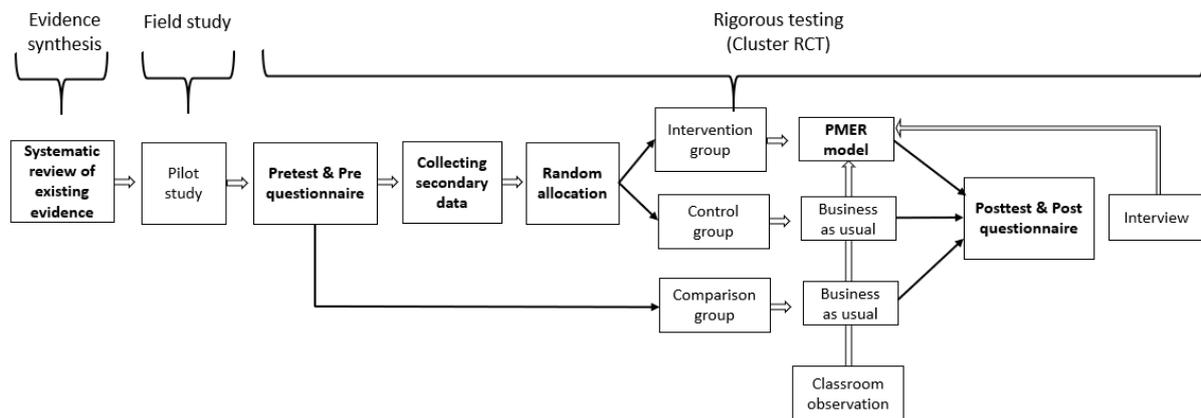
4.2 Ethical considerations

Ethical considerations were maintained to follow the ethical guidelines by the British Educational Research Association (BERA) (2018), Durham University School of Education's Ethics Policy (2018) and Thailand's National Policy Guidelines for Human Research (NRCT, 2015).

To maintain transparency, a letter of approval request for research was sent to the participating institutions to inform them of the general purposes and procedures of the trial. An information sheet with the general purposes and procedure of the research and an opt-out consent form were given to participants before the commencement of the trial in 2020 (see Appendices 7 and 8). Participants were informed of their right to withdraw from the study at any time. Moreover, the students were made aware that participation in or withdrawal from the trial would not have any effect on their course grades.

To protect participants' confidentiality, the names of participating universities, students and teachers have been kept anonymous throughout the thesis and will continue to be so in any publications drawn from the study. Since the research setting was in Thailand, the research guidelines of the local authority were consulted to assure compliance with local regulations. As the researcher has worked at the research site for several years, issues and misunderstandings related to cultural differences were minimal. In addition, students were informed of their right to give or refuse to give any information in the questionnaire and other data collection instruments.

4.3 Preliminary trialling: Pilot study

Before conducting the main trial, a pilot study was implemented with a group of tertiary EFL learners in the south of Thailand. A pilot study conducted as a feasibility analysis is essential for a rigorous trial, as it can estimate the likelihood of success (Gorard, 2013) and evaluate the efficacy of the intervention being developed (Torgerson & Torgerson, 2008). For the current study, whose aim is to implement and assess the effectiveness of a metacognition-based intervention, the pilot study would reveal the efficacy of the intervention and provide practical information for revision and improvement of the intervention before the actual trial. In addition, it would assess the comprehensibility of the metacognitive awareness questionnaire, which is a measurement tool for the secondary outcome. The pilot study was conducted in the second semester of the academic year 2019 at one of the target universities in the southernmost area of Thailand. The research questions for the pilot study were:

1. How feasible is the intervention for improving the learning outcomes of the pilot participants?
2. What may be the challenges for implementing the intervention and how to address them?

To reach the aims and answer the above questions, a pre-test/post-test controlled trial was conducted. The design was chosen because it included a comparison group to provide counterfactual evidence and was the most practical at the time. Classroom observations were administered to obtain information for the second research question. Moreover, the questionnaire included open-ended questions that could elicit further explanations for both research questions.

4.3.1 Participants

The pilot was conducted at the university, which would also join the main trial, since participants in the pilot intervention may share certain characteristics with those taking part in the actual trial. Two groups of education students in cohort 2018 (Year 2 students) were recruited for the study. The two groups were selected by the lead of the English programme at the participating university from amongst students registering for a general English course during that semester. The intervention participants included 26 Health Education students, while 19 Physical Education students joined as a control group, making 45 participants in total. Two teachers took part, one being in the intervention group and the other in the business-as-usual control group.

4.3.2 Intervention and data collection tools

The pilot intervention was not a full PMER model which was completed later as a product of an exhaustive review of the literature and the empirical lessons learnt from this pilot study. The pilot intervention was a metacognitive instruction featuring metacognitive strategies (MS) and reflective tasks to encourage students to regulate their English listening before, during and after the task. The intervention was scheduled to be implemented in five sessions throughout the remaining time of the second semester of the academic year 2019. As part of the intervention, the teacher who delivered it was provided with a teacher's guidebook and class materials, while students were given practice books which included both background information of the concept and practice tasks (see Photos 4.1 and 4.2).

Photo 4.1: Students' practice books

Photo 4.2: Teachers' guidebook
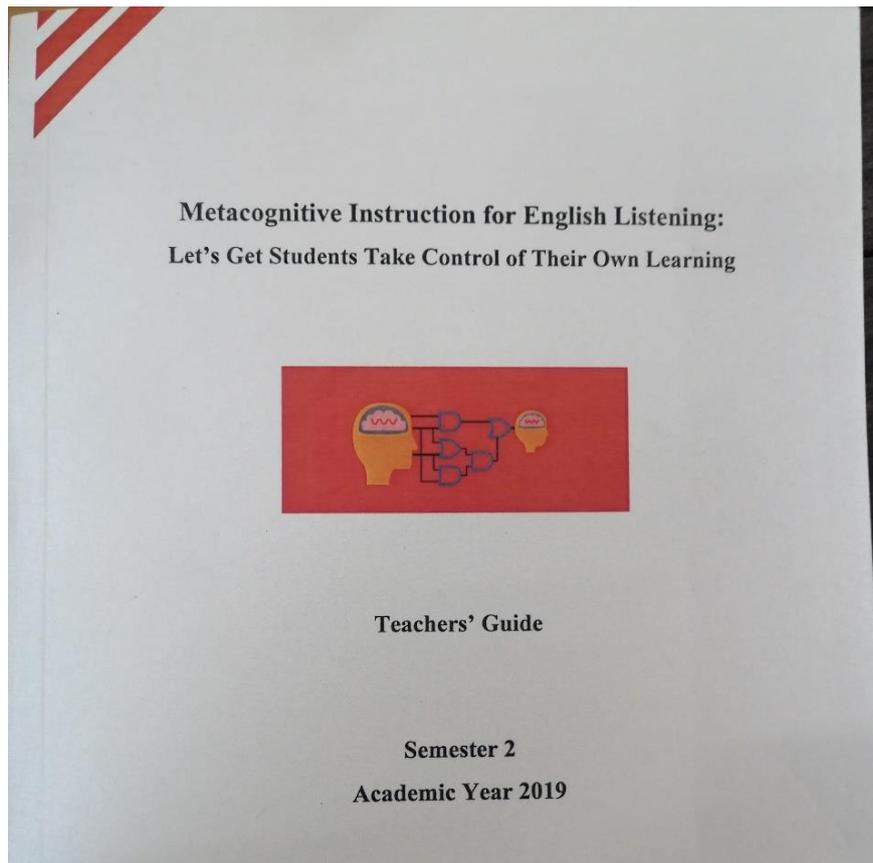
Among the five metacognitive sessions, the first session introduced strategies for effective listening comprehension, including listening strategies, cognitive strategies, affective strategies before the general discussion of metacognition was presented and linked to specific metacognitive strategies for listening. The second, third and fourth sessions focused on each metacognitive process, namely, planning, monitoring and evaluating, respectively. The final session brought together the content of all five sessions as a summary.

Two main instruments, i.e., English tests and questionnaires, were also piloted. English tests were used as a pre-test and a post-test to assess the primary outcome, which is the listening and English achievement of the students. Test items were adopted from the standardised test operated by the participating university. The test contains 35 items, with 20 items on listening comprehension and 15 items for language usage. The test could be marked objectively, and this would circumvent bias in marking.

The second instrument was a metacognitive awareness questionnaire adapted from the MALQ by Vandergrift et al. (2006). The questionnaire was used to assess the secondary outcome, which is the learners' metacognitive awareness in listening. As participants in the

main trial would have a mixed level of English proficiency including a low level of proficiency, this pilot study was a good opportunity to check whether the questionnaire items were comprehensible to the participating EFL students. Moreover, open-ended items were included in the post-questionnaire for intervention students to explore their perceptions and feedback about the intervention, which would be informative for its revision for the main trial.

Another method of data collection was class observation. Observations were planned to occur twice in the first and fourth sessions in both groups. However, observations actually occurred twice during the first and second sessions in the intervention group and only once in the first session of the control group due to the COVID-19 outbreak, which brought a halt to on-site teaching and learning after session 4. Instead of observation, phone conversations were made with the intervention teacher after session 4 to ask about adaptations made and challenges found.

4.3.3 Implementation of the pilot study

The metacognitive intervention was implemented in five sessions. The administration of the pilot study is summarised in Figure 4.4.

Figure 4.4: Implementation of the pilot intervention
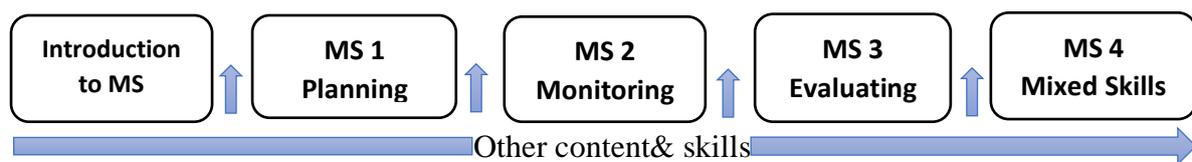
Prior to the first metacognitive session, a meeting with participating teachers in the intervention group and control group was arranged on different dates. The meeting with the control teacher was brief and contained information about what he and his students would be asked to take part, namely, to take a pre-test and pre-questionnaire, a post-test and post-questionnaire and to be given a classroom visit. The meeting with the intervention teacher lasted for two hours, as it involved a detailed discussion about the intervention she would be using with her students. A PowerPoint presentation and sample materials were used to give the teacher a clear idea of the intervention.

A pre-test and pre-questionnaire were administered to both groups in the second week of February 2020. Both instruments were administered by the class teachers and the researcher was present in the classes as an assistant. The test lasted for 30 minutes and was followed by a

questionnaire which took approximately 20 minutes to complete. The questionnaire in the pilot study was in English only. While completing the questionnaire, several students asked for a Thai translation of many questionnaire items. Thus, the researcher decided to permit the teacher to translate certain sentences so that students could gain a better understanding of the questions, which would lead to them giving a more valid answer. After the end of the session, the researcher collected the tests for marking and the questionnaires to tally answers. Pre-test results and an answer key were given to all students in both groups a week later.

Each metacognitive session was set to last for 75 minutes so that there was time left for the teacher to work on her usual class content. The first session, which was an introductory session, occurred in the fourth week of February. The researcher planned to join the first metacognitive session as an observer, but was requested by the teacher to join as a co-teacher since she wanted to build up more confidence and ensure she was conducting the session appropriately. The researcher, therefore, joined as an assistant in order to support the teacher and to observe the class as planned. The session finished within the set timeframe and students seemed to find the presentation interesting, possibly because the concept was novel to them.

The second session focusing on the first metacognitive strategy, i.e., planning, took place the following week. On this occasion, the teacher was fully responsible for delivering the entire session using the materials provided. An introductory video was presented to the students, followed by the teacher's presentation of the strategy. This took approximately 40 minutes, which was longer than planned, as the teacher frequently rechecked students' understanding to make sure they could follow. Next, the listening task was introduced, and a chart for planning listening was given to students so that they could prepare themselves before the task. At this stage, they were allowed to discuss with peers. The students, then, listened for the first round and were given a short pause to reflect on their performance and prepare themselves for the second round. They, then, proceeded to follow each step until the end.

After the planning session, a number of concerns were raised and discussed by the class teacher and the researcher. One was the time-consuming instruction, which may have been caused by excessive repetition during the presentation, and how it could be reduced so that the session would not be too exhausting for students. As a solution for following sessions, a comprehension-check task for the metacognitive strategy was designed for students to complete before introducing a listening task. Another concern was that students seemed reluctant to provide answers when completing the metacognitive planning chart, which also

resulted in taking up time. The reason for this could have been because the students were not used to working with a task that asked them to reflect on their own thinking. To improve this in following sessions, students would be asked to discuss in small groups first before reflecting on their own and the teacher would explain the worked example available in the student book.

The third session which involved the second metacognitive strategy, monitoring, took place two weeks later in mid-March. In this session, the teacher was on her own with students and followed the session procedures with adaptations learnt from the second session. After the session, the teacher revealed that she thought she performed better on time management and group discussion seemed to ease students' anxiety. It could also be speculated that learning the metacognitive strategy for a second time helped students feel less nervous and become more capable of accomplishing the metacognitive task.

The fourth session on evaluating listening was scheduled to take place in the last week of March. However, this coincided with the outbreak of the COVID-19 disease worldwide and the situation in Thailand started to become increasingly serious. As a result, the government announced the closure of educational institutions and the pilot university announced a short break for a fortnight before teaching was resumed in early April and all face-to-face teaching shifted to online modes. As a consequence, the researcher had to create an online platform via Google Classroom so that the instruction of the fourth session could be conducted virtually and remaining materials distributed online.

Despite the unexpected interruption to classroom teaching, the pilot university managed to support teaching staff by providing online accounts for a number of virtual classroom platforms such as Google Classroom, MS Team and Zoom for teachers and offering teacher training on the effective operation of those platforms. Thus, teachers and students could continue the pilot and resume the fourth session in mid-April. The PowerPoint presentations for this session were posted on the online classroom with audio explanations. Students were requested to study the materials online and complete the listening task, metacognitive task and practice test online. The teacher monitored students' progress and were available to give students support for any difficulties they might experience.

The fifth session was also conducted online. It comprised a revision and application of all strategies the students had learnt. As an adaptation to the current COVID challenge, repetition of the learnt strategies was kept to a minimum and the session mainly entailed the independent practices of all the strategies.

The post-test and post-questionnaire were conducted online in mid-April. Students in both groups were given 30 minutes to complete the post-test online under teacher supervision. Meanwhile, completion of the questionnaire was self-paced with a set dateline of 18th April. All intervention students turned up for the post-test but only 11 students in the control group submitted their post-test answers.

4.3.4 Pilot study results

The primary outcome of the pilot study was the students' achievement in listening comprehension and overall English competence. The primary outcomes are presented in effect sizes, estimating the difference between the two groups. (More detail of what an effect size is can be found in section 4.6.1). The primary outcomes of the pilot study are presented in Tables 4.1 and 4.2 below.

Table 4.1: Pilot overall English outcome: Pre-test, post-test and gain scores

Group	Pre-test (full=35)	Pre-test SD	Post-test (full=35)	Post-test SD	Gain score	Gain SD	Effect size
Intervention (n=26)	13.62	4.41	16.54	7.86	2.92	7.33	0.01
Control (n=11)	10.27	2.15	13.09	8.02	2.82	7.03	

Table 4.2: Pilot listening outcome: Pre-test, post-test and gain scores

Group	Pre-test (full=20)	Pre-test SD	Post-test (full=20)	Post-test SD	Gain score	Gain SD	Effect size
Intervention (n=26)	7.38	2.47	9.85	4.34	2.46	4.30	0.33
Control (n=11)	6.73	2.57	7.33	4.43	1.00	4.65	

The effect sizes reported in Tables 4.1 and 4.2 were based on the gain scores and the gain standard deviations since the participants were not randomised. Table 4.1 shows no difference between intervention and control participants in their gain attainment, indicating no effect of the intervention for overall English achievement. However, in terms of listening outcome, which was the primary target of the intervention, the intervention group shows a slightly, but meaningfully, higher attainment, as indicated by a small effect size of 0.33. This indicates the potential of the intervention for improving EFL listening and its potential for use in the main trial.

The secondary outcome obtained from the self-reported questionnaire was the students' metacognitive awareness in listening. Based on interpretation guidelines by the Vandergrift et al. (2006), responses were grouped into two categories, namely, positive and negative behaviours. The higher rating of positive behaviours and the lower rating of negative behaviours would reflect the desirable metacognitive awareness of learners.

Table 4.3: Pilot outcomes of metacognitive awareness in listening

Group	Positive Behaviours				Negative Behaviours			
	Post Mean	Post SD	Gain Mean	Gain SD	Post Mean	Post SD	Gain Mean	Gain SD
Intervention	3.45	0.75	-0.29	0.84	3.47	0.73	-0.16	0.62
Control	3.14	0.75	0.08	0.75	3.07	0.78	-0.06	0.83
Effect Size	0.40		-0.44		0.52		-0.14	

The secondary outcome showed mixed and unclear results. From Table 4.3, the gain-rating effect size on positive behaviours is -0.44 in favour of the control group, while the gain-rating effect size on negative behaviours is -0.14 which slightly favours the intervention group. By comparison, the post-rating effect size on positive behaviours is 0.40, favouring the intervention group, while the post-rating effect size on negative behaviours of 0.52 favours the control groups. Therefore, there is no clear evidence for the effect of the intervention on metacognitive awareness.

The results of the primary outcome suggest potential for the intervention to improve listening comprehension. However, results in the secondary outcome are unclear, as they indicate positive results post-questionnaire, but show negative results in gain-rating in positive and negative behaviours before and after the trial. This means that the intervention has the potential to be used in the main trial, but requires revision to better cater for the metacognitive awareness of learners.

4.3.5 Discussion and reflections from the pilot study

- **Promises of the intervention**

From the pre/post-test results, the intervention seems promising for improving listening comprehension, although the effect on overall language skills is trivial. However, the sample size is evidently too small to draw a conclusion. The intervention has potential worth for implementation with larger EFL learner cohorts in the main trial. Moreover, there was no barrier to the uptake of this relatively cheap intervention.

However, results of the secondary outcome are utterly unclear. One of the reasons for the non-effective results may be due to the short time span of the intervention. As speculated by Veenman et al. (2006), adequate length of the intervention is one of the suggested conditions for metacognitive activity. The presentation and practice of the intervention conducted in only five sessions over half a semester may not have been adequate. In addition, the confidence of the teacher delivering the intervention may have affected the result. The two-hour orientation before the implementation may not have been sufficient to support the teacher in delivering it confidently. More preparation before the trial, constant and consistent communication during the trial and other pedagogical supports for teachers such as materials could have been helpful. Another reason for the non-effective results may have been the abrupt change in teaching plan from on-site to an online platform caused by the COVID-19 outbreak.

- **Communications and support for the trial teachers**

In the pilot study, orientation with the intervention teacher before the start of term about the intervention and how it should be used was important, but still seemed insufficient for the teacher to use the intervention with confidence. The researcher communicated with the teacher outside the classroom several times throughout the pilot. To ease communication in the main trial to be conducted later, the researcher's connection with all participant teachers via a social media application would be helpful in allowing instant access to each other. Teacher-researcher discussion would also be arranged after each session to allow teachers to reflect on the session and prepare for ensuing sessions.

- **Students' reaction**

From the open-ended questionnaire responses, most students had positive comments about the intervention. Comments frequently indicate that the intervention helps students to learn better and plan their listening effectively. The intervention was also seen as being interesting in content and systematic in its method of teaching and was perceived to help students be aware of their own learning.

However, students also expressed some negative feedback on the intervention. The greatest issue was the complicated contents and difficult terms used in the intervention. Some students also thought that their limited English proficiency made learning and application of the intervention even more difficult.

- **Teacher reaction**

The teacher recognised the benefits of the intervention especially in terms of changing the attitudes of most students who think they are not good at English. She felt that some students liked the ideas presented in the intervention and started to open their minds to English. However, there were students who thought the intervention was too difficult for them because of their poor English proficiency. The teacher recommended simplifying some terms and providing materials for students to practise in their own time.

- **Adaptations of the intervention**

As the COVID situation was still unpredictable, it was decided that teaching in the main trial would still be conducted online. Thus, the online platform for intervention materials and instruments needed to be arranged to enable the online implementation of the intervention.

Revisions of the materials for students and teachers were required to make the intervention simpler and more concise. Some difficult terms were changed or eliminated. Notes were added in the student book to remind students not to worry about the correctness of their answers in the metacognitive tasks. Notes were also added to emphasise that the metacognitive strategies were adjustable to suit personal use in future listening. Retrieval practices were included in the pilot study, but were possibly not given adequate emphasis. The role of retrieval practices should be emphasised more in the main trial, as it would help students develop a more accurate awareness of their knowledge (Karpicke, 2009).

The spiral sequence would be used with each metacognitive session in the main trial. This means the following session would repeat the concept presented in the previous session. Such a sequence is aimed at helping students become more accustomed to each metacognitive strategy and recall what they have learnt.

Additional materials would be designed to support teachers. These would include a teacher's guidebook, online teaching platforms, and a multimedia package for instruction. Moreover, additional practices were designed for students to practise independently. This was is in line with the pilot teacher's recommendation. For each of the five metacognitive sessions, two independent practices were designed for students to practise further in their own time. The independent practices could be accessible by intervention teachers and students online to support independent learning despite the COVID outbreak.

- **Adaptations of the questionnaire**

Due to the students' difficulty in comprehending items in the English questionnaire, a Thai translation was added to each item. This was to ensure that students understood each item clearly, so that they could provide the most relevant responses that would genuinely reflect their opinion and behaviour. Back-translation was used. The English questionnaire was translated into Thai by the researcher and the Thai version was sent to an independent translator to translate it back into English. An English teacher not involved in the project was asked to compare both English versions of the questionnaire to check for congruence between them.

In addition, four closed-ended questions asking for students' overall perceptions about the intervention were added to the intervention group's post-questionnaire. Furthermore, open-ended items in the intervention group's questionnaire were edited to elicit more insightful opinions about the intervention. The open questions asked their general opinion about the intervention, positive and negative aspects of the intervention and suggestions. These questions allowed answers in Thai in order to encourage students to freely express their opinions and avoid a situation where students might not answer due to the language barrier.

4.4 Main study: A cluster RCT

The main study took the form of an effectiveness trial, whose ultimate aim was to investigate the impact of a model of metacognitive instruction on improving learning outcomes and enhancing the metacognitive awareness of English learners in Thai universities located in the disadvantaged area of southern Thailand. Specifically, the trial was implemented to address the following research questions:

RQ 1. To what extent does the metacognitive instruction have an impact on the listening and overall English achievement of EFL learners in southern Thai universities?

RQ 2. To what extent does the metacognitive instruction have an impact on metacognitive awareness for the listening of English learners in southern Thai universities?

RQ 3. In what manner is the impact of metacognitive instruction associated with differences in biographical variables such as gender, first language background, socio-economic backgrounds and pre-existing proficiency levels?

RQ 4. What are the teachers and students' perceptions towards the metacognitive instruction?

117

In order to study the development of learners' listening strategy, the two most common approaches are, first, to look at the learners' proficiency levels cross-sectionally and compare them with strategy use and, secondly, to implement a strategy instruction intervention and examine development (Graham et al., 2011). Of these approaches, the second approach, namely, the implementation of an intervention, can provide the most convincing and valid results, if information on strategy use is gathered both before and after the intervention (Graham et al., 2011). To answer the main research questions above on the impact of a metacognitive intervention on learners' outcomes, the research design must encompass the requisite elements to allow counterfactual evidence and causal inference. Therefore, the main empirical study of this thesis was based on a cluster Randomised Controlled Trial (RCT) highlighting both impact and process evaluations. The detail of the study is elaborated further below.

4.4.1 Study design: A cluster RCT

The primary empirical study was based on a cluster RCT because it is among the strongest designs for investigating the impact of a programme and is usually labelled as a 'gold standard' for assessing 'what works' (Hutchison & Styles, 2010). This is because an RCT holds all the essential elements for answering causal questions, namely, time and sequence, observations, intervention, comparison and randomisation (Gorard, 2013), as already discussed in section 4.1. The design of the study is notated and illustrated in Figure 4.5.

Figure 4.5: Notation of the design of the main trial
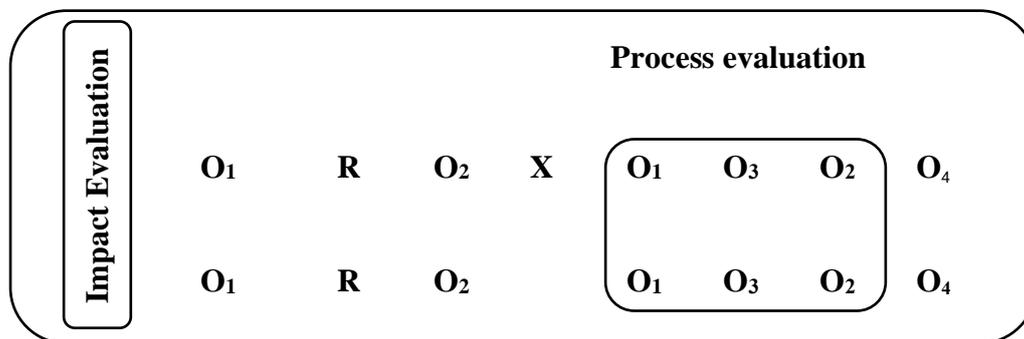
Note:

R = Randomisation   X = Intervention
Before intervention: O_1 = University Test data O_2 = Pre-test / Questionnaire
Process evaluation: O_1 = Observation O_2 = Interview
After intervention: O_3 = Post-test / Questionnaire O_4 = University Test data

As can be seen from Figure 4.5, randomisation (marked R) was operated to allocate students to either an intervention group or a control group. Random allocation occurred at a cluster level. This means that the clusters to which students belonged, rather than the individual students, were randomly allocated to either of the two conditions. Before randomisation, students' University Test results (O_1), were obtained, where applicable, to provide baseline data. After randomisation, participants were asked to take a pre-test and complete a questionnaire (O_2). Next, the intervention was introduced to the intervention group, while the control group continued with the normal course. During the implementation of the intervention, class observations were conducted for the purpose of process evaluation (O_1 in the box). Subsequently, a post-test and questionnaire were administered to measure primary and secondary outcomes (O_3). Later, on completion of the intervention, interviews were carried out as another measure of process evaluation. Finally, the results of the University Progress Test (O_4) were obtained as a delayed assessment.

Despite being a robust design, a RCT cannot solely guarantee to provide secure findings, since sources of bias can occur at several stages throughout the trial (Torgerson & Torgerson, 2003). The following sections will summarise other elements of the study and precautions taken to minimise bias and maintain validity.

4.4.2 Population, recruitment of participants and group allocation

The issue of who participants are and how they are recruited is an important consideration, as it can affect the credibility of results (Cohen et al., 2018). Results can be spurious if the characteristics of the samples do not match the aim of the study. To begin with, it is useful to understand basic terms and concepts related to participant recruitment.

Population refers to all of the units (human or non-human) that fit the context under study, while *sample* refers to the smaller group or subset of the population selected for investigation (Bryman, 2012). How this subset is sampled or recruited needs to be conducted in such a way that the recruited participants meet the aim of the study. Ideally, the whole population should be targeted where possible (Gorard, 2013). However, this is usually not possible in practice due to limitations in time, finance and other resources. Such limitations create the need for recruiting samples for investigation. A major question, therefore, is how many samples are needed? Numbers such as thirty have been suggested as a minimum. Indeed, there are many factors to consider, such as research questions and design, in determining the required sample size. A general rule of thumb is: the larger, the better (Cohen et al., 2018). For

impact evaluation, at least hundreds of participants per arm would be needed to allow for a reliable conclusion (Gorard et al., 2017).

4.4.2.1 Recruitment of participants

A main aim of the study was to implement and examine the impact of a metacognitive intervention on EFL learners in a disadvantaged tertiary education setting in Thailand (RQs 1, 2, 3). The research context was restricted to the southern border provinces as this region has a unique socio-cultural entity and endemic challenges (Pherali, 2021). With such an aim and context, three higher education institutions were considered eligible for the study. All of the three universities were contacted and invitations were sent to the head or person responsible for the English programme of the universities. Only two universities agreed to take part. The third declined to join because the semester there had already begun and the detailed syllabus of all courses had been finalised. The total number of students at the time of trial was approximately 6,500 at university A and 7,000 at university B. Both universities regulate the Exit English Exam scheme as demanded by central government policy.

A criterion sampling was applied in recruiting students who were taking a general English course at the universities in the first semester of the academic year 2020, so that the intervention could be implemented with all clusters at the same time. According to these criteria, six groups of second-year students in Political Sciences (Pol Sci) and eight groups of first-year students in Education at university A were eligible and recruited. Each group of Education students consisted of students from the same subject major. This would minimise dilution bias, since the chances would be slim for participants to share the treatment with people outside their groups (Torgerson & Torgerson, 2008) as they usually study together. The groups of Pol Sci students were also mainly from the same subject major. From university B, two large groups consisting of hundreds of second-year students with different subject majors were eligible to participate. However, by the time the researcher contacted university B, their semester was about to begin and there was insufficient time for teachers to embed the intervention into the course. Therefore, these two groups could not be included in the random allocation and could only take part as a comparison group. In total, 16 groups of students were recruited from both universities, forming 16 clusters. The members of each group were identified from each university's registration name list before random allocation. Teachers normally responsible for the English language courses at each university participated in and

delivered the intervention in order to minimise the experimenter's effect (Gorard et al., 2017) and enhance ecological validity (Higgins et al., 2005).

4.4.2.2 Random allocation

The random allocation occurred at the beginning of the semester before the launch of the intervention. It was administered by an independent third party who was undertaking a PhD in research methodology at the university's faculty of Science and Technology. This postgraduate student was blind to the intervention, the participants and the research focus with the intention of minimising bias. He was given a protocol (see Appendix 9) to work with. The allocation was administered via the R programme. Using a programme to assign recruited clusters to conditions would apparently avoid using personal judgment and, therefore, minimise selection bias.

In the allocation protocol, the recruited clusters were stratified by faculty to ensure that students from each faculty were proportionately present in each condition. This was done with the hope of levelling out any pre-existing difference in key characteristics, such as proficiency levels and instruction variables in the intervention and control conditions. Such restricted randomisation is particularly suitable for studies with a small sample size so as to minimise selection bias from the overall imbalance in the key covariates between the two groups (Torgerson & Torgerson, 2008).

Another point to note in the allocation protocol is that there were two teachers responsible for the six Pol Sci clusters, with each teacher taking three clusters. In order to avoid either of the teachers taking both intervention and control clusters - a situation which would have been at risk of contamination - the randomisation of Pol Sci clusters was based on teachers instead. This means that only one teacher was randomly assigned to the intervention group, while the other teacher would be in the control condition. The three clusters of students assigned to each teacher would follow their teacher into their corresponding group.

For the Education clusters, the faculty was experiencing a serious challenge of teacher shortage due to the resignation of some teachers and excessive workload during the COVID-19 outbreak. There was only one teacher taking care of all eight clusters. It would not have been possible to avoid having the teacher teach in both intervention and control conditions, unless the researcher joined as a teacher for the intervention clusters. However, such a situation would have caused another kind of bias from the researcher's effect (Gorard et al., 2017). Thus, this option was not chosen, since the potential bias would be more serious compared to having

one teacher taking all the clusters which may or may not cause contamination. This was one of the major limitations of the study. To minimise the effect of having only one teacher, the researcher emphasised that the teacher must be careful not to apply any part of the intervention to the control clusters until the post-trial data collection was complete. Moreover, the researcher joined as a teacher in one intervention cluster to allow comparison of the results with the regular teacher's groups. Therefore, each Education cluster could be randomly assigned to conditions. Three clusters were randomly allocated to the intervention group and five clusters were assigned to the control group. Among the three intervention clusters, one was randomly assigned to the researcher and the other two to the usual teacher. The randomisation procedure and results are illustrated in Figure 4.6.

After randomisation, three clusters of Education (EDU) students and three clusters of Political Sciences students (Pol Sci) were assigned to the intervention group, while five EDU clusters and three Pol Sci clusters were assigned to the control group. At the same time, two non-randomised clusters from university B were included for comparison. In total, there were 16 clusters with a total number of 920 students. Four teachers participated in either intervention or non-intervention clusters, except for one teacher who taught in both groups due to teacher shortage, as already explained above.

Figure 4.6: Random allocation procedures
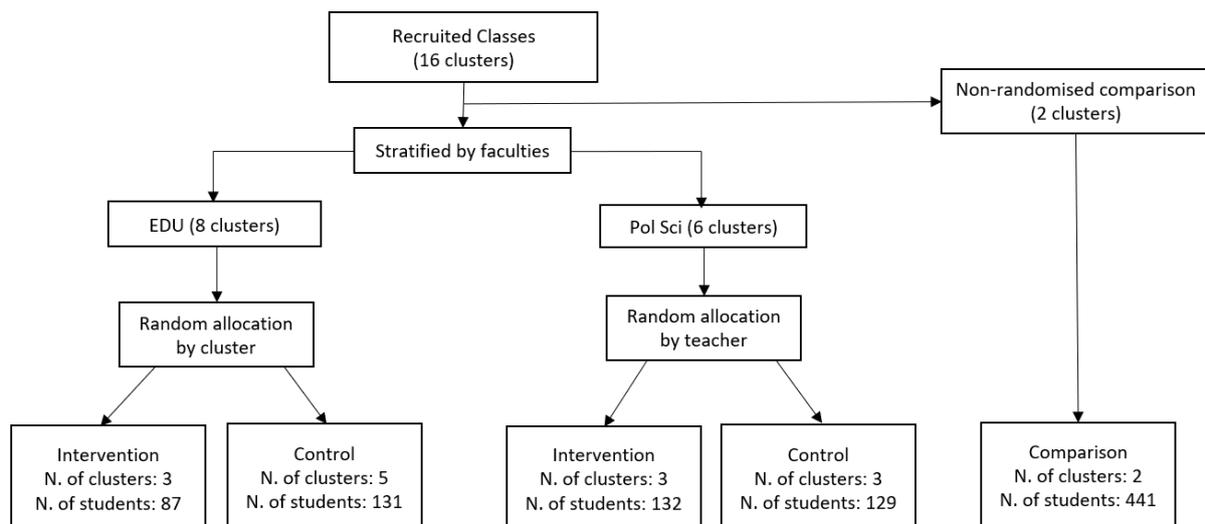
As the trial started in mid-2020 when there was still a serious concern about the COVID-19 outbreak at the trial site, another complication arose with the EDU clusters. Due to the COVID-19 outbreak, not every course in each faculty started at the same time. The re-opening of classroom teaching was carried out in phases. The EDU clusters began later than

the Pol Sci clusters. When it was time to begin the sessions for the EDU clusters, there were three clusters having to study during the same time period owing to students' timetables clashing with other subjects, as a result of which there was only one teacher available to teach the course. For these reasons, there was little opportunity to move things around. Among the three clusters, one had been previously assigned to the intervention group and the other two to the control group. This necessitated a re-randomisation of the EDU clusters and the three clusters studying during the same time period had to be treated as one big cluster. This unavoidably affected the scale of the study. Although the individual participants remained the same, the total number of education clusters decreased from eight to six, with three clusters being in the intervention group and another three in the control group (see Figure 4.7).

All students were concealed from the group allocation to avoid dilution and remorseful demoralisation caused by not being given the intervention or vice versa (Torgerson, Torgerson & Taylor, 2015). Therefore, although re-randomisation took place with the EDU clusters, there was no obvious effect on the students as they were unaware of the group allocation. Re-randomisation was unlikely to cause dilution bias, since it was carried out before the intervention was introduced. The chances of contamination between the EDU and Pol Sci students who received the intervention earlier were also slim because the students were from two different faculties and different years.

Figure 4.7: Re-randomisation of clusters
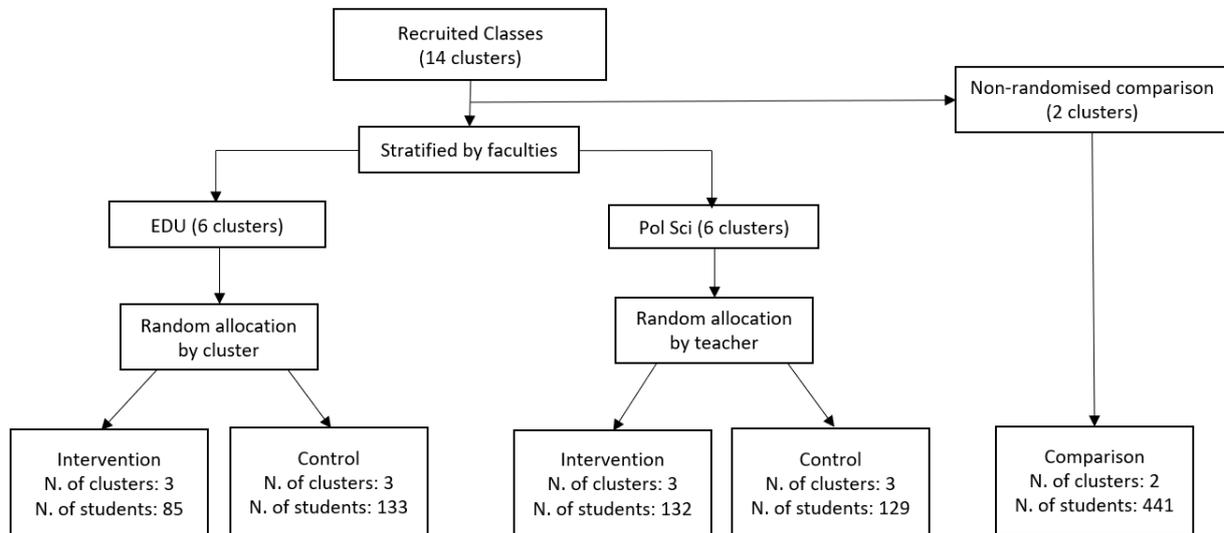
Approximately half of the randomised students were freshmen, while the other half of the randomised participants as well as the non-randomised comparison group were in their second year. By the time they took part in the trial, the students would have studied English for

at least 12-13 years. Among all participants, approximately three-fourths were of intermediate proficiency level or below, while less than a quarter were of upper-intermediate proficiency level, as suggested by their pre-test scores. With regard to randomised participants, almost two-thirds were of intermediate proficiency level or below, while around one third were of upper-intermediate proficiency level.

4.4.3 Comparability of the trial participants

How similar participants in the intervention and control groups are regarding the main outcome variable before commencement of the trial is an important issue to consider if any confounding variables which can be a threat to internal validity are to be avoided (Hedges, 2012). For this reason, it is common to see a comparison of baseline data or pre-tests to examine compatibility between the groups. However, some researchers argue that the randomisation, if done properly, will balance out the differences in any known and unknown variables in the two groups by chance and using tests to confirm this is not necessarily helpful if attritions occur during the trial (Torgerson & Torgerson, 2008).

In the current trial, participants were randomised by clusters and the total number of clusters included in the randomisation was small. Therefore, consideration and analysis of baseline data of the participants can provide an indication of how comparable the two groups are. A few sets of data, namely, scores of English admission tests, scores of previously taken university tests and pre-trial English tests were collected to provide baseline information about the participants. To shed light on their English proficiency, which was the primary outcome variable, pre-test results were used to examine how similar the participants were before the trial. Since other tests had been taken a long time ago, they may not represent the current English competence of students before the trial, since students might have developed or changed since taking those tests. Moreover, pre-test scores can demonstrate how effective the randomisation is in balancing participants in this key variable.

The pre-test results of both intervention and control students suggested that the difference between the two groups regarding English proficiency at the outset was small. The intervention group's mean score was 16.63 with an *SD* of 6.40, while the control group's mean was 16.06 with an *SD* of 6.92. To establish greater certainty, an effect size showing the difference in scores was calculated. The size of the difference was small (+0.09), with the intervention group being slightly ahead. Therefore, the main results will be based on gain score, which indicates progress, rather than absolute attainment. Further details of the pre-test scores can be found in Chapter Five, where the results are presented.

Certainly, any attrition occurring during the trial would affect analysis and inevitably cripple pre-intervention analysis (Torgerson & Torgerson, 2008). Bearing this in mind, the pre-test analysis was treated merely as a piece of baseline information. Factors such as attrition and other validity threats would be taken into consideration in the actual analysis.

4.4.4 Sample size and the projected effect size

Overall, there were 14 clusters with 920 students in total who agreed to participate. Among these, six clusters with 216 students were allocated to the intervention group and six clusters with 258 students to the control group, plus two non-randomised clusters with 441 students as a comparison group. As a cluster trial, having a total cluster of 14 is not considered a large trial. However, this number of clusters can be adequate to level out any chance of cluster-level covariation (Murray, 1998 cited in Torgerson & Torgerson, 2008 p.102).

As regards the scale of a controlled trial, it is important that the number of recruited participants is large enough to detect any projected effect of the intervention. Torgerson and Torgerson (2008) suggested that the minimum sample size of 128 with 64 members in each group would be required to reach 80 per cent power to detect an effect size of 0.5 in an individually randomised trial, and around 256 per arm for an effect size of 0.25. Gorard (2018) proposed that, to reach the NNTD (number of counterfactual cases needed to disturb a finding) of 50, which would avoid the chance of having the effect size equalling zero despite attritions, a minimum of 200 cases per arm or 400 in total are needed to detect a projected effect size of 0.25, or 250 per arm for an effect size of 0.2, and approximately 170 cases per arm for an effect size of 0.3, for example. Meanwhile, the EEF (2019)'s criterion suggests that EEF studies should have an adequate sample size to detect a minimum detectable effect (MDES) of 0.2, which would require hundreds of participants per arm.

For this trial, the projected effect size was based on the average effect size of a previous similar study by Chou (2017), which had an effect size of 0.52, and the results from the pilot trial of this study, which had an effect size of 0.33, giving an average of 0.43. This means a minimum of 128 cases per arm would be required to detect such an effect size. However, as a cluster trial, the required number of participants would be larger than an individually randomised trial to account for a possible clustering effect (Campbell et al., 2000). For this purpose, the intra-cluster correlation coefficient (ICC), which shows the extent to which cluster members are dependent on each other, needs to be considered. A higher ICC indicates a higher possible effect of clustering. As the ICC was not found in reports of previous studies of the

same intervention, Killip et al. (2004) suggest that, as a general rule of thumb, the ICC values for human studies are generally between 0.01-0.02, while Torgerson & Torgerson (2008) argue that the ICC of 0.05 is not uncommon.

Based on Torgerson and Torgerson (2008), a cluster trial with an ICC of 0.05 to detect an effect size of 0.5 where the cluster size is 25, the numbers needed for the trial would increase from 128 in individual allocation to 282 for cluster allocation. Based on this estimate, to detect the effect size of 0.4 while accounting for the ICC of 0.05, the number of 128 cases per arm required in the individually randomised trial would increase to 377.6. Considering that participants of this trial in the intervention and control groups numbered 216 and 258 respectively, the sample size was a little under power. However, the inclusion of two clusters of 441 students would make the numbers of intervention/non-intervention participants reach a 1:2 ratio, which could increase the likelihood of detecting the projected effect size to 90 per cent power (Torgerson et al., 2015).

Despite several complications, challenges and modifications caused by the outbreak of the COVID-19 disease and the limitations within the trial sites, this trial is the largest study on the topic compared to any trials found in the systematic review. The researcher believes that the recruited number of participants would suffice to allow successful implementation of the trial and provide meaningful results. Thus, the next section will discuss the intervention and how it was implemented.

4.4.5 The intervention and delivery procedures

The metacognitive intervention for the study was based on the PMER model which has been discussed in detail in section 4.1.3 and Chapter Two. The intervention comprised three phases, i.e., orientation, regulation and facilitation. The orientation phase was to provide the students with an awareness of different types of strategies for listening comprehension and an introduction to metacognitive awareness. As the EEF's guideline by Quigley et al. (2018) suggests, to be metacognitive, other kinds of competences are required. Moreover, the orientation was to provide students with backgrounds to the concept of metacognition and how it can be applied to English listening. Next was the regulation phase, a pivotal part of the intervention which provided opportunities for learners to apply such knowledge and strategies in real listening tasks. The six-stage sequence shown in Figure 4.2 above was designed to guide students in their metacognitive regulation. In their attempts to regulate their learning metacognitively, facilitation and supports are highly advantageous for the learners. Materials

and worked examples were provided for students to support their self-regulation. Moreover, lesson plans, class materials and instant support were provided to teachers to facilitate them in appropriately supporting their students. Such supports were largely based on suggestions from the pilot study and the systematic review discussed in Chapter Three.

Apart from a right balance of strategy sets, the metacognitive processes involved, and a clear pedagogical sequence, the degree of interaction between the listener and the speaker is another factor which influences comprehension (Buck, 2001). The listening tasks in the PMER model only included non-collaborative listening. This was because non-collaborative listening could be more easily and objectively assessed via standardised tests. It is also featured in the university's standardised test which the participants need to take in real life.

As a part of the facilitation, online resources for independent practices were also prepared for the students. This was motivated by several reasons. Quigley et al. (2018) encourage the inclusion of independent practices in metacognition-based programmes to promote self-regulation. This was confirmed by findings from the systematic review presented in Chapter Three. From an empirical perspective, the teacher in the pilot study suggested that having materials for students' independent practice would be helpful. With the fact that the intervention was focused in only five sessions throughout the semester, having support materials for students to practise at their own pace seemed essential. Samples of the online independent practices are shown in Appendix 10.

Teachers play a significant role in delivering the intervention and supporting students. However, many teachers perceived that they had insufficient knowledge and understanding of metacognitive instruction (Wilson & Bai, 2010). For this reason, a training session, materials and tools were arranged to support the intervention teachers. A two-hour training session and a teachers' manual were provided to help them understand the fundamentals of metacognitive instruction in their listening sessions. An online chatroom was created and phone numbers of the researcher and the teachers were exchanged to allow instant and constant discussion and support for the teachers throughout the course. In addition, lesson plans, class materials and multimedia tools for each metacognitive session were supplied as suggested by the pilot study teacher. The teachers were advised to use the given materials to ensure fidelity, but were allowed to change or adapt the material to fit the syllabus. As the main trial was implemented during the COVID-19 outbreak, an online platform for distance teaching and learning was also

created for participants. All these supports for teachers were provided so that they could efficiently support their students.

Figure 4.8: Delivery sequence of the metacognitive intervention
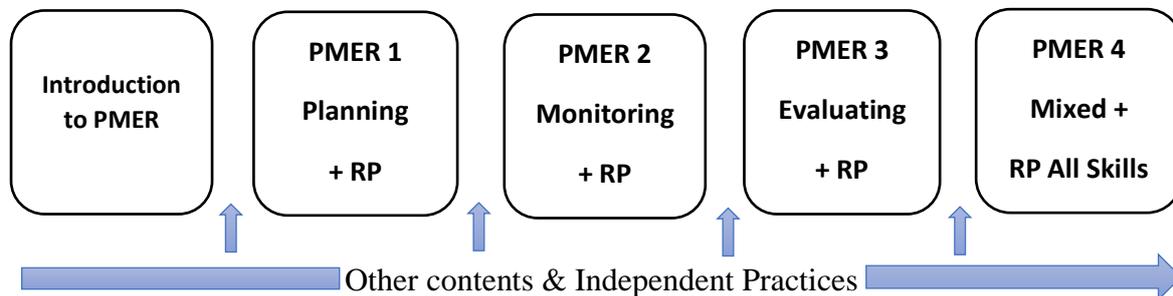
As illustrated in Figure 4.8, the intervention was implemented with the intervention clusters over five classroom sessions throughout the first semester of the academic year 2020. In addition, two independent practices for each session were provided, giving ten practices in total. These practices were available online for students to learn at their own time outside the classroom.

The first session was an introduction guiding students on different strategies for effective listening. Listening strategies, cognitive strategies and affective strategies were presented before metacognitive strategies were introduced. This was aimed at raising students' awareness of the various strategies for listening, since the awareness and application of available relevant strategies can make the listening process easier and more successful (Graham, 2007). The session was also to provide a fundamental understanding of metacognition and self-regulation and adequate relevant skills to be practically applied in ensuing sessions. The subsequent three sessions focused on each metacognitive process, namely, planning, monitoring and evaluating with a retrieval practice (RP) in each of them. The final session was a conclusion of all the strategies. Online independent practices were available for self-paced revision after each session. The detail of each session is presented below.

Session 1 Introduction
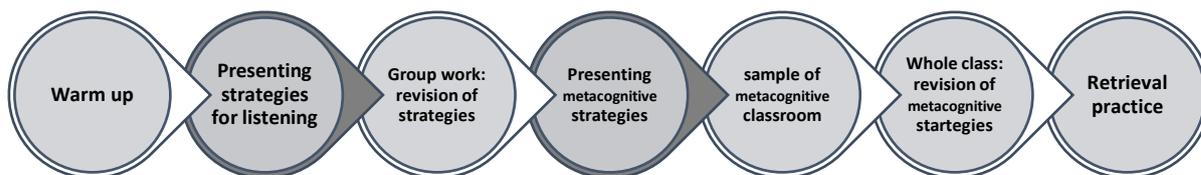
In the first session, the main aim was to activate background knowledge about different types of listening strategies and introduce the concept of metacognition. In the warm-up, a short movie scene with a character reflecting on himself was shown to the students. Next, different types of strategies for listening comprehension, i.e., listening, cognitive and affective strategies were presented. Group work to review the presented strategies was launched before the metacognitive strategies were introduced. Then, a video of a sample metacognitive classroom was displayed for a clearer picture. This was followed by a whole class discussion led by the teacher to reiterate the presented metacognitive strategies. Finally, students were asked to complete a retrieval task and discuss answers together to help them realise how much they had learnt from the lesson. Students were also encouraged to carry out the Independent Practices online to consolidate their knowledge and skills.

Session 2 Planning
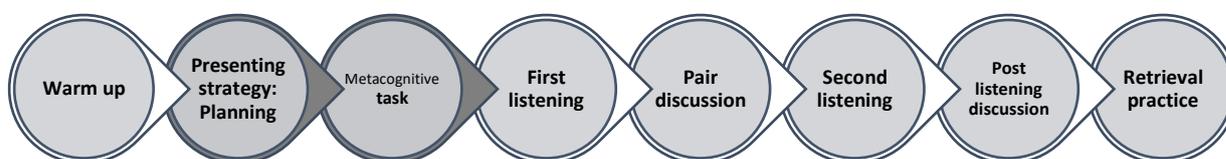
Session 2 focused on the first metacognitive strategy, i.e., planning. In the warm-up, the listening topic and the listening task were introduced. Before listening, the planning strategy was presented and a pair-work metacognitive task was given to aid students' planning of their listening. Next, the students tried the first listening task and attempted to use their plan to benefit their listening. After the first listening, there was a brief time gap for pair discussion to reflect on their first attempt and modify strategies for the second listening. In the second listening, the students were guided on how to approach the listening task with more selective attention and modify some strategies. This was followed by a whole class discussion to

encourage students to evaluate their performance and understanding. Finally, a retrieval task was introduced, followed by a discussion of answers. Reminders were made about the online Independent Practices.

Session 3 Monitoring
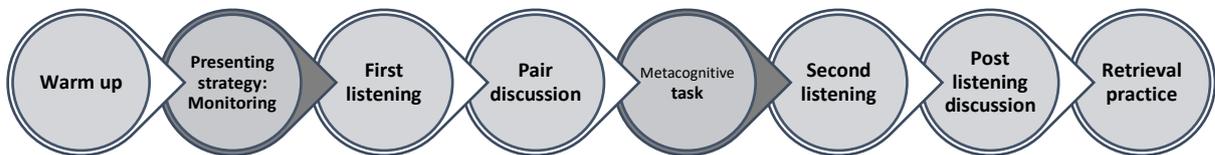
Session 3 focused on the second metacognitive strategy, namely, monitoring. In the warm-up, the listening topic and the listening task were introduced. The monitoring strategy was presented before the first listening so that students could apply the strategy to monitor their listening. After the first listening, pair-work discussion was initiated to prompt students' reflection while completing the metacognitive task. The task elicited students' reflection on their first attempt and modification of strategies for the next attempt. In the second listening, students were guided on how to approach the listening task with more selective attention and modify some strategies. Next, a whole class discussion was launched, followed by a retrieval practice, discussions of answers and a reminder of the Independent Practices.

Session 4 Evaluating
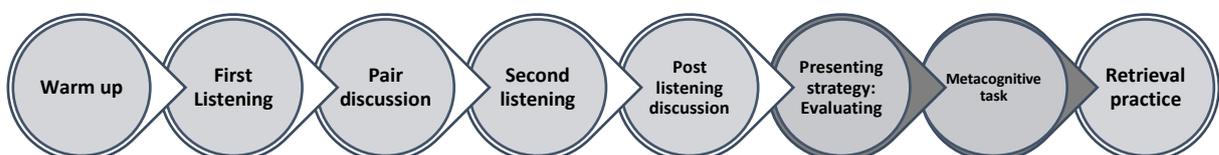
Session 4 focused on the third metacognitive strategy, i.e., evaluating. The listening topic and the listening task were introduced in the warm-up phase. Next, the students were given some time to prepare for the first listening. After the first listening attempt, there was a brief break for pair discussion to reflect on their first attempt and modify strategies for the second listening. In the second listening, the students tried to approach the listening task with more selective attention and modified strategies. Then, a whole class discussion was prompted to discuss how they performed in the task. Next, the evaluating strategy was presented,

followed by a metacognitive task to urge students to evaluate their understanding and performance. Finally, a retrieval practice was given, followed by discussions of answers and a reminder of the Independent Practices.

Session 5 Put it all together

Warm up → Planning listening → First listening → Monitoring listening → Second listening → Evaluating listening → Warp up → Retrieval practice
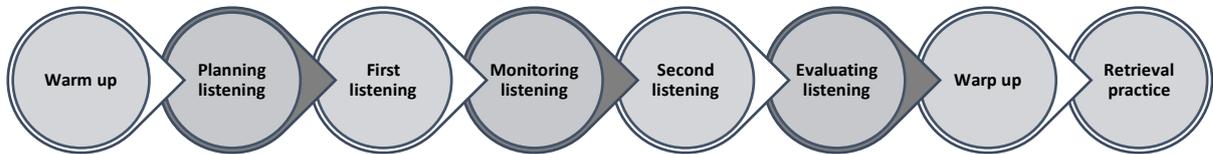
Session 5 was aimed to promote the regulation of all strategies in a complete cycle. The listening topic and the listening task were introduced in the warm-up. Before listening, the students completed a metacognitive task to plan their listening. Next, the students tried the first listening and attempted to use what they had planned to benefit their listening. After the first listening, pair discussion was prompted for students' reflection of their performance and completion of the monitoring task. In the second listening, the students tried to approach the listening task with more selective attention and modified strategies. Then, a whole class discussion was launched to prompt students to reflect on their performance and understanding and use it to complete an evaluation task. At the end of this, all the presented strategies were summarised again. Finally, a retrieval task was introduced, followed by a discussion of answers. Reminders were made about the online Independent Practices.

Independent Practices

To provide the students with further support to practise metacognitive regulation in listening at their own pace, two independent practices were created for each session. Both were available online via Google Classroom. The online platform allowed learners to receive feedback after submission of their completed work and permitted teachers to monitor students' participation. The practices were based on self-questioning prompts similar to the classroom tasks. The practices were not made mandatory but consistently encouraged by the teachers.

Despite the extensive details presented about the metacognitive intervention, it was emphasised to the participants that the aim of metacognitive instruction is not about the specific terms or strategies used in the intervention, but rather self-efficacy, self-reflection and strategic thinking which would be positive dispositions in assisting them to overcome challenges. As

Perkins and Salomon (2012) and Wegerif et al. (2015) suggest, the focus on teaching thinking should be shifted from teaching a set of skills to enduring positive dispositions, because such positive dispositions are transferable across contexts.

4.4.6 Outcome measures and procedures

To derive answers to all the research questions, both impact and process evaluations were carried out in the main trial. For impact evaluation, standardised English tests adapted from Tell Me More (TMM) software were used as a pre-test and a post-test before and after the intervention. The TMM sample tests were obtained by permission from authorised staff at one of the participating universities. The pre-test and post-test items of the trial were proportionately taken from six samples of the TMM test. Both the pre-test and post-test contained 35 items with 20 items on listening comprehension and 15 items on language use. They were the main instrument for measuring the primary outcomes of the study, i.e. the students' listening comprehension and overall English achievement. The tests could be marked objectively, which would avoid marker's bias. Moreover, in the main trial, the tests were administered online under the teacher's supervision. The online platform was set to mark students' answers to avoid possible mistakes from hand marking and eliminate assessment bias.

A questionnaire adapted from Vandergrift et al. (2006) was used to assess the secondary outcome, i.e. metacognitive awareness in listening. The self-report questionnaire is the most common tool for assessing metacognition (Gascoine et al., 2017). As learnt from the pilot study, some students had problems understanding certain items. Thus, the questionnaire items were translated into Thai and underwent a back-translation into English by an independent translator to check for congruence. A comparison of the original English version and the back-translation was analysed by an English teacher who was not involved in the research project. Four additional closed items were included in the post-questionnaire of the intervention group to assess students' overall satisfaction and perception about the intervention. To measure the accuracy of students' self-reported metacognitive judgment, a relative accuracy measurement was conducted by investigating the correlation between the students' metacognitive judgment and the relevant performance outcomes (Schraw, 2009), which were the students' test score in this study.

For process evaluation, class observations took place in both intervention and control clusters to observe fidelity, contamination and difficulties. As part of the process evaluation, open-ended questions were included in the intervention clusters' post-questionnaire to elicit

participants' reflections and opinions further to provide insights about the intervention. Furthermore, semi-structured interviews were conducted with all teachers and volunteering intervention students for more in-depth expressions and suggestions about the intervention, the exit exam policy and English education in Thailand in general. Details of the data collection tools have been discussed at length in section 4.1.4.

The administration of each tool was conducted at different times throughout the trial. The pre-test and pre-questionnaire were administered by the class teachers at the beginning of the semester. The pre-test was conducted online under the teachers' supervision due to the COVID-19 outbreak and the suspension of onsite teaching. The pre-questionnaire was also administered online with permission for students to complete it at their own time to allow time to consider and provide genuine responses.

The classroom observations were conducted by the researcher. These were initially planned to occur twice in each teacher's classes but, due to the COVID-19 outbreak, there were a number of modifications, allowing only one-time observation in each teacher's classes. However, communication with the intervention teachers occurred frequently outside the classroom throughout the semester. The comparison teacher was observed online, since university B only allowed online teaching while the remaining observations happened in class. Because of the limited chances of observations, the observation of some classes occurred early at the beginning of the course and others took place later in the semester, so that both early stages and later stages of implementation were observed.

Semi-structured interviews were administered after the trial was over so that the participants had experienced the entire course of the intervention. Moreover, these took place after the post-test was complete, so that students could observe how much the intervention had improved their learning. The intervention students were asked to join the interviews on a voluntary basis. Interviews were conducted on the phone in the hope of minimising students' pressure. Interviews were also conducted with all teachers in person. The aim of the interviews was to explore participants' perceptions about the intervention, the difficulties they may have experienced during its implementation, the benefits and the drawbacks of the intervention and suggestions which could improve the implementation of metacognitive instruction. Moreover, the interviews sought their opinions about the exit exam policy and English teaching in Thailand in general. All the interviews were conducted by the researcher and the Thai language

was used to allow respondents to express their thoughts more freely. The interviews were then translated into English and transcribed for analysis.

4.4.7 Tracing impact: the logic model of the PMER intervention

To trace changes that may be caused by a programme or an intervention, a logic model can be a helpful tool. The logic model represents a logical flow of how a programme or an intervention might work under certain conditions and bring about changes to solve identified problems (McLaughlin & Jordan, 1999; Renger & Titcomb, 2002). The model exhibits the logical linkage among inputs, activities or intervention, outputs, audiences and outcomes (McCawley, 2001). It is useful for conceptualising, planning, and communicating about the programme and helps to identify which important elements of the programme to focus on, what questions to consider and what measures of performance to use (McLaughlin & Jordan, 2015). With such capacity, the logic model is often referred to when describing the theory of change in programme evaluation.

The logic model consists of three fundamental parts: programme structure, outcome structure and context (McLaughlin & Jordan, 2015). The *programme structure* comprises resources, activities and outputs. *Outcomes* are the changes or benefits resulting from exposure to activities. The outcomes can be multiple and sequential, forming an outcome structure, i.e., the short-term, intermediate and long-term outcomes. *Contextual* factors that occur at the start or during the implementation of the model should be taken into consideration as they can influence the success of a programme.

Based on the components suggested by McLaughlin and Jordan (2015), a logic model for tracing the impact of the current study's metacognitive intervention is illustrated in Figure 4.9. Students in cohorts 2019 and 2020 were the inputs of the trial. Their pre-existing English competence indicators were used as baseline data. After randomisation, the pre-test and the MALQ questionnaire were administered to provide more apparent pre-intervention data about the target outcomes. During the trial in which one group received the intervention and the control and the non-randomised comparison group were in a normal classroom setting, classroom observations were conducted in all groups. After receiving the training, which yielded development in knowledge and skills as the outputs, all trial participants took the post-test and the MALQ questionnaire to measure their language outcomes and metacognitive awareness in listening. As a delayed post-test, the results from the university English test which students in one of the universities had to take were used to examine the long-term effect of the

intervention. The changes in the pre/post English tests and the MALQ questionnaire before and after the intervention and the results of the university English test among both intervention and non-intervention participants would determine the impact of the intervention.

Figure 4.9: Logic model for tracing impact of the metacognitive intervention
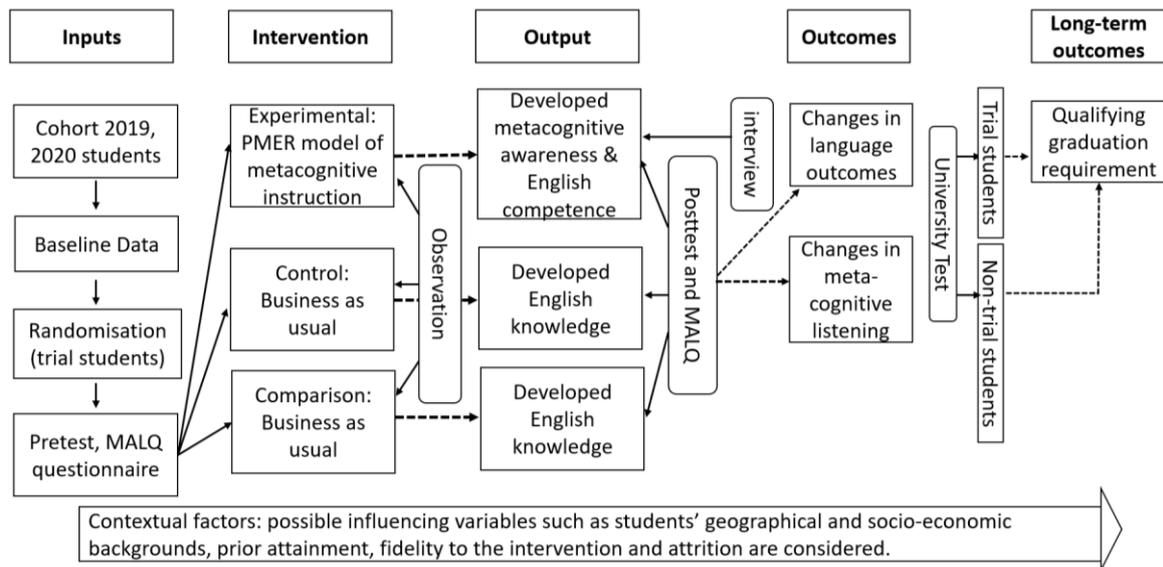
As illustrated in Figure 4.9, the PMER intervention is the primary factor for driving a distinguishable change in the outcomes. Thus, how promising the expected outcome will be is essentially determined by the capacity of the intervention. The mechanisms of the PMER intervention that will bring about the expected outcome of the intervention students are presented in Figure 4.10.

Figure 4.10: Mechanism of the PMER model for improving target outcomes
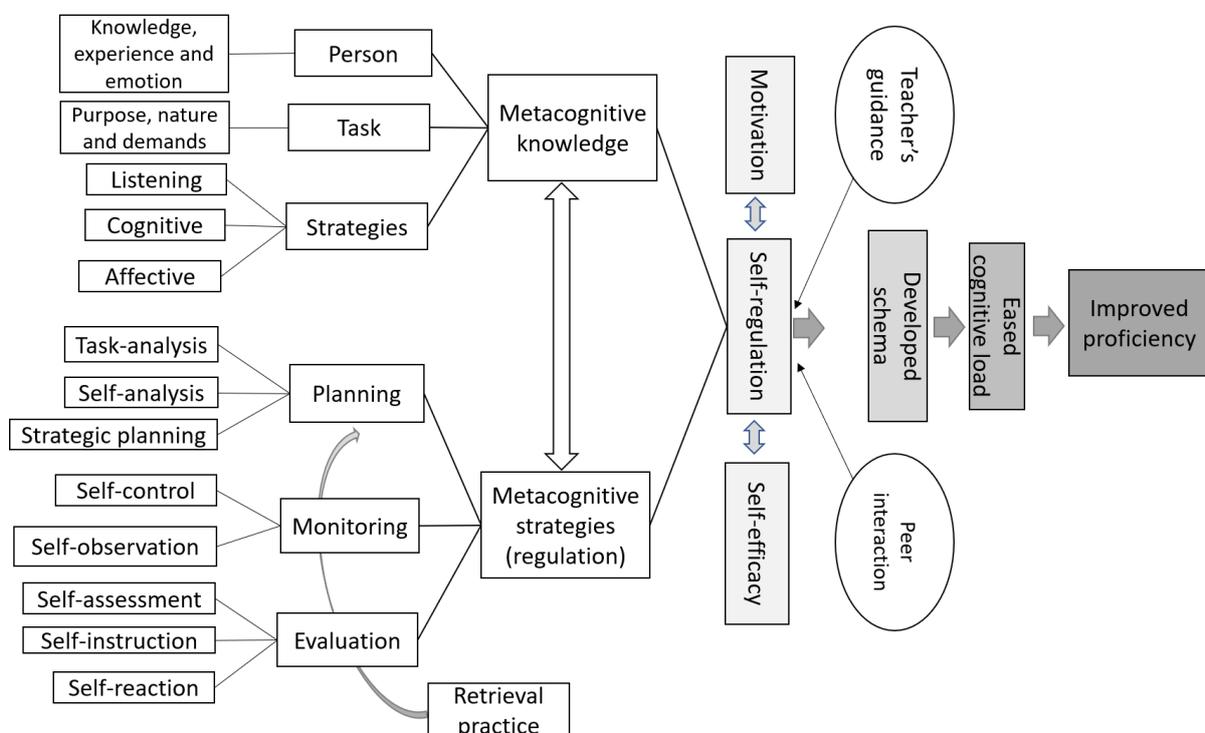
For effective learning, a learner should be aware of the metacognitive knowledge about the learning task at hand, personal knowledge relevant to the task and the available strategies that will be helpful for the task. This knowledge is required in metacognitive regulation, which occurs through the processes of planning, monitoring and evaluating one's own learning. The knowledge and skills of these processes are strengthened by retrieval practices. Such metacognitive awareness of strategies and guidance on how and when to apply them can support learners to be more strategic, which in turn promotes self-efficacy and motivation to self-regulate. Self-regulation requires the learners' agentic role in those metacognitive and self-regulatory processes and could be enhanced by teachers' support and peer interactions. Self-regulatory processes lead to developed schemas, which help ease the cognitive load during information processing. The mechanism of these interconnected processes can bring about development in English proficiency.

4.5 Secondary Data

The availability of official data such as admission scores is useful for public policy evaluation, as it allows secondary analysis to be conducted (Smith, 2008). Secondary analysis of large-scale data can provide empirical evidence for educational research, predict people's lifelong education trajectories (Smith, 2012) and provide a more objective indication about policy and practice than local perception-based commentaries (See et al., 2004).

In this study, the relevant secondary data which could be obtained was the results of the Admission English Tests of the whole 2019 student cohort, parental income data and the full-cohort results of the university's English tests. These data sets could be obtained from only one of the participating universities. The data can provide a clearer picture of the characteristics of the participants and population, which are tertiary EFL learners in the southernmost area of Thailand. Moreover, it sheds more light on the impact of the intervention, because comparisons of the trial and non-trial data are enabled.

4.6 Data analysis

As the main trial involved impact and process evaluations, both statistical analysis and thematic analysis were operated. The effect sizes of the intervention on the learning outcomes were calculated in answering Research Question 1. The standardised mean difference of the questionnaire responses provided an answer to Research Question 2 regarding the impact on metacognitive awareness in listening. The correlation and regression analysis of key background variables such as income, gender, first language backgrounds and English

proficiency backgrounds supplied an answer to Research Question 3. For Research Question 4 on teachers and students' perceptions towards the intervention, descriptive statistics and thematic analysis of the questionnaire, observation notes and interviews were conducted as a process evaluation. How each analytical measure provides answers to the research questions is summarised in Figure 4.11. Details of the measures of analysis are presented in the sessions below.

Figure 4.11: Overall procedures of data analysis
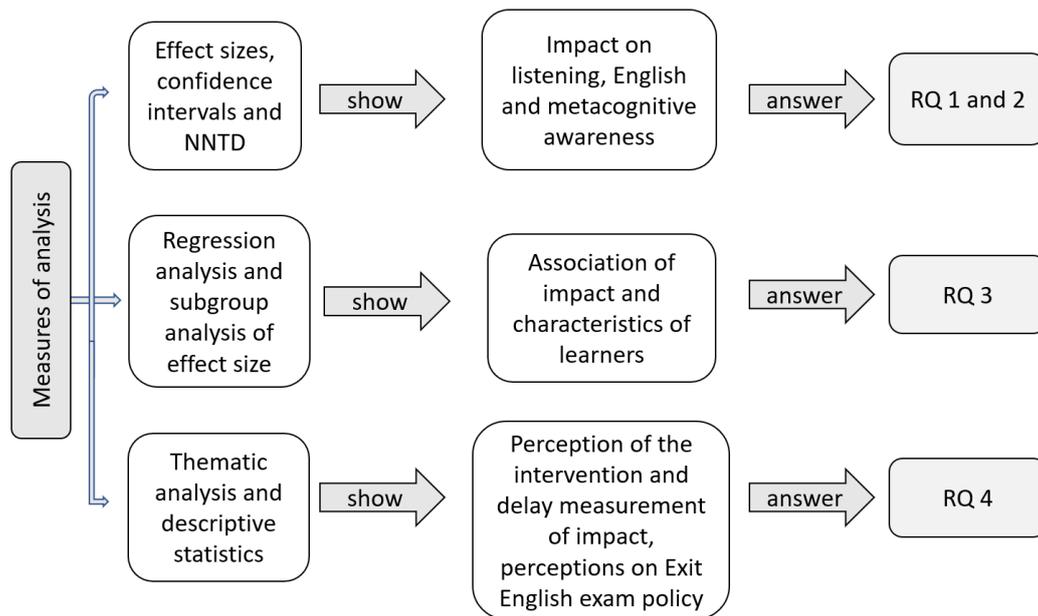
4.6.1 Measuring impacts: Effect size calculation

To investigate the impact of the metacognitive intervention on language outcomes and metacognitive awareness to answer research questions 1 and 2, effect sizes were calculated. In a broad sense, the effect size is "the degree to which the phenomenon exists" (Cohen, 1988 p.4). There are two main types of effect sizes. One is the standardised mean difference and the other is about an association of variables (Ellis, 2010) or variance accounted for (Elliot & Sammons, 2004). The first type, also known as the *d* family, measures the difference between groups, while the second type, also referred to as the *r* family, measures associations or correlations between variables. For impact evaluation trials, the effect size provides an estimate of the difference between groups and indicates the practical significance of the results (Coe, 2002).

Effect size is an appropriate tool for reporting and interpreting effectiveness. Reporting an effect size tells us not only 'does it work?' but the estimated size of the effect also indicates

'how well it works' (Coe, 2002). This information is more meaningful than the p-value for decision-making in policy and practice (Elliot & Sammons, 2004). The reporting of effect sizes, which does not usually present sophisticated statistical values, makes the research report more accessible to a broader audience (Ellis, 2010). Moreover, the reporting of effect sizes facilitates the processes of synthesising the effect of interventions, which aids a systematic and rigorous approach to the accumulation of knowledge (Higgins, 2018). To emphasise the importance of the effect size, the American Psychological Association (APA) has encouraged authors to report effect size since 1994 (Wilkinson et al., 1999).

Significance testing, which is commonly seen in research reports in social sciences, is not included in this study for several reasons. In terms of appropriacy, significance testing does not fit the sampling method of this trial. The calculation of significance testing assumes that the samples were randomly selected from the population (Gorard, 2010). However, this study placed more attention to random allocation to maintain internal validity and recruited the samples based on criterion, rather than random sampling. Regardless of this sampling issue, significance testing is still not the best option. Significance testing can only tell if the difference between groups is 'statistically significant', but does not estimate the size of the effect which would inform how well the intervention works (Coe, 2002). Moreover, significance testing, which is usually demonstrated by the p-value, depends essentially on the size of the sample and the size of the effect. Therefore, there are chances that one would achieve a significant result if the sample is very big, regardless of the actual effect (Higgins, 2018). With the aforementioned reasons and issues, significance testing is considered irrelevant for analysis in this study.

Regarding the effect size in terms of standardised mean difference, several values are commonly used to provide an effect estimate such as Cohen's *d* or Hedges' *g*. The effect size is derived by subtracting the mean of one group from the other (M1 – M2) and dividing the result by a standard deviation (Coe, 2002). One point to consider, then, is which standard deviation to use? For Cohen's *d*, the pooled SD is used, while Hedge's *g* uses a pooled standard deviation of each group which is weighted by its sample size (Ellis, 2010). The pooled effect size is "essentially the average of the standard deviations of the experimental and control groups" (Coe, 2004 p.90). What makes *g* distinct from *d* is the use of the sample size of each group to weight the results. As this trial has different numbers of participants in the two groups, Hedges' *g* was chosen for this study as it can account for the difference. In calculating effect sizes, the gain score and gain SD were used because there was a slight pre-existing difference

between the groups. The effect size based on post-test scores only was also calculated for comparison, as the approach is less affected by the additional measurement error from the pre-test and post-test (Gorard, 2013).

The reporting of effect size provides meaningful interpretations of the results. Rather than merely telling if the difference between groups is statistically significant, it indicates an estimate of the size of the effect which can inform how well an intervention works. Cohen (1988) proposes a guideline for interpreting the effect sizes. Based on this guideline, 0.2 shows a small effect size, 0.5 is a medium size and 0.8 is a large size. For educational interventions, Hattie (2009) suggests that an effect size of 0.4 or larger provides a desirable benchmark for promising replications. This guideline is helpful for interpreting results but should not be treated as an interpretation dogma, since there are other factors to consider, such as the design, sample size, measurement reliability as well as the cost and feasibility of the intervention (Coe, 2004; Higgins, 2018).

Coe (2002) provides a thorough guideline for interpreting effect sizes and what they can mean in a practical sense (Table 4.4). Given that the participants in the two groups are balanced in the target outcome variable before the intervention, the effect size of 0.0 indicates that the percentage of students in the control groups who would be below an average person in the experimental group would be around 50. In other words, half of the control group is below the mean of the experimental group. The effect size of 0.2, for instance, would suggest the mean of the intervention group became higher, making more students (58%) in the control group perform below the intervention group's average. The larger effect sizes indicate a higher percentage of control students who are outperformed by the intervention group. It should also be noted that the effect size can be below zero, which shows the effect is moving in the opposite direction.

Table 4.4: Interpretation of effect sizes (Coe, 2002 p.4)

Effect Size	Percentage of control group who would be below average person in the experimental group
0.0	50%
0.1	54%
0.2	58%
0.3	62%

0.4	66%
0.5	69%
0.6	73%
0.7	76%
0.8	79%
0.9	82%
1.0	84%
1.2	88%
1.4	92%
1.6	95%
1.8	96%
2.0	98%
2.5	99%
3.0	99.9%

For Research Question 2, the effect size derived from the standardised mean difference can also be calculated to inform the difference in the participants' perception about their awareness and application of metacognitive strategies. One advantage of reporting effect sizes in standardised terms is that the results are scale-free, which means that the effect sizes across studies can be compared (Ellis, 2010). Participants' responses to a questionnaire can be transformed into effect sizes to show the effect of a particular intervention (Coe, 2004). Moreover, as the secondary outcome data was elicited from a self-report questionnaire, an accuracy of judgment analysis was also conducted. For this purpose, a relative accuracy measurement was conducted by examining the correlation between the students' metacognitive judgment and the relevant performance outcomes (Schraw, 2009).

In reporting effect sizes, another value that is frequently reported together is the confidence interval (CI). The rationale for reporting CI is that it provides an indication of the uncertainty and precision associated with effect sizes (Ellis, 2010). The CI is reported in a range with a lower end and an upper end, indicating the range within which the effect sizes can possibly fall in varied measurements of similar samples (Coe, 2022).

However, the reporting of CI has been questioned by some commentators. Gorard et al. (2017) argue that the CI is based on assumptions similar to significance testing, which has several flaws as discussed earlier, and is often misinterpreted. One of the misinterpretations of

the CI is to conclude that a level of confidence such as 95% would suggest there is a 95% probability that the interval contains the parameter or the mean population (Ellis, 2010; Gorard et al., 2017). With the multiple assumptions that can lead to misinterpretations, the CI was not reported in this study. Alternatively, the number of counterfactual cases needed to disturb the findings (NNTD) was calculated and reported as a measure of the robustness of the results. Further details of NNTD are presented in the next section.

4.6.2 Dealing with attrition: sensitivity analysis

Attrition or missing cases can commonly occur in any trial. Missing data and how they are dealt with are an important consideration in examining the trustworthiness of the findings, because they can indicate the degree of bias which undermines the estimate of the treatment effect (Gorard et al., 2017). Even if the missing cases are few and balanced between comparator groups, such attrition needs to be dealt with carefully, because the characteristics of the missing cases can still be different, which inevitably leads to bias (Dong & Lipsey, 2011). Such bias can subvert the result and threaten the validity of an interpretation or a conclusion (Shadish et al., 2002).

To examine the missing cases and their possible effect on the result, an analysis of data pattern was conducted to compare the pre-test scores of the students who remained until the end of the trial and the pre-test scores of the students who missed their post-test. This can allow a projection of how the result might be affected if the post-test results of the missed cases could have been obtained, based on their pre-test scores.

As a measure to account for the missing cases, Gorard and Gorard (2016) propose a calculation of the number of counterfactual cases needed to disturb the findings (NNTD) as a way to estimate the sensitivity of the results or the potential effect of the missing data on the obtained results. The NNTD indicates how many counterfactual cases would be needed to change the obtained results to the opposite direction. It assumes that all the missing cases showed negative progress and determines whether the effect will be reversed had such missing cases been included. The NNTD can be a measure of robustness in the findings, similar to the CI, as its calculation includes the mean, standard deviation and the number of cases, and it can be a preferable option because it requires fewer assumptions and can cater for missing data (Gorard, 2021).

To calculate the NNTD, one simple way is to multiply the effect size by the number of cases of the group with smaller remaining participants (Gorard et al., 2017). The ensuing number is the number of counterfactual cases needed to subvert the effect to the opposite

direction. This means that if the number of missing cases is smaller than the NNTD value, the obtained result would not be reversed despite all the missing cases having negative outcomes. On the contrary, if the number of missing cases is higher than the NNTD, there is a possibility that the results can be reversed with the inclusion of those missing data.

4.6.3 Examining association: regression analysis and subgroup standardised mean difference

Research Question 3 seeks to examine the association between the impact of metacognitive instruction and the differences in gender, pre-existing proficiency levels, first language and other socio-economic backgrounds. For this purpose, regression analysis and subgroup analysis of standardised mean difference by each variable were computed. The regression analysis yields a correlation coefficient or r, which describes the relationship between two or more variables (Gorard, 2012). A positive correlation means that the two variables increase or decrease in the same direction while a negative correlation means the two variables have different trends, i.e., one moving up and the other moving down (Mertens, 2010). More importantly, the regression analysis provides the r^2, which is another type of effect size indicating whether the change in one variable leads to the change in another variable. In the multiple regression analysis of this study, hierarchical models, where blocks of variables were entered in biographical stages, were analysed by the forward entry method. The regression models from this approach can indicate the strongest predictors for the outcome and can examine if the treatment is effective, net of background and prior attainment (Gorard, 2021).

To provide a triangulation of the findings from the regression analysis on the impact of the intervention, a subgroup analysis of the standardised mean difference of the intervention students' primary outcomes in relation to different key variables was also conducted. This investigation was operated by calculating effect sizes from the intervention group data to examine how the outcomes differ by different characteristics of the intervention participants, i.e., their gender, year of intake, hometown, school type, first language, parental income, parental education, admission English score and pre-test performance.

4.6.4 Exploring processes and perceptions: thematic analysis

The classroom observations and semi-structured interviews were both used as means of process evaluation. The questionnaire has both closed and open questions, and the latter can also inform about the processes of the trial. Concurrently, the data from classroom

observations, interviews, and questionnaires can provide insights into the participants' perception about the intervention, difficulties they may have experienced and how much they complied with the intervention guidelines. For the data from these instruments, a thematic analysis was conducted. Thematic analysis is a fundamental method for identifying, analysing and reporting patterns or themes from textual data (Braun & Clarke, 2006).

The intervention group post-questionnaire had four additional close-ended questions eliciting students' overall perceptions about the intervention. This data underwent simple descriptive statistical analysis to provide information on the overall students' perceptions of the metacognitive approach to teaching and learning. The observation notes, interview transcriptions and the open-ended answers from the questionnaire were coded and categorised to derive overarching themes of the data. The coding (i.e. attaching keywords to the data) and categorising (or drawing connections between codes into larger ideas) can encapsulate the underlying themes of the obtained data (Roulston, 2014). The results can inform about the fidelity, contamination and other validity issues in implementing the intervention and can illuminate the participants' perceptions and experiences using the intervention. Both descriptive statistics and thematic analyses can provide answers to Research Question 4 and reassure the validity of the results in the previous research questions.

4.6.5 Secondary data analysis

The analysis of large-scale data from a larger number of students of the same population could be used to inform about education policy and practice (Smith, 2012). Moreover, the relevant information obtained from wider non-trial participants can be used to compare with the equivalent data of the intervention students to shed more light on the impact of the intervention. In this study, secondary data which could be obtained came from the university English test scores of student cohorts 2019 and 2020, parental income data from cohorts 2019 and 2020 and the English admission scores from cohort 2019.

In the analysis, the percentages of each subcategory under each data set were calculated to exhibit the characteristics of the students in the target research context, namely, the southernmost areas of Thailand. The percentage of students who have passed the university tests can indicate how many students remain unqualified for the graduation criteria. The post-trial percentage of the trial participants passing or failing the university English test can be a delayed measurement of the impact of the intervention. Moreover, a comparison of the pass

rates between the full-cohort data and the intervention student data can further examine the impact of the intervention.

4.7 Chapter summary

The chapter presented fundamental elements in evidence-based research as background understanding to the randomised controlled trial, which was the design of this study. Before the main trial, evidence synthesis and a pilot study were conducted, which guided the implementation of the intervention in the main trial. The PMER model of metacognitive intervention was designed for the main trial with tertiary EFL learners in two of Thailand's southernmost universities. Six clusters were randomly assigned to the invention while another six clusters were in a business-as-usual group. Two non-randomised clusters were included as a comparator. Multiple data collection methods, namely, standardised English tests, metacognitive awareness questionnaire, class observations and interviews were used for both impact and process evaluations. Secondary data analysis was also included for richer evidence.

For the analysis, effect sizes were calculated to estimate the impact of the intervention on the primary and secondary outcomes and the NNTD was used as a sensitivity analysis to account for any attrition. Moreover, regression analysis was conducted to examine the association between the impact of the intervention and the key characteristics of the participants, namely, gender, hometown, first language background, high school type, year of intake, parental education, parental income, English admission scores and pre-test scores. In-depth data derived from the process evaluation methods underwent thematic analysis to explore the perceptions of the participants and provide more insights into the impact evaluation. Secondary data from larger samples who were not in the trial were included to portray a bigger picture of the consequences of the Exit English Examination policy on the students as well as the impact of the intervention.

CHAPTER FIVE

RESULTS: IMPACT ON PRIMARY OUTCOMES

This chapter presents analysis and results which answer the first research question on the impact of the intervention on the primary outcomes, followed by discussions of the findings. The chapter begins with a presentation of the data included in the final analysis. Next, impact evaluation is presented by descriptive statistics of the data and the analysis of standardised mean differences of the outcomes among the participants. After that, a sensitivity analysis of the results is conducted. Then, discussions of the findings are presented, followed by a conclusion of the chapter.

5.1 Data for analysis

The data used in the analysis was drawn from the trial participants and the secondary data obtained from the participating universities and relevant agencies. Figures 5.1 and 5.2 show the number of students at recruitment until the follow-up stage when the post-intervention measurement was completed. The flowcharts are based on the CONSORT guidelines by Moher et al. (2010).

Figure 5.1: Flowchart of number of cases at the start and the end of the trial: Randomised clusters
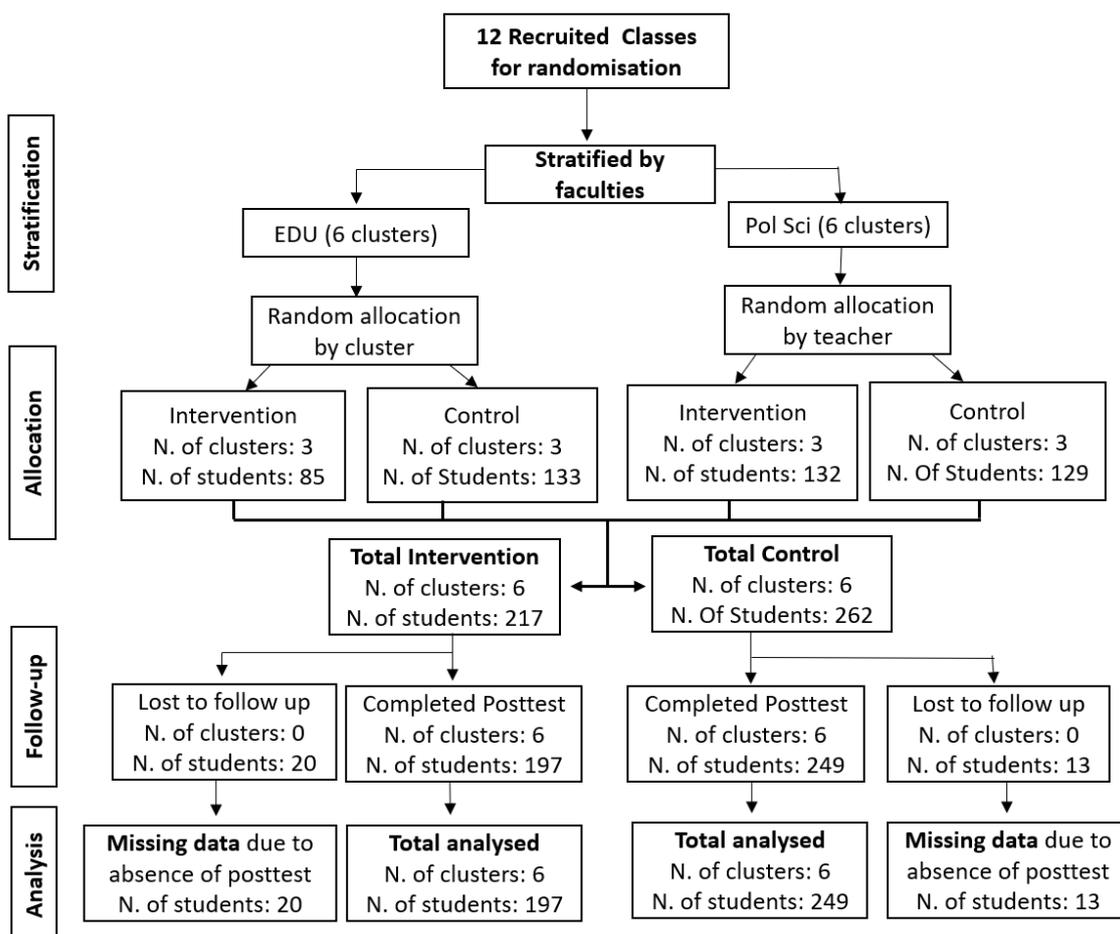
Figure 5.2: Flowchart of number of cases at the start and the end of the trial: Non-randomised clusters
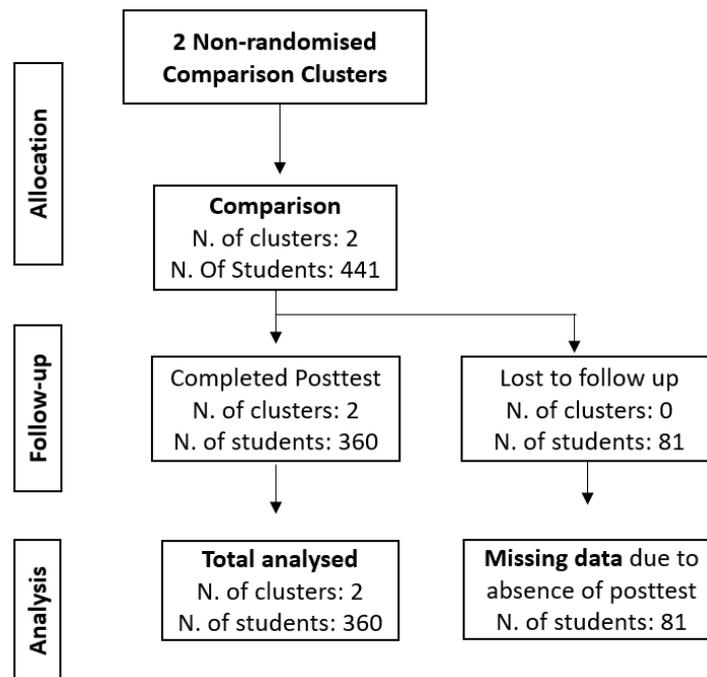
After the recruitment, 6 clusters with 217 students were allocated to the intervention group and 6 clusters with 262 students were in the control group. In addition, 2 non-randomised clusters with 441 students joined as a comparison group. After the trial, all the clusters were retained with some missing cases. Eventually, 197 students from the intervention group, 249 students from the control group and 360 comparison students completed their post-trial measurement of the primary outcome which could be used for the analysis. These students have their data ready for the primary outcome but their data for the secondary outcome or secondary data analysis may be missing. Details of these data will be discussed further when each of them is presented.

5.2 Impact evaluation of primary outcomes

The target learning outcome of this study is students' listening comprehension and overall English achievement. This type of data was collected from English tests which have a listening section as one of the main constructs among other English skills. Thus, the listening scores as well as the overall English scores were collected, analysed and reported in this section.

5.2.1 Descriptive statistics of primary outcomes

The mean scores and the standard deviation are presented here because they are common figures for presenting test results and are essential for further investigation. The results of overall English scores and listening scores from the pre-tests and post-tests of all recruited students are shown in Tables 5.1, 5.2, 5.3. The characteristics of the trial participants will be presented in Chapter Seven where the relationship between the participants' learning outcome and their background variables is discussed.

Table 5.1: Pre-test results of the trial participants: overall score & listening score

Groups	N of recruited students	Pretest Mean	Pretest SD	Pre Listening Mean	Pre Listening SD
Intervention	217	16.63	6.40	10.16	4.42
Control	262	16.06	6.92	9.52	4.62
Comparison	441	10.70	3.89	6.36	2.94

Table 5.1 illustrates the number of participants recruited to the trial and their pre-test results. Both overall pre-test scores and the scores of the listening section are presented. It is noticeable that the participants who were randomly assigned to the intervention and the control groups have only slightly different pre-test scores (16.63/16.06) and pre-listening scores (10.16/9.52) with the intervention group being slightly ahead. However, the non-randomised comparison group has noticeably lower scores than the previous two groups in both overall English and listening comprehension.

As not all students completed the post-test, some of these pre-test data were excluded from the analysis of the impact of the intervention. Only the data from students who completed both pre-test and post-test were included in the analysis.

Table 5.2: Overall scores of pre-tests and post-tests of trial student participants

Groups	N	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Gain Mean	Gain SD
Intervention	197	16.75	6.54	21.97	7.27	5.22	7.05
Control	249	16.16	6.89	15.18	8.35	-0.98	8.81
Comparison	360	10.66	3.95	11.59	4.16	0.93	4.84

Table 5.3: Listening scores of pre-tests and post-tests of trial student participants

Groups	N	Pre Listening Mean	Pre Listening SD	Post Listening Mean	Post Listening SD	Gain Mean	Gain SD
Intervention	197	10.20	4.48	12.59	4.65	2.39	4.69
Control	249	9.56	4.62	8.33	5.03	-1.22	5.80
Comparison	360	6.31	3.00	5.96	2.77	-0.36	3.49

The data in Tables 5.2 and 5.3 shows the test scores of the participants in the pre-test, the post-test and gain scores which were derived by subtracting the post-test scores by the pretest scores. From these tables, their pre-test scores which reflect the pre-existing English proficiency of the participants indicate that the English proficiency between the intervention and the control groups remain quite similar for both overall English (16.75/ 16.16) and listening comprehension (10.20/ 9.56) with the intervention group being already slightly ahead (ES +0.09 overall and ES+0.13 listening). This indicates some degree of comparability between the intervention and the control groups even after excluding the participants with no post-test data. Considering the data from the non-randomised comparison group, the scores of the remaining participants in both overall English and listening comprehension are clearly lower than the intervention group at the outset. With these pre-existing differences between groups, the gain scores, instead of the post-test scores, were used for calculating the headline effect sizes, which look at progress rather than absolute attainment.

Another important point noticeable from Table 5.3 is that the listening gain scores in the intervention group is positive while it is negative in both the control and the comparison groups. This indicates some degree of positive impact of the intervention. However, a precise estimate of the impact is not clear from looking at the mean score alone. How large the impact is can be illustrated by effect size estimation. This is presented in the following section.

5.2.2 Effect sizes on learning outcomes

In this section, effect sizes are calculated to estimate the impact of the intervention on the participants' listening comprehension and overall English achievement. This will provide answers to the first research question which enquires;

To what extent does the metacognitive instruction have an impact on the listening and overall English achievement of the tertiary EFL learners in the southern Thai context?

Effect sizes show how large the outcome difference is between the students after the intervention has been given. The effect sizes from the gain score mean will be presented to show the difference in improvement. The effect sizes reported in this section are Hedges' g because the numbers of participants in different groups are not the same (fuller explanation of Hedges' g was presented in section 4.6.1 of Chapter Four). The comparison of test scores was made between the intervention and control groups, the intervention and comparison groups, and the intervention and a combined group of control and comparison participants. The results are presented in the following tables.

Table 5.4: Effect size by overall English gain scores: intervention and control groups

Groups	N	Gain Mean	Gain SD	Effect Size
Intervention	197	5.22	7.05	0.77
Control	249	-0.98	8.81	

Table 5.5: Effect size by listening gain scores: intervention and control groups

Groups	N	Gain Listening Mean	Gain Listening SD	Effect Size
Intervention	197	2.39	4.69	0.68
Control	249	-1.22	5.80	

Tables 5.4 and 5.5 show that the intervention group has clearly greater improvement than the control group in both overall English competence and listening comprehension. The effect size of overall English is slightly bigger than the effect size of listening comprehension.

Table 5.6: Effect size by overall English gain scores: intervention and comparison groups

Groups	N	Gain Mean	Gain SD	Effect Size
Intervention	197	5.22	7.05	0.75
Comparison	360	0.93	4.84	

Table 5.7: Effect size by listening gain scores: intervention and comparison groups

Groups	N	Gain Listening Mean	Gain Listening SD	Effect Size
Intervention	197	2.39	4.69	0.69
Comparison	360	-0.36	3.49	

Similarly, Tables 5.6 and 5.7 suggest that the intervention group has clearly higher improvement compared to the non-randomised participants in both overall English competence and listening comprehension. The effect size of overall English is slightly bigger than the effect size of listening comprehension.

Table 5.8: Effect size by overall English gain scores: intervention group and the control plus comparison groups

Groups	N	Gain Mean	Gain SD	Effect Size
Intervention	197	5.22	7.05	0.74
Control and Comparison	609	0.15	6.81	

Table 5.9: Effect size by listening gain scores: intervention group and the control plus comparison groups

Groups	N	Gain Listening Mean	Gain Listening SD	Effect Size
Intervention	197	2.39	4.69	0.67
Control and Comparison	609	-0.71	4.59	

In the same vein, Tables 5.8 and 5.9 indicate the intervention group has clearly greater improvement even when compared to the combined group of control and comparison students. The effect sizes favour the intervention groups in both overall English competence and

listening comprehension. The effect size of overall English is slightly bigger than the effect size of listening comprehension.

As the pre-test scores of the intervention and the control students were slightly different, the calculation of the effect by post-test mean may provide interesting results for comparison. It should be noted that the scores of the intervention and the non-randomised comparison groups were substantially different at the pre-test. Comparing their post-test may not yield an accurate estimate of the effect. Therefore, the effect size by post-test will be presented only from the comparison of the intervention and the control groups.

Table 5.10: Effect size by post-test scores of overall English: intervention and control groups

Groups	N	Posttest Mean	Posttest SD	Effect Size
Intervention	197	21.97	7.27	0.86
Control	249	15.18	8.35	

Table 5.11: Effect size by post-test listening scores: intervention and control groups

Groups	N	Post Listening Mean	Post Listening SD	Effect Size
Intervention	197	12.59	4.65	0.87
Control	249	8.33	5.03	

Tables 5.10 and 5.11 show that the intervention group has substantially higher post-trial outcomes than the control group in both overall English competence and listening comprehension. Interestingly, the effect size of overall English is essentially similar to the effect size of listening comprehension, different from the trend in the gain score effect sizes.

As the researcher took part as a teacher in one of the intervention groups, it is possible that the effect sizes can be influenced by the experimenter effect. The effect size is usually bigger in the trials with the researcher as the deliverer of the intervention (Dignath & Büttner, 2008). To investigate this, two measures were applied. One is to exclude the data of the researcher-taught cluster from the intervention group to calculate the effect size. The other

measure is the comparison of the researcher-taught intervention cluster and other intervention clusters.

Table 5.12: Effect size of overall English gain scores: researcher IG and non-researcher IG

Groups	N	Gain Mean	Gain SD	Effect Size
Researcher-taught Intervention (IG)	26	5.65	5.06	0.07
Non-researcher Intervention (IG)	171	5.15	7.31	

Table 5.13: Effect size of listening gain scores: researcher IG and non-researcher IG

Groups	N	Gain Listening Mean	Gain Listening SD	Effect Size
Researcher-taught Intervention	26	2.38	3.50	-0.00
Non-researcher Intervention	171	2.39	4.86	

Tables 5.12 and 5.13 suggest that having the researcher deliver the intervention does not make meaningfully better gain scores, compared to other intervention groups taught by the usual teachers. No meaningful difference is found in both overall English and listening comprehension.

Table 5.14: Effect sizes comparing intervention group with and without the researcher-taught cluster and control group

Groups	N	Overall Gain Mean / SD	Listening Gain Mean / SD	Effect Sizes
IG + R	197	5.22 / 7.05	2.39 / 4.69	**Overall English** IG + R vs Control: 0.77
IG - R	171	5.15 / 7.31	2.39 / 4.86	IG - R vs Control: 0.74

Control	249	-0.98 / 8.81	-1.22 / 5.80	**Listening**
				IG + R vs Control: 0.68
				IG - R vs Control: 0.66

Notes: IG + R = all intervention clusters *including* the researcher-taught cluster
         IG – R = intervention clusters *without* the researcher-taught cluster

Table 5.15: Effect sizes comparing intervention group with and without the researcher-taught cluster and comparison group

Groups	N	Overall Gain Mean / SD	Listening Gain Mean / SD	Effect Sizes
IG + R	197	5.22 / 7.05	2.39 / 4.69	**Overall English**
				IG + R vs CPR: 0.75
IG - R	171	5.15 / 7.31	2.39 / 4.86	IG - R vs CPR: 0.73
				Listening
Comparison (CPR)	360	0.93 / 4.84	-0.36 / 3.49	IG + R vs CPR: 0.69
				IG - R vs CPR: 0.69

Notes: IG + R = all intervention clusters *including* the researcher-taught cluster
         IG – R = intervention clusters *without* the researcher-taught cluster

Table 5.16: Effect sizes comparing intervention group with and without the researcher-taught cluster and the combined group of control and comparison students

Groups	N	Overall Gain Mean / SD	Listening Gain Mean / SD	Effect Sizes
IG + R	197	5.22 / 7.05	2.39 / 4.69	**Overall English**
				IG + R vs CG + CPR: 0.74
IG - R	171	5.15 / 7.31	2.39 / 4.86	IG - R vs CG + CPR: 0.72
				Listening
CG + CPR	609	0.15 / 6.81	-0.71 / 4.59	IG + R vs CG + CPR: 0.67
				IG - R vs CG + CPR: 0.67

Notes: IG + R = all intervention clusters *including* the researcher-taught cluster
         IG – R = intervention clusters *without* the researcher-taught cluster
         CG + CPR = control and comparison groups

From tables 5.14, 5.15 and 5.16, no meaningful differences are observed in any effect sizes whether the intervention group includes the researcher-taught cluster or not. The differences between effect sizes with and without the researcher-taught intervention cluster are

less than 0.04 in all comparisons. Therefore, it could be concluded that there is no evidence of the experimenter effect in the effect sizes reported earlier in this section in both overall English competence and listening comprehension.

5.3 Sensitivity analysis

As in almost all empirical studies in social sciences, this trial has some missing data at the post-test. This section reports the missing cases in each group and a sensitivity analysis to account for such attrition. First, the number and percentage of the missing cases are presented below.

Table 5.17: Missing cases in all groups

Group	Pre-test (N)	Post-test (N)	Missing (N)	Missing (Percent)
Intervention (IG)	217	197	20	9.22
Control (CG)	262	249	13	4.96
Comparison (CPR)	441	360	81	18.37
Control and comparison (CG & CPR)	703	609	94	13.37
Overall of randomised participants (IG & CG)	479	446	33	6.89
Overall	920	806	114	12.39

Regarding the randomised participants, 20 cases were missing in the intervention group and 13 cases in the control group, equalling 9.22 per cent and 4.96 per cent respectively (see Table 5.17). The overall missing cases of the randomised groups are 33 (6.89%). Based on the EEF (2019)'s criteria for judging research findings, such percentages of attrition can be considerably low. For the comparison group, there is a higher attrition rate as 81 (18.37%) of the cases were missing at the post-test. Overall, 114 cases among all groups missed their post-test, making a 12.39% attrition rate which is not adversely high.

Being aware of and reporting the attrition rate is essential but may not be sufficient to assure the security of the reported results. The attrition can affect the results and lead to a biased conclusion even if they are proportionally spread between the two groups (Dong & Lipsey,

2011). Two attempts, i.e., analysis of data pattern and NNTD, were conducted to account for the missing cases. Such analyses are presented below.

5.3.1 Analysis of data pattern

The analysis of the data pattern can provide an indication of how the results may have been affected if the post-test scores of the missing cases could have been obtained. The findings from this analysis are as the following.

Table 5.18: Overall pre-test scores of dropout students and remaining students

Group	Pre-test of dropout students			Pre-test of remaining students		
	N	Mean	SD	N	Mean	SD
Intervention	20	15.45	4.89	197	16.75	6.54
Control	13	14.15	7.43	249	16.16	6.89

Table 5.18 shows that the pre-test scores of the intervention students who remained are slightly higher than the pre-test scores of the intervention students who dropped out (16.75/ 15.45 respectively). For the control group, the pre-test scores of the remaining students are also slightly higher than the pre-test scores of the dropout students (16.16/ 14.15 respectively). However, the difference in the control group is wider than in the intervention group. This suggests that if the dropout students in the control group had remained and taken the post-test, there is a chance that the post-test mean of the control group with those students could have been lowered. This may also similarly occur in the intervention group but the influence would have been lower than in the control group. Therefore, the missing post-test scores could not have subverted the results of the study regarding the overall English competence.

Table 5.19: Listening pre-test scores of dropout students and remaining students

Group	Pre-test of dropout students			Pre-test of follow-up students		
	N	Mean	SD	N	Mean	SD
Intervention	20	9.8	3.83	197	10.2	4.48
Control	13	8.77	4.85	249	9.56	4.62

The results from the listening section have a similar pattern to the overall English scores above with a smaller difference in the mean scores between the remaining and dropout students. Table 5.19 shows that the pre-test listening scores of the intervention students who remained

are slightly higher than the pre-test listening scores of the intervention students who dropped out (10.2/ 9.8 respectively). The pre-test scores of the remaining students in the control group are also slightly higher than in the dropout peers (9.56 / 8.77). However, the difference in the control group is wider than in the intervention group. Similar to the results from the overall English score above, if the dropout students in the control group had taken the post-test, there is a chance that the post-test listening mean of the control group with those students could have been lowered. Therefore, the missing cases could not have subverted the results of the study regarding the listening comprehension.

The analysis of the data pattern can demonstrate a tendency of the result if the post-test scores of the missing cases could have been obtained based on their pre-test performance. For a more robust method to account for the attrition, the number of counterfactual cases needed to disturb the finding (NNTD) was also conducted which will be presented next.

5.3.2 NNTD

Another measure of sensitivity analysis was conducted to account for the missing cases by calculating the number of counterfactual cases needed to disturb the finding (NNTD) as discussed in section 4.6.2 in Chapter Four. The NTTD is calculated by multiplying the effect sizes by the number of participants in the smaller arm. The NNTD for all main comparisons for effect sizes reported earlier is presented in Table 5.20.

Table 5.20: Number of counterfactual cases needed to disturb the finding (NNTD)

Outcome	Effect size	N of smaller cell	NNTD	N of missing cases
Overall English gain: IG /CG	0.77	197	152	33
Listening gain: IG /CG	0.68	197	134	33
Overall English gain: IG /CPR	0.75	197	148	101
Listening gain: IG /CPR	0.69	197	136	101
Overall English gain: IG /CG+CPR	0.74	197	146	114
Listening gain: IG /CG+CPR	0.67	197	132	114
Overall post-test IG /CG	0.86	197	169	33
Listening post-test IG /CG	0.87	197	171	33

The calculation of NNTD in Table 5.20 is based on the effect sizes reported in section 5.2.2 and the recruitment and attrition data in the Table 5.17. In the first row, the NNTD of 152 means that it would take 152 missing cases to eliminate the reported effect. In fact, there were only 33 missing cases across the intervention (IG) and control (CG) groups. Even if all the 33 missing cases had negative outcomes, that would not be sufficient to reverse the obtained results. Therefore, it is reasonably safe to accept that the results are not affected by the missing data and can be attributed to the intervention.

The effect sizes in section 5.2.2 were calculated for multiple comparison pairs. Thus, the calculation of NNTD was applied to all of the comparisons to check if the missing data can subvert the results. It was found that even in the case where there are the highest missing cases (i.e., in comparing the intervention group with the control plus comparison groups), the NNTD is still higher than the actual missing cases. Therefore, it is justifiable to conclude that all the results presented in section 5.2.2 are secure and not affected by the missing cases.

5.4 Discussion of the primary outcome results

The results presented in section 5.2 provide solid evidence for the effectiveness of the intervention on the primary outcomes, i.e., the listening comprehension and the overall English capacity. All effect sizes clearly favour the intervention group compared with the control group (CG), comparison group (CPR) and the combined group of control and comparison students (CG+CPR). Such effect sizes are secure from the missing cases. This evidence answers the first research question as the medium effect of the intervention is affirmed. Such findings are in accordance with the systematic review's results in Chapter Three which also reported the effectiveness of metacognitive interventions. However, the average effect size of this study is markedly smaller than the average effect size in the systematic review. The relatively smaller effect size on average could be influenced by the sufficiently large number of samples which can minimise the chance of inflating the result (Coe, 2002). Arguably, the result of this study is relatively secure because it is drawn from a fair design with a substantially larger sample size compared to the existing studies. These two elements, namely, research design and sample size, highly relate to the credibility of research findings (Gorard, 2014). Moreover, the attrition was reported and accounted for by a sensitivity analysis. This enhances the security of the reported results (Gorard et al., 2017). The results from the delayed measurement (to be presented in full in Chapter Nine) which was a standardised test administered by an independent third party re-affirms the impact of the intervention.

In this study, several attempts and precautions were taken to maintain the internal validity. Apart from the fair design, careful precautions need to be taken throughout the study to minimise bias (Torgerson & Torgerson, 2003). In this study, standardised measurements which could be marked objectively were used. The randomisation was done by an independent third party. Analysis of the results from the clusters with and without the researcher as a teacher were compared to investigate the experimenter's effect. These attempts were made to minimise bias and assure the validity of the results. Such validity precautions can enhance the trustworthiness of the findings (EEF, 2019). With the secure results being affirmed, it is reasonable to maintain that the metacognitive approach which guides the learners to the process of listening and promotes their self-regulation throughout such process has merit for the teaching and learning of English listening for EFL learners.

Several factors can contribute to the observed impact. The various types of strategies introduced in the intervention can be one of the contributing factors. The consciousness-raising of the available strategies to be used promotes positive self-perception (Paris & Winograd, 1990) and helps the learners to be more strategic in their problem-solving (Graham, 2007). The combination of listening, affective, cognitive and metacognitive strategies supports the learners to cope with the demands in the task more effectively, rather than teaching metacognitive strategies in isolation (Graham & Macaro, 2008). Meanwhile, the explicit instruction raises language learners' awareness of the available strategies which in turn enhances their self-efficacy and ability to regulate their learning (Vandergrift & Goh, 2012). Moreover, the self-reflection prompts and the self-questioning supported the learners' agency to activate prior knowledge, apply various strategies and regulate their emotional responses (Bown, 2009). In addition, the retrieval practices could strengthen the knowledge and skills to which the students have been introduced. The retrieval tasks improved learning and the ability to retain information (Agarwal et al., 2012) and provided corrective feedbacks to the learners (Butler et al., 2008).

The intervention in this study primarily aims at the listening comprehension of the learners of English as a foreign language. The results in this chapter show that the intervention group has improved listening outcomes as well as the overall English scores. This seems to signify a relationship between the improvement in listening comprehension and the improvement in overall English proficiency.

To investigate this relationship further, a simple regression analysis between the listening gain scores and the overall English gain scores of the intervention students was conducted and reported below. In the analysis, the listening gain score was used a predictor of the overall English score.

Table 5.21: Simple regression of gain English scores explained by listening gain scores

Predictors	R	R Square	Unstandardised Coefficients		Standardised beta Coefficients
			B (Constant)	B(Predictor)	
Listening gain	0.87	0.77	2.08	1.31	0.87

Table 5.21 shows a strong association between the listening outcome and the overall English outcome. The *r* correlation coefficient of 0.87 indicates a strong positive correlation between the two outcomes. Moreover, in the regression analysis where the listening outcome was used as a predictor, the r^2 of 0.77 shows a strong likelihood that the improvement in listening outcome leads to the improvement in the overall English outcome.

The triangulation of the results from both types of effect size i.e., standardised mean difference and association provides evidence that the listening ability is highly associated with overall language proficiency. As Feyten (1991) suggested, listening is the foundation of language acquisition, so improvement in listening skills can benefit the overall language learning. In the test used as a measurement tool in this study, similar to most English tests, reading comprehension is another important section. As both listening and reading largely deal with comprehension, they share several processes and skills (Pressley, 2002). This means that many skills used in listening can be transferable to reading and vice versa. Thus, improvement in listening can lead to better reading and the improved comprehension skills can be a key determinant in the improvement of the overall English scores.

Due to teacher shortage, the researcher unavoidably had to take part as a teacher in one intervention cluster. Some studies observe that the effect of an intervention is usually larger when the training is conducted by researchers than by regular teachers (Dignath & Buttner, 2008). Thus, the possible bias from the experimenter effect needs to be considered (Gorard, 2014). Tables 5.12 and 5.13 obviously demonstrate that the group taught by the researcher does not make a meaningfully higher gain than other intervention clusters taught by the usual teachers. This exhibits no sign of the experimenter's effect. In fact, the regular teachers know

more than the researcher about the characteristics and the learning styles of the learners. Thus, they are better aware of context-specific considerations to organise their lessons to suit the learners. If the teachers can appropriately apply the metacognition principles in their teaching with such awareness, it is unsurprising that they can make a higher impact on the students than an ad hoc outsider teacher.

5.5 Chapter summary

This chapter presents relevant data used in the analysis of the impact of the intervention on the participants' listening comprehension and overall English proficiency. From the results, the impact of the intervention on the language outcomes of the Thai EFL learners is affirmed. Such findings are secure from the missing cases and the experimenter's effect. Therefore, the metacognitive approach which guides students to various types of strategies in the listening process and promotes self-regulation is helpful for improving second language learning. The multi-strategy components, self-reflection and retrieval practices included in the intervention can be attributed to such impact. Moreover, the association between the improvement in listening comprehension and overall English ability were found. This underlines the significant role of comprehension skills in the development of learning English as foreign language.

As the impact of the intervention is evidenced in this chapter, some explanations of such impact have been given. The impact of the intervention on the primary outcomes will be investigated further in Chapter Seven where such impact is examined in relation to the various characteristics of the participants. Moreover, the process evaluation in Chapter Eight will illuminate clearer mechanisms of the benefits and some drawbacks which the intervention may have. The results from this chapter, together with more in-depth information from these two chapters will provide comprehensive findings about the role of metacognitive instruction in the tertiary Thai EFL context.

CHAPTER SIX

RESULTS: IMPACT ON SECONDARY OUTCOME

This chapter examines the impact of the intervention on the secondary outcome. It begins with a description of the data used in the analysis. Correlation analysis of question categories in the questionnaire is conducted to investigate the relationship between categories, followed by an analysis of standardised mean differences which indicate the impact. Next, a discussion of the findings is presented, followed by a chapter summary.

6.1 Data for impact evaluation of the secondary outcome

The secondary outcome of the study is the participants' metacognitive awareness in listening as enquired in the second research question;

To what extent does the metacognitive instruction have an impact on metacognitive awareness for listening of the tertiary EFL learners in southern Thai context?

The secondary outcome was measured by a questionnaire for the students to reflect on their thinking process and listening behaviours. The questionnaire was adapted from the MALQ by Vandergrift et al. (2006). The details of the questionnaire were discussed in Chapter Four and can also be found in Appendix 1.

6.1.1 Response rates of the metacognitive questionnaire

The descriptive statistics of the students' responses to the questionnaire are presented below.

Table 6.1: Number of students responding to the questionnaires

Group	Recruited students (N)	Students included in analysis (N)	Students responding to Pre-questionnaire (N)	Students responding to Pre/Post questionnaire (N)
Intervention (IG)	217	197	195	165
Control (CG)	262	249	212	205
Comparison (CPR)	441	360	247	210
Control and comparison (CG & CPR)	703	609	459	415
Overall of randomised participants (IG & CG)	479	446	407	370
Overall	920	806	654	580

Table 6.2: Response rates of students' questionnaires

Group	Pre-questionnaire per recruited students (%)	Pre-questionnaire per analysed students (%)	Pre/ Post questionnaire per analysed students (%)	Pre/ Post questionnaire per pre-responses (%)
Intervention (IG)	89.86	98.98	83.76	84.62
Control (CG)	80.92	85.14	82.33	96.70
Comparison (CPR)	56.01	68.61	58.33	85.02
Control and comparison (CG & CPR)	65.29	75.37	68.14	90.41
Overall of randomised participants (IG & CG)	84.97	91.26	82.96	90.91
Overall	71.09	81.14	71.96	88.69

As suggested in Tables 6.1 and 6.2, the response rates from the randomised participants are generally high. The response rates are higher than 80 per cent of all analysed students and of the number of students who answer the pre-questionnaire. Obviously lower response rates are observed in the comparison group which generally has low response rates except for the rate of students who responded to both pre and post questionnaires per pre-questionnaire rate (85.02%). As the participants were requested to complete the questionnaire online at their own time outside the classroom, this can be the reason for the low response rate. How the missing cases can affect the result will be discussed in section 6.3.

6.1.2 Descriptive statistics of questionnaire responses

The questionnaire items urge the respondents to reflect on their behaviours in listening. Each item elicits a response to a 1-5 opinion scale. The total 21 items represent the participants' positive behaviours and negative behaviours in listening. In Table 6.3 below, the first four categories are desirable behaviours which represent metacognitive awareness while the following three categories represent low metacognitive awareness in listening. Overall scores are provided at the end of all positive and negative categories. The data was only from the participants with both pre and post responses. This data will later be used for effect size calculation.

Table 6.3: Mean and SD of the pre and post questionnaire responses of the trial participants

Components of metacognitive awareness	Intervention (N:165)		Control (N: 205)		Comparison (N: 210)	
	Pre (M/SD)	Post (M/SD)	Pre (M/SD)	Post (M/SD)	Pre (M/SD)	Post (M/SD)
Planning and Evaluation	3.52 / 0.82	3.38 / 0.97	2.99 / 0.82	3.10 / 1.00	3.14 / 0.82	3.05 / 0.90
Directed Attention	3.52 / 0.87	3.36 / 1.02	3.04 / 0.87	3.05 / 1.02	3.16 / 0.89	3.04 / 0.95
Person Knowledge	2.88 / 0.96	2.99 / 0.97	2.92 / 0.95	2.85 / 1.05	3.10 / 0.98	3.02 / 0.98
Problem Solving	3.37 / 0.82	3.36 / 0.96	3.01 / 0.82	3.03 / 0.95	3.16 / 0.85	3.01 / 0.98
Overall Positive Behaviours	3.42 / 0.75	3.34 / 0.89	3.00 / 0.79	3.04 / 0.92	3.15 / 0.78	3.03 / 0.88
Giving up Attention	2.96 / 1.01	2.90 / 1.05	3.14 / 1.03	2.80 / 1.10	3.04 / 1.01	2.89 / 1.04
Perceived difficulty	3.22 / 0.98	3.15 / 1.10	3.19 / 0.88	3.04 / 1.11	3.17 / 1.00	3.06 / 1.08
Mental Translation	3.26 / 0.85	3.23 / 0.94	3.11 / 0.88	2.95 / 0.98	3.11 / 0.87	3.00 / 0.96
Overall Negative Behaviours	3.19 / 0.69	3.15 / 0.81	3.14 / 0.81	2.96 / 0.92	3.12 / 0.80	3.00 / 0.89

Table 6.3 suggests that on average, the scores rated by the intervention and comparison groups after the trial are similar to or lower than the pre-trial ratings in all categories except the intervention group's person knowledge. On the contrary, the control students have their post-trial rating higher in most categories. How meaningful these differences are will be explored further in the effect size calculation in section 6.3.

6.2 Correlation of responses in each category of metacognitive awareness

Before calculating the effect sizes to examine the impact of the intervention on the secondary outcome, an analysis of the correlation between each category of the questionnaire items is informative because it can portray an overview of the relationship between each category of metacognitive awareness. According to the interpretation guidelines from the authors of the MALQ, the first four categories in the Tables 6.4 and 6.5 are positive behaviours encouraged by the metacognitive instruction and the last three categories are the indicators of low metacognitive awareness in second language listening. Because of the low response rate in the comparison group, the results below were drawn only from the randomised participants.

Table 6.4: Correlation (*r*) of pre-questionnaire responses by question categories from randomised participants (N=407)

Category	Plan & Evaluation	Directed Attention	Person Knowledge	Problem Solving	Giving up Attention	Perceived difficulty	Translation
Plan & Evaluation	1	0.85	0.50	0.83	0.41	0.55	0.68
Directed Attention		1	0.46	0.81	0.35	0.59	0.66
Person Knowledge			1	0.59	0.41	0.36	0.60
Problem Solving				1	0.42	0.57	0.79
Giving up Attention					1	0.47	0.47
Perceived difficulty						1	0.54
Translation							1

Note: first row = pre- trial questionnaire result, second row = post-trial questionnaire result

Before the trial, the perceptions of all randomised participants on planning and evaluation highly correlate with their perceptions of their ability for directing their attention (*r* = 0.85) and problem-solving (*r* = 0.83). This means students who thought they did well in planning and evaluation also believed that they were good at directed attention and problem-solving. Moreover, students' perceptions of directed attention highly correlate with problem-solving (*r* = 0.81), suggesting students with the ability to direct their attention also perceived to have a high potential to problem-solve. Interestingly, students' perceptions of three of the positive behaviours, namely, planning and evaluation, directed attention, and problem-solving substantially correlate with one negative behaviour, mental translation (*r* = 0.68, 0.66 and 0.79 respectively). This suggests that students who perceived they did well in planning and evaluation, directed attention and problem-solving still relied heavily on translating words in their head. They do not see translation as a negative behaviour to avoid. Moreover, students' perceptions of planning and evaluation, directed attention, and problem-solving moderately correlate with their perception of the difficulty of English (*r* = 0.55, 0.59 and 0.57). This suggests that considerable number of students who believed they were good at these skills can still have a perception that English and English listening are difficult for them.

Table 6.5: Correlation (*r*) of post-questionnaire responses by question categories from randomised participants (N=370)

Category	Plan & Evaluation	Directed Attention	Person Knowledge	Problem Solving	Giving up Attention	Perceived difficulty	Translation
Plan & Evaluation	1	0.87	0.50	0.81	0.45	0.56	0.70
Directed Attention		1	0.59	0.85	0.50	0.59	0.78
Person Knowledge			1	0.59	0.53	0.48	0.60
Problem Solving				1	0.49	0.56	0.84
Giving up Attention					1	0.52	0.52
Perceived difficulty						1	0.59
Translation							1

After the trial, the correlations between the randomised students' perceptions of planning and evaluation and their perceptions of directed attention and problem-solving remain high (*r* = 0.87 and 0.81 respectively). Students who reported they did well in planning and evaluation believed that they were also good at directed attention and problem-solving. Furthermore, students' perceptions of directed attention highly correlate with problem solving (*r* = 0.85), similar to the pre-trial perceptions. Surprisingly, students' perceptions of three of the positive behaviours, namely, planning and evaluation, directed attention, and problem-solving has slightly higher correlations with mental translation (*r* = 0.70, 0.78 and 0.84 respectively). This suggests that whether or not the metacognitive strategies are introduced, students tend to rely highly on translating words in their head. Moreover, the moderate correlation between students' perceptions of the difficulty of English and their perceptions of planning and evaluation, directed attention, and problem-solving remain unchanged (*r* = 0.56, 0.59 and 0.57 respectively). This suggests that their perceptions that English and English listening are difficult has not notably changed even after the intervention.

6.3 Impact on the secondary outcome: Effect sizes of metacognitive awareness

To investigate the impact of the intervention on the students' metacognitive awareness, calculation of effect sizes could be computed to provide answers to the second research question. One advantage of reporting effect sizes in standardised terms is that the results are scale-free, so they can be compared across studies (Ellis, 2010). Students' responses to the

scale-based questionnaire can be used to compute means and standard deviations which can be utilised for effect size calculation (Coe, 2004).

The questionnaire items can be grouped into seven categories. The first four categories represent positive behaviours and the last three categories are the indicators of the low metacognitive awareness in second language listening. For the positive behaviours, the positive effect sizes would indicate the positive impact of the intervention on metacognitive awareness. On the contrary, for the negative behaviours, effect sizes in a negative manner would show positive impacts of the intervention while the positive effect sizes would suggest that the non-intervention groups reported better outcomes. The effect sizes from the students' questionnaire responses are reported below.

Table 6.6: Students' self-reported metacognitive awareness by gain rating of post/pre responses: intervention and control groups

Components of metacognitive awareness	Group	Pre/Post Questionnaire Difference					
		N	Mean	SD	Effect size	Missing	NNTD
Planning and Evaluation	Intervention	165	-0.14	1.09	**-0.23**	37	38
	Control	205	0.11	1.10			
Directed Attention	Intervention	165	-0.16	1.09	**-0.14**	37	23
	Control	205	0.01	1.21			
Person Knowledge	Intervention	165	0.11	1.20	**0.14**	37	23
	Control	205	-0.07	1.28			
Problem Solving	Intervention	165	-0.01	1.09	**-0.03**	37	5
	Control	205	0.02	1.09			
Overall Positive Behaviours	Intervention	165	-0.08	0.98	**-0.12**	37	20
	Control	205	0.04	1.02			
Giving up Attention	Intervention	165	-0.05	1.35	**0.20**	37	33
	Control	205	-0.33	1.43			
Perceived Difficulty	Intervention	165	-0.07	1.14	**0.07**	37	12
	Control	205	-0.15	1.20			
Mental Translation	Intervention	165	-0.03	1.15	**0.11**	37	18
	Control	205	-0.15	1.14			
Overall Negative Behaviours	Intervention	165	-0.04	0.94	**0.14**	37	23
	Control	205	-0.18	1.04			

According to Table 6.6, among the positive behaviours, the differences between the intervention and the control groups are quite ambiguous. On planning and evaluation, it is the control group which reported slightly higher improvement (effect size 0.23) and the result is quite secure. On person knowledge, the intervention group reported a slightly greater improvement (ES 0.14) but the NNTD is lower than the missing cases, signifying insecurity of the result. On overall positive behaviours, the results suggested that the control group reported a slightly higher improvement (ES 0.12) but that result is not secure.

Regarding the negative behaviours, all effect sizes seem to favour the control group. However, all effect sizes are small and none of the effect sizes is secure as the missing cases are higher than the NNTDs.

Overall, the results from Table 6.6 suggested that the control group reported a slightly greater improvement generally in both positive and negative behaviours. However, the results are not secure due to a large volume of missing cases which are higher than the NNTDs.

Table 6.7: Students' self-reported metacognitive awareness by gain rating of post/pre responses: intervention and comparison groups

Components of metacognitive awareness	Group	Pre/Post Questionnaire Difference					
		N	Mean	SD	Effect size	Missing	NNTD
Planning and Evaluation	Intervention	165	-0.14	1.09	**-0.05**	67	8
	Comparison	210	-0.09	0.99			
Directed Attention	Intervention	165	-0.16	1.09	**-0.04**	67	7
	Comparison	210	-0.11	1.06			
Person Knowledge	Intervention	165	0.11	1.20	**0.16**	67	26
	Comparison	210	-0.09	1.24			
Problem Solving	Intervention	165	-0.01	1.09	**0.14**	67	23
	Comparison	210	-0.15	1.00			
Overall Positive Behaviours	Intervention	165	-0.08	0.98	**0.05**	67	8
	Comparison	210	-0.12	0.93			
Giving up Attention	Intervention	165	-0.05	1.35	**0.08**	67	13
	Comparison	210	-0.16	1.27			
Perceived Difficulty	Intervention	165	-0.07	1.14	**0.03**	67	5
	Comparison	210	-0.10	1.25			
Mental Translation	Intervention	165	-0.03	1.15	**0.08**	67	13
	Comparison	210	-0.12	1.07			

Overall Negative Behaviours	Intervention	165	-0.04	0.94	**0.08**	67	13
	Comparison	210	-0.12	0.96			

For the comparison between the intervention and the comparison groups in Table 6.7, the comparison groups reported slightly higher improvements in planning and evaluation and directed attention but the effect sizes are trivial (ES 0.05 and 0.04 respectively) and not secure from the missing cases. Meanwhile, the intervention group reported a slightly greater improvement in person knowledge and problem-solving (ES 0.16 and 0.14) but such results are also not secure. The overall effect size of positive behaviours (ES 0.05) which very slightly favours the intervention group can be considered trivial and such result is not secure as the NNTD is far lower than the missing cases.

For negative behaviours, the control group reported a very slightly higher improvement in all categories. However, all effect sizes are very small and are not secure from the missing data.

Overall, an unclear picture is seen about the metacognitive awareness between the intervention and the comparison groups because the results on positive behaviours slightly favour the intervention group while the results from the negative behaviours slightly favour the control groups. Moreover, the effect sizes are very small and not secure from the missing data.

Table 6.8: Students' self-reported metacognitive awareness by gain rating of post/pre responses: intervention and control plus comparison groups

Components of metacognitive awareness	Group	Pre/Post Questionnaire Difference					
		N	Mean	SD	Effect size	Missing	NNTD
Planning and Evaluation	Intervention	165	-0.14	1.09	**-0.14**	74	23
	Control and comparison	415	0.01	1.05			
Directed Attention	Intervention	165	-0.16	1.09	**-0.09**	74	15
	Control and comparison	415	-0.05	1.14			
Person Knowledge	Intervention	165	0.11	1.20	**0.15**	74	25
	Control and comparison	415	-0.08	1.26			
Problem Solving	Intervention	165	-0.01	1.09	**0.05**	74	8
	Control and comparison	415	-0.07	1.05			

Overall Positive Behaviours	Intervention	165	-0.08	0.98	**-0.04**	74	7
	Control and comparison	415	-0.04	0.98			
Giving up Attention	Intervention	165	-0.05	1.35	**0.14**	74	23
	Control and comparison	415	-0.24	1.36			
Perceived Difficulty	Intervention	165	-0.07	1.14	**0.05**	74	8
	Control and comparison	415	-0.13	1.23			
Mental Translation	Intervention	165	-0.03	1.15	**0.10**	74	17
	Control and comparison	415	-0.13	1.10			
Overall Negative Behaviours	Intervention	165	-0.04	0.94	**0.11**	74	18
	Control and comparison	415	-0.15	1.00			

From the comparison between the intervention group and the combined group of control and comparison participants in Table 6.8, mixed results are found as the combined group reported slightly greater improvements in planning and evaluation (ES 0.14) and direct attention (ES 0.09) while the intervention group reported a slightly higher improvement in person knowledge (ES 0.15) and problem-solving (ES 0.05). Again, the effect sizes are very small and not secure from the missing data which are far higher than all NNTDs.

For negative behaviours, the combined group of control and comparison students reported a slightly bigger improvement throughout. However, all effect sizes are very small and not secure.

Overall, the improvement in the metacognitive awareness between the intervention and the non-intervention groups has a mixed picture in positive behaviours but the non-intervention groups reported better improvement in negative behaviours. However, such effect sizes are very small and not secure due to a large volume of missing cases which are largely higher than the NNTDs.

Different from test scores, the responses to the questionnaire are self-reported and the students judged their metacognitive awareness by themselves. The question of how reliable their judgements are may need to be considered. To investigate this, the Accuracy of Judgement Analysis was conducted and will be presented later in section 6.4. The results of this analysis suggest that the students' judgement of their metacognitive awareness from the post questionnaire rating is slightly more accurate than the gain rating of the post/pre questionnaire responses. With such results from the accuracy of judgement analysis, examining how different

the post-questionnaire responses are between the intervention and the non-intervention students is also useful. Therefore, the effect sizes of their post-trial questionnaire responses are also presented below.

Table 6.9: Students' self-reported metacognitive awareness in post-trial responses: intervention and control groups

Components of metacognitive awareness	Group	Pre/Post Questionnaire Difference					
		N	Mean	SD	Effect size	Missing	NNTD
Planning and Evaluation	Intervention	165	3.38	0.97	**0.29**	37	48
	Control	205	3.10	1.00			
Directed Attention	Intervention	165	3.36	1.02	**0.31**	37	51
	Control	205	3.05	1.02			
Person Knowledge	Intervention	165	2.99	0.97	**0.14**	37	23
	Control	205	2.85	1.05			
Problem Solving	Intervention	165	3.36	0.96	**0.34**	37	56
	Control	205	3.03	0.95			
Overall Positive Behaviours	Intervention	165	3.34	0.89	**0.33**	37	54
	Control	205	3.04	0.92			
Giving up Attention	Intervention	165	2.90	1.05	**0.09**	37	15
	Control	205	2.80	1.10			
Perceived Difficulty	Intervention	165	3.15	1.10	**0.09**	37	15
	Control	205	3.04	1.11			
Mental Translation	Intervention	165	3.23	0.94	**0.29**	37	48
	Control	205	2.95	0.98			
Overall Negative Behaviours	Intervention	165	3.15	0.81	**0.22**	37	36
	Control	205	2.96	0.92			

Table 6.9 indicates that the intervention group reported a slightly higher metacognitive awareness in the post-trial questionnaire in all positive behaviours with meaningful effect sizes overall (ES 0.33) and in planning and evaluation (ES 0.29), directed attention (ES 0.31) and problem-solving (ES 0.34). Such effect sizes are secure as the NNTDs are higher than the missing cases.

For negative behaviours, the control students reported a slightly higher improvement overall (ES 0.22). The control group reported greater improvement of mental translation (ES

0.29) but no meaningful differences in other negative behaviours. This effect size on the negative behaviour of metal translation is also secure, showing the control group reporting greater improvement on mental translation.

Overall, the intervention group has a greater improvement in positive behaviours while the control group has a better improvement in mental translation. These results are secure from the missing cases.

Table 6.10: Students' self-reported metacognitive awareness in post-trial responses: intervention and comparison groups

Components of metacognitive awareness	Group	Pre/Post Questionnaire Difference					
		N	Mean	SD	Effect size	Missing	NNTD
Planning and Evaluation	Intervention	165	3.38	0.97	**0.36**	67	59
	Comparison	210	3.05	0.90			
Directed Attention	Intervention	165	3.36	1.02	**0.32**	67	53
	Comparison	210	3.04	0.95			
Person Knowledge	Intervention	165	2.99	0.97	**-0.03**	67	5
	Comparison	210	3.02	0.98			
Problem Solving	Intervention	165	3.36	0.96	**0.36**	67	59
	Comparison	210	3.01	0.93			
Overall Positive Behaviours	Intervention	165	3.34	0.89	**0.35**	67	58
	Comparison	210	3.03	0.88			
Giving up Attention	Intervention	165	2.90	1.05	**0.02**	67	3
	Comparison	210	2.89	1.04			
Perceived Difficulty	Intervention	165	3.15	1.10	**0.08**	67	13
	Comparison	210	3.06	1.08			
Mental Translation	Intervention	165	3.23	0.94	**0.25**	67	41
	Comparison	210	3.00	0.96			
Overall Negative Behaviours	Intervention	165	3.15	0.81	**0.17**	67	28
	Comparison	210	3.00	0.89			

From Table 6.10, which compares the post-trial responses of the intervention group and the comparison group, it could be seen that the intervention students reported a slightly higher metacognitive awareness in overall positive behaviours (ES 0.35) and the positive behaviours they reported to do better are planning (ES 0.36), directed attention (ES 0.32) and problem-

solving (ES 0.36). Such effect sizes are meaningful but are not secure as the NNTDs are slightly lower than the missing cases.

For negative behaviours, the comparison group reported a slightly higher improvement overall (ES 0.17) and the category they reported meaningful improvement is mental translation (ES 0.25). However, such results are not secure as the NNTDs are higher than the missing cases.

Overall, the intervention group reported a slightly higher improvement in positive behaviours while the comparison group reported a better improvement in mental translation. However, these results are not secure from the missing cases.

Table 6.11: Students' self-reported metacognitive awareness in post-trial responses: intervention and control plus comparison groups

Components of metacognitive awareness	Group	Pre/Post Questionnaire Difference					
		N	Mean	SD	Effect size	Missing	NNTD
Planning and Evaluation	Intervention	165	3.38	0.97	**0.33**	74	54
	Control and comparison	415	3.07	0.95			
Directed Attention	Intervention	165	3.36	1.02	**0.32**	74	53
	Control and comparison	415	3.05	0.99			
Person Knowledge	Intervention	165	2.99	0.97	**0.05**	74	8
	Control and comparison	415	2.93	1.02			
Problem Solving	Intervention	165	3.36	0.96	**0.35**	74	58
	Control and comparison	415	3.02	0.94			
Overall Positive Behaviours	Intervention	165	3.34	0.89	**0.34**	74	56
	Control and comparison	415	3.04	0.89			
Giving up Attention	Intervention	165	2.90	1.05	**0.05**	74	8
	Control and comparison	415	2.85	1.07			
Perceived Difficulty	Intervention	165	3.15	1.10	**0.09**	74	15
	Control and comparison	415	3.05	1.09			
Mental Translation	Intervention	165	3.23	0.94	**0.27**	74	45
	Control and comparison	415	2.98	0.97			
Overall Negative Behaviours	Intervention	165	3.15	0.81	**0.19**	74	31
	Control and comparison	415	2.98	0.91			

Similarly, in Table 6.11, the intervention students reported a slightly higher metacognitive awareness than the non-intervention students in overall positive behaviours (ES 0.34) and the positive behaviours they reported to do better are planning (ES 0.33), directed attention (ES 0.32) and problem-solving (ES 0.35). However, the NNTDs are slightly lower than the missing cases, undermining the security of such results.

On negative behaviours, a small effect size (ES 0.19) favours the non-intervention groups in overall negative behaviours. The non-intervention students reported a slightly higher improvement of mental translation (ES 0.27). However, the NNTDs are largely lower than the missing cases, undermining the security of both effect sizes.

Overall, the intervention group reported a slightly greater improvement in positive behaviours while the non-intervention group reported a slightly higher improvement in mental translation. However, these results are not secure from the missing cases.

6.3.1 Missing cases

There were quite a lot of missing cases in the questionnaire responses especially among the non-randomised comparison group. This can affect the results about the effect of the intervention on metacognitive awareness. The NNTD and missing cases are reported in all tables in section 6.3. The results are undermined by the missing cases in almost all tables because the missing data is higher than the NNTD. Only Table 6.9 which compares the post-questionnaire responses between the intervention and control groups has the results which are secure and accounted for the missing cases.

6.4 Accuracy of judgement analysis

To measure the accuracy of students' self-reported metacognitive judgement, a relative accuracy measurement (more details were presented in section 2.1.5) as suggested in Schraw (2009) was conducted by investigating the correlation between the student's metacognitive judgement and the relevant performance, i.e., their English language outcomes. Tables 6.12 and 6.13 below report the correlations of students' English test performance and the gain rating of their metacognitive judgement before and after the trial (columns 3 and 4). It also examines the correlations between students' test performance and their post-trial judgement of metacognitive awareness (columns 5 and 6). In Table 6.12, students' overall gain scores and gain listening scores were used in the analysis. In Table 6.13, their post-test scores and post-listening scores were used.

Table 6.12: The correlation of students' metacognitive judgement and gain test scores

Group	Correlation pairs	Gain score and post/pre questionnaire	Gain Listening and post/pre questionnaire	Gain score and post questionnaire	Gain Listening and post questionnaire
IG	Test scores and positive behaviours	-0.01	-0.04	0.13	0.09
IG	Test scores and negative behaviours	-0.03	-0.02	0.14	0.09
CG	Test scores and positive behaviours	0.12	0.06	0.09	0.06
CG	Test scores and negative behaviours	0.10	0.07	0.09	0.07
CG+CPR	Test scores and positive behaviours	0.10	0.03	0.08	0.02
CG+CPR	Test scores and negative behaviours	0.09	0.05	0.07	0.02

Table 6.13: The correlation of students' metacognitive judgement and post-test scores

Group	Correlation pairs	Post-test and post/pre questionnaire	Post listening and post/pre questionnaire	Post-test and post questionnaire	Post listening and post questionnaire
IG	Test scores and positive behaviours	0.02	0.00	0.27	0.25
IG	Test scores and negative behaviours	-0.04	-0.06	0.11	0.09
CG	Test scores and positive behaviours	0.12	0.10	0.18	0.15
CG	Test scores and negative behaviours	-0.03	-0.02	0.08	0.06
CG+CPR	Test scores and positive behaviours	0.12	0.08	0.16	0.12
CG+CPR	Test scores and negative behaviours	-0.01	-0.03	0.07	0.03

From Table 6.12 which examines the correlations of students' metacognitive judgement in the questionnaire responses and their overall English gain scores and listening gain scores,

weak correlations are observed in all groups and all variable pairs. Such weak correlations suggest that the accuracy of their metacognitive judgement is low. This is not unusual for the non-intervention students as they were not explicitly exposed to the notion of metacognition. For the intervention students, the correlations are also low. However, while the accuracy of judgements of the non-intervention groups in the gain rating of the post/pre questionnaire responses and the post-questionnaire responses are not much different (below 0.04 in all variables), the intervention group has a slightly more accurate judgment in the post-questionnaire responses compared to the post/pre gain rating, with higher correlation coefficients of more than 0.1 in the positive behaviours.

Similar trends are seen in Table 6.13 which examines the correlations of students' metacognitive judgement and their post-test scores and post listening scores. Weak correlations are observed in all groups and all variable pairs, suggesting that the accuracy of their metacognitive judgement is low. However, while the accuracy of metacognitive judgements of the non-intervention groups in the gain rating of the post/pre questionnaire responses and the post-questionnaire responses are not much different (0.08 or below in all variables), the intervention group has a slightly more accurate judgment in the post-questionnaire responses compared to the post/pre gain rating, with higher correlation coefficients of more than 0.2 in the positive behaviours.

The data from Tables 6.12 and 6.13 suggest that students in all groups have considerably low accuracy of metacognitive judgements both before and after the trial, regardless of whether they have been exposed to the intervention. The data also suggest that the intervention students have a slightly higher competence to accurately judge their metacognitive awareness after the trial. With a higher accuracy of judgement in the post questionnaire, their post-questionnaire responses can be more accurately indicative of the students' metacognitive awareness than the gain rating of the responses to the post-trial and pre-trial questionnaires.

The low accuracy of judgement of the intervention group can be caused by several factors. First, it can be caused by the negative perceptions some students may have due to a number of unfamiliar terms used in the intervention, as expressed in the open-ended questionnaire responses. In addition, the lengthy and detailed processes may have caused some students to have a negative impression of the intervention, as one of the teachers suggested. These factors can confound their retrospective judgements, making them not proportionately

correspond to their performance. These judgements also contradict their overall satisfaction towards the metacognitive intervention and the open-ended questionnaire responses which report benefits and frequent use of strategies such as planning, self-motivation, attention focussing and schema activation. Further details of students' open-ended questionnaire responses and other means of process evaluation are presented in Chapter Eight.

6.5 Discussion of the findings on secondary outcomes

From the results comparing students' self-report of metacognitive awareness before and after the trial, the evidence of the impact of the intervention is rather ambiguous and unclear because the intervention group reported slightly better improvement in some categories while the non-intervention groups reported higher improvement in some categories and the results are mostly not secure from the missing cases. For instance, in Table 6.6 which compares responses from the intervention and the control groups, the results suggest that the intervention may slightly improve person knowledge of the intervention students but worsen planning and evaluation because it is the control group which reported higher improvement in this category. This negative result on planning and evaluation is also more secure than the positive result on person knowledge. Moreover, considering the responses on negative behaviours especially on mental translation, the non-intervention groups reported a slightly higher improvement throughout in Tables 6.6, 6.7 and 6.8 but the effect sizes are not secure from attrition in all of those tables.

The most secure result from the gain rating of metacognitive awareness is on the planning and evaluation between the intervention and the control groups in Table 6.6 which has a small negative effect. The results from negative behaviours, despite being not very secure, also show the non-intervention students reported a slightly higher improvement throughout. This suggests that the intervention may have a negative impact on students' metacognitive awareness. The results from the gain rating of responses can also be interpreted as no evidence of the impact of the intervention on metacognitive awareness because the results are quite mixed with the intervention group reported a slightly higher improvement than the non-intervention groups in some aspects and vice versa and the effect sizes are generally insecure. In both cases, no secure evidence indicates a positive effect of the intervention on metacognitive awareness.

With such results from the gain rating of questionnaire responses, the logic model presented in figure 4.9 in Chapter Four should be revisited. In the model, students' developed

metacognitive awareness is a distinctive feature which will lead to the improvement in the primary and secondary outcomes. However, according to the gain rating results, the intervention does not lead to improvement in metacognitive awareness. Thus, the metacognitive awareness cannot be the explanatory factor for the change in the post-intervention outcomes.

On this issue, there may be a number of possible reasons. First, it could be that it was the different types of strategies presented in the intervention, rather than the metacognitive awareness, which led to a clear improvement in the language outcomes of the intervention students while their metacognitive awareness remains ambiguous. So, it is possible to say that it is strategy-based instruction, not metacognitive awareness, which leads to improvement in the primary outcomes. Because developing language competence requires different types of strategies, both linguistic and non-linguistics (O'Malley & Chamot, 1990), explicit instruction of various strategies to students, especially the less proficient ones, can help them develop language competence (Griffiths, 2008).

Another reason for the non-effective result can be due to the underlying subtlety in the notions of awareness and competence. These two concepts can create hierarchical stages of competence because one can be unaware of their incompetence at the beginning, become aware of their incompetence and competence later when they develop and can be unaware of their competence when they have gained mastery of something or some skills (Hansen, 2012). From this model of competence hierarchy, unawareness can occur in both situations where a learner has or does not have competence in something. While the questionnaire in this study sought to assess students' metacognitive awareness, it is possible that some students might have metacognitive competence but are not aware of it while some other students may not be aware that they do not really have metacognitive competence. Perhaps, it might be the metacognitive competence which should be the focus in the secondary outcome, rather than metacognitive awareness. From the students' responses to the open-ended questionnaire, they reported using some metacognitive strategies. Moreover, their answers to retrieval practice tasks in each session were mostly accurate. This indicates that some students may have some degree of metacognitive competence but they may not necessarily be aware of such competence. Thus, the data from the questionnaire which elicited students' judgement of their own awareness may have caused such ambiguous results.

Such complexity in assessing metacognitive awareness leads to the next question of how reliable the students' judgement of their own awareness is. Therefore, the accuracy of judgement analysis was conducted in this chapter. From this analysis, students' metacognitive awareness from the post-trial responses seems to be slightly more accurate for several reasons. First, the results from the post-trial responses are in line with the results from the correlation analysis of the question categories reported in section 6.2. The correlation analysis of different question types in Table 6.5 indicates that students' responses on planning and evaluation, directed attention and problem-solving are highly correlated, corresponding to the post-trial questionnaire outcomes which observe meaningful effect sizes on these three question categories. Second, the analysis of the accuracy of judgement presented in section 6.4 suggests that the post-trial responses of the intervention group are slightly more accurate than the gain-rating of the post/pre questionnaire responses. Moreover, when taking the missing questionnaire responses into account, only the results from the post-trial responses reported in Table 6.9 remain secure. Therefore, only the results from this table can provide secure evidence of the effect of the intervention on the metacognitive awareness in listening.

With the above reasons, the results from the post-questionnaire responses can also be considered. The results suggested that the intervention has a slight effect on improving positive metacognitive behaviours but does not have any impact on reducing negative listening behaviours such as mental translation which was reported to be used slightly more frequently by the intervention group even after the intervention. Contrary to the findings in many studies which report improvement in metacognitive awareness (e.g., Chou, 2017; Tavakoli & Koosha, 2016), the impact of the intervention on metacognitive awareness in this study has mixed results. Several relevant factors can explain these mixed findings. Regarding the small effect on positive behaviours, the intervention, despite highlighting reflections of strategy use, did not adequately draw students' reflection to the strategy employment and how each strategy could affect comprehension. As Graham et al. (2011) suggested, explicit strategy instruction which involves students' reflection on the relationship between strategy use and comprehension is more likely to develop higher awareness in their strategy use. Another reason could be that the majority of the trial participants have low proficiency in English (more details on participants' characteristics are presented in Chapter Seven). Low proficient learners tend to have low awareness of their own thinking and listening processes (Goh & Hu, 2014).

Regarding the intervention students' higher application of mental translation even after the intervention, the following reasons can be relevant. The first reason could be that the

intervention attempted to promote learners' metacognitive awareness but did not adequately raise the awareness of the negative behaviours or attitudes which the students should minimise. In the questionnaires, the questions were also not specified whether a behaviour enquired in the questions is considered positive or negative in order to avoid leading the responses by the 'positive' and 'negative' or 'good' and 'bad' terms. In addition, the grammar-translation method, which has been a prominent teaching approach in the Thai EFL context (Wongdaeng & Hajihama, 2018), can make students become accustomed to translation and do not see the word-by-word translation as a negative behaviour to avoid in listening. Relying too much on translation can hinder listening fluency (Vandergrift & Goh, 2012).

Another lesson learnt from the study is the complication in assessing metacognitive awareness. As a tool in this study as well as in most studies on the topic, the self-report questionnaire was used. The main disadvantage of the self-report tool is the uncertainty of how accurate such responses are (Veenman et al., 2006). Thus, if a self-report assessment is used, the accuracy of judgement analysis should be incorporated (Schraw, 2009). In addition, the unclarity from the questionnaire results also underlines the significance of process evaluation which provides a description of how the participants actually performed and reacted. Furthermore, the low correlation between students' metacognitive awareness judgement and their performance suggests that some, if not most, students may still not be profoundly aware of their metacognition. For these reasons, multi-method assessments with both online and offline measurements may be used in future studies (Saraç & KaraKelle, 2012; Raoofi et al., 2013). This would allow cross-method comparisons as a triangulation and reliability check of the results. Additionally, vignettes or short stories or scenarios may be included to elicit another set of responses related to students' self-report awareness (Siddiqui et al., 2017). These techniques can elicit relevant responses from participants which allow independent examiners or the researchers to make judgements whether a particular response represents students' metacognitive awareness or competence. This can provide more reliable results if independent examiners or transparent pre-specified criteria are used.

Another observation from the questionnaire administration is that how the rating options are scaled needs careful consideration. In this study, the questionnaire items on students' metacognitive awareness were accompanied with a rating scale of 1 to 5 in each item, where one means 'strongly disagree' and five is 'strongly agree' (see Appendix 1). Even though the numerical values of 1-5 seem to align with the descriptions and can be used to calculate means and standard deviations and further used for effect size calculation, such pre-specified

descriptions may be limited to describe the averaged results of the obtainable responses (Bishop & Herron, 2015). For example, it would be outlandish to describe an average result of 4.5 as 'agree and a half'. Alternatively, using a scale with dichotomous descriptors at the two ends and a wide range of options in between may be more representative of the numerical values in the range. For instance, a scale of 0-10 where zero means 'not at all true' and ten means 'completely true' may be used (Siddiqui et al., 2017). This may better represent the numerical value of the number and may provide a more evident outcome of the impact of the intervention on students' metacognitive awareness.

6.6 Chapter summary

From the findings in this chapter, rather mixed results were found on the impact of the intervention on metacognitive awareness. The results from post-intervention responses indicate a small effect of the intervention on improving planning and evaluation, directed attention, problem-solving but the intervention does not minimise the negative behaviours such as mental translation and the perceived difficulty of English and English listening. From the analysis of correlations between metacognitive judgement and their performance, the participants, either receiving or not receiving the intervention, have low accuracy of their metacognitive judgment, indicating the low awareness of metacognition.

As the results from the close-ended questionnaire appear to have a mixed outcome, the results from a process evaluation can provide additional information about the role of metacognitive intervention and the participants' metacognitive awareness. The process evaluation in Chapter Eight which explores the participant's perceptions deeper from the open-ended questionnaire items, interviews from different perspectives and the researcher's observation can provide a fuller understanding of the benefits and drawbacks of the intervention which may be more illuminating about their metacognitive awareness.

CHAPTER SEVEN

RESULTS: PRIMARY OUTCOMES AND LEARNERS' BACKGROUNDS

This chapter investigates the relationship between the primary outcomes and the background characteristics of the participants. Initially, descriptive statistics of the participants' characteristics are presented. Then, such data was used in the regression analysis in section 7.2. The same data from the intervention students was used in the subgroup analysis in section 7.3. Then, discussions of the findings were presented, followed by a summary at the end of the chapter.

7.1 Data for analysing mediating factors for the impact of the intervention

This section examines the relationship between the impact of the intervention and some key characteristics of the participants. The data about the students' characteristics were elicited via the first part of the questionnaire (appendix 1). The results from analyses in this chapter can provide answers to the third research question;

> In what manner is the impact of metacognitive instruction associated with differences
> in biographical variables such as gender, first language background, socio-economic
> backgrounds and pre-existing proficiency levels?

The impact investigated in this chapter is only on the primary outcomes which are the students' listening comprehension and overall English competence. The impact on the secondary outcome presented in Chapter Six is somewhat ambiguous and, therefore, is not included in this analysis. The results from this chapter can provide clearer explanations for the impact of the intervention discussed in Chapter Five.

For triangulation of results, the analysis was conducted via regression analysis to examine the association between the students' characteristics and the primary outcome and the subgroup analysis of the standardised mean difference of the intervention students' primary outcomes in relation to those characteristics. The primary outcomes used in both analyses are students' overall English gain scores and their listening gain scores.

Before the regression analysis, the descriptive statistics of each variable used in the regression models are presented. Then, the results of simple regression analysis and multiple regression analysis are displayed, followed by the subgroup analysis of the standardised mean differences by each variable.

7.1.1. Descriptive statistics of participants' characteristics

The characteristics of the participants can be explanatory variables for the change in the outcomes, i.e., their overall English gain scores and listening gain scores. The key variables considered in the following analyses are gender, year of intake, hometown, school type, first language, parent income, parental education, admission English scores and pre-test scores. Each characteristic of the participants in each group is presented in the following tables. Some characteristics of the comparison group, i.e., year of intake, parent income and admission scores are either not available or not relevant, thus, cannot be included.

Table 7.1: Descriptive statistics of the outcomes by gender

Group	Gender	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	Female	145	4.89	6.76	2.29	4.65
	Male	52	6.13	7.79	2.67	4.85
Control	Female	176	-0.65	8.55	-1.09	5.43
	Male	73	-1.77	9.41	-1.56	6.63
Comparison	Female	285	0.75	4.77	-0.47	3.42
	Male	75	1.63	5.09	0.07	3.73

From Table 7.1, female students are the large majority in all condition types. Generally, male students in the intervention and comparison groups have slightly higher overall gain score and listening gain score than their female peers. Conversely, the female students performed better in the control group.

Table 7.2: Descriptive statistics of the outcomes by year of intake

Group	Year	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	2019	113	6.95	7.28	3.12	4.93
	2020	83	2.96	6.02	1.48	4.14
Control	2019	126	0.75	9.89	-0.29	6.37
	2020	121	-2.99	6.81	-2.33	4.80

Table 7.2 shows the students in 2019 intake who are in the second year of their study in both randomised groups have better improvement in both overall gain and listening gain scores. This suggests that having been in a higher year of study is beneficial for improvement.

Table 7.3: Descriptive statistics of the outcomes by hometown

Group	Hometown	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	Deep South	102	5.72	7.04	2.67	4.85
	Others	60	3.87	6.75	1.58	4.39
Control	Deep South	133	-1.62	8.24	-1.68	5.50
	Others	69	0.46	9.19	-0.45	6.10
Comparison	Deep South	231	0.64	4.81	-0.39	3.57
	Others	13	0.54	5.16	-0.92	3.75

Table 7.3 suggests that the students whose hometowns are from the southernmost provinces which are also all the 'deep South' are the majority in all condition types but the ratio in the comparison group is multiple-time different. The students in the deep south in the intervention and the comparison groups perform slightly better than those from other areas in both overall gain and listening gain scores but an opposite trend is seen in the control group.

Table 7.4: Descriptive statistics of the outcomes by school type

Group	School Type	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	Religious	82	5.79	7.52	2.54	5.40
	Non-Religious	73	4.11	6.25	1.88	3.88
Control	Religious	110	-0.28	9.71	-0.85	6.26
	Non-Religious	90	-1.91	6.91	-1.93	4.90
Comparison	Religious	152	0.53	4.61	-0.47	3.47
	Non-Religious	81	0.73	5.23	-0.38	3.83

According to Table 7.4, the students in the intervention and the control groups who are from religious schools have a slightly higher improvement in overall and listening gains. Meanwhile, the comparison group which has students from religious schools almost two times higher than non-religious schools sees the non-religious group having slightly higher gains.

Table 7.5: Descriptive statistics of the outcomes by first language

Group	First Language	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	Dialect Malay	70	6.21	7.07	2.81	4.97
	Southern Thai	65	4.54	6.88	1.75	4.52
	Standard Thai	27	3.15	6.60	2.07	4.41
Control	Dialect Malay	96	-1.53	8.62	-1.48	5.86
	Southern Thai	77	0.31	9.06	-0.60	5.93
	Standard Thai	29	-2.10	7.15	-2.31	4.58
Comparison	Dialect Malay	198	0.52	4.70	-0.53	3.53
	Southern Thai	24	1.63	3.59	0.17	3.02
	Standard Thai	21	1.10	6.71	0.29	4.60

As shown in Table 7.5, dialectal Malay is spoken by most participants, followed by southern Thai. A mixed picture is seen in terms of which first language mostly relates to the improved outcomes. Malay speakers in the intervention group seem to do slightly better than southern Thai and standard Thai speakers. Meanwhile, for the control group, southern Thai speakers perform slightly better than Malay speakers who, in turn, have slightly higher gain than standard Thai speakers. The results from the comparison group slightly favour students with non-Malay mother tongue but the numbers of Malay speaking and non-Malay speaking groups are multiple-time different.

Table 7.6: Descriptive statistics of the outcomes by parent income

Group	Parent Income	N	Overall Gain Mean	SD	Listening Gian Mean	SD
Intervention	Below minimum	90	5.49	6.74	2.48	4.67
	Meet minimum	105	5.26	7.08	2.51	4.53
Control	Below minimum	119	-1.92	8.33	-1.90	5.30
	Meet minimum	126	-0.32	9.03	-0.73	6.10

In Table 7.6, the minimum wage of approximately 9,000 baht per month, as announced in the government gazette in 2020, was used as a benchmark for dividing students into groups based on their parent income. The table suggests that the outcomes are mixed in the intervention group while the control students whose parents' income meet the minimum wage have higher outcomes in both overall and listening gains.

Table 7.7: Descriptive statistics of the outcomes by parental education

Group	First Gen Higher. Ed.	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	Yes	76	6.47	6.79	3.29	4.40
	No	85	3.76	6.97	1.33	4.81
Control	Yes	96	-0.84	8.99	-1.21	5.83
	No	105	-0.95	8.33	-1.30	5.68
Comparison	Yes	147	1.07	4.51	-0.11	3.46
	No	92	-0.01	5.30	-0.76	3.75

According to Table 7.7, the number of students who are and are not the first generation to attend higher education is not notably different. Both the overall gain score and the listening gain score of students who are FGHE are higher than the non-FGHE in all groups. This means being the first generation admitted to higher education does not hinder their development.

Table 7.8: Descriptive statistics of the outcomes by admission scores

Group	Admission Score range	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	Lower	45	6.27	7.78	2.29	5.32
	Higher	65	7.68	6.34	3.86	4.24
Control	Lower	80	0.60	10.22	-0.45	6.78
	Higher	44	1.05	9.46	0.00	5.71

The division of students by their admission scores was based on the criterion which the university where the data was collected from used as a minimum score of English for recruiting students to different programmes at the campus where the data was collected. The minimum requirement ranges from 15, 20, 25, 30 and 35 from the total score of 80. It was decided that 25, the middle score of that range, would be used as a benchmark for dividing the participants. Thus, students whose scores were above 25 were included in the 'higher' group. From Table 7.8, the students in the intervention and control groups with higher admission scores have slightly better gain scores than their peers with lower admission scores in both overall and listening gains.

Table 7.9: Descriptive statistics of the outcomes by pre-test

Group	Pretest	N	Overall Gain Mean	SD	Listening Gain Mean	SD
Intervention	Lower	115	7.68	7.09	3.97	4.90
	Higher	82	1.77	5.38	0.43	3.57
Control	Lower	155	1.82	7.22	0.88	4.57
	Higher	94	-5.60	9.27	-4.41	6.02
Comparison	Lower	338	1.28	4.54	5.61	1.97
	Higher	22	-4.36	6.28	13.77	1.56

The division of students' pre-test performance was based on a 50% cut-off point. The pre-test has a total score of 35. To pass a 50% cut-off, students need to get at least 18. To be included in the higher proficiency group, it was decided that they should get at least one score above the cut-off point. Thus, students whose pre-test scores were above 18 were included in the 'higher' group. From Table 7.9, the numbers of students with lower pre-test scores and pre-listening scores are higher in all condition groups but the number of low pre-test scorers in the non-randomised group is multiple-time higher than the number of students having high pre-test scores. This indicates that the majority of the trial participants have low English proficiency. Regarding the outcomes, the students with lower pre-test scores have a

substantially greater improvement in overall gain and listening gain in the intervention and control groups.

It is noticeable that the number of obtained data (N) in each variable in the previous nine tables is not the same. This is because not all students provided relevant data or some of them were not accessible. The missing data of each variable is presented in Table 7.10 below.

Table 7.10: Missing information

Variables	Intervention			Control			Comparison		
	Valid	Missing	Total	Valid	Missing	Total	Valid	Missing	Total
Gender	197	0	197	249	0	249	360	0	360
Year of intake	196	1	197	247	2	249	N/A	N/A	N/A
Hometown	162	35	197	202	47	249	244	116	360
School type	155	42	197	200	49	249	233	127	360
First language	162	35	197	202	47	249	243	117	360
Parent income	195	2	197	245	4	249	N/A	N/A	N/A
First gen H. Ed.	161	36	197	201	48	249	239	121	360
Admission score*	110	97	197	124	125	249	N/A	N/A	N/A
Pretest	197	0	197	249	0	249	360	0	360

Notes: -The N/A in the comparison group means such data could not be obtained.
-The admission score data is with * because only students in cohort 2019 have data on admission scores accessible. This means such data from cohort 2020 were missing in both the intervention and control groups.

7.2 Regression analysis of primary outcomes and background variables

The regression analysis below was conducted to examine the association between the primary outcome and participants' backgrounds. Such background variables are gender, hometown, school type, first language, year of intake, parent income, parent education, admission scores, pre-test scores and group allocation. In the first stage, a simple regression analysis of the primary outcomes, i.e., the overall gain scores and listening gain scores was conducted to examine how much such outcomes are explained by each background variable. Then, such variables were put into a multiple regression analysis. Hierarchical models where blocks of variables were entered in biographical stages were operated. This means the variables were entered into the model in steps. First, all background variables were put into the first block using forward entry. Then, the English admission score was added, followed by their pre-test score. The last step was the addition of the group allocation variable. Coefficients of the multiple regression models in this section can be found in Appendix 11. The regression models from this approach can illuminate the strongest predictors explaining the change in the outcomes and can examine if the treatment is effective, net of backgrounds and prior attainment (Gorard, 2021).

186

The first set of analyses applied the data from the intervention and control groups. The results are shown in Tables 7.11 – 7.14 below.

Table 7.11: Simple regression of overall gain scores explained by background variables: intervention and control groups

Predictors	r^2	Unstandardised Coefficients		Standardised beta Coefficients
		B (Constant)	B(Predictor)	
Gender	0.00	2.18	-0.33	-0.02
Year of Intake	0.06	7.93	-4.25	-0.25
Hometown	0.00	1.08	0.48	0.03
School Type	0.00	2.04	-0.62	-0.04
First Language	0.00	2.00	-0.39	-0.03
Parent Income	0.00	0.33	0.94	0.05
First Gen H. Ed.	0.01	3.62	-1.23	-0.07
Admission Score	0.02	0.28	2.36	0.13
Pretest	0.16	10.17	-0.51	-0.40
Group	0.13	11.42	-6.20	-0.36

Table 7.12: Multiple regression models predicting overall gain scores: intervention and control group

Models and Predictors	r^2	r^2 Change
1. Admission Score	0.02	0.02
2. Admission Score, Pretest	0.36	0.34
3. Admission Score, Pretest, Group	0.39	0.03

Table 7.11 can provide an indication of how much each variable associates with the overall gain score of the participants in the intervention and control groups. The r^2 of each variable in Table 7.11 suggests that most variables have a very low association with the overall gain score as they are 0.06 or below. Thus, they cannot be considered influential factors explaining the outcome of the intervention and control participants. Two variables, i.e., the pre-test score and the group to which students were allocated, can be explanatory factors for the overall gain score but the association is rather slim (r^2 = 0.16 and 0.13 respectively). As the low achieving group was coded as 1 and the higher achieving group as 2, the beta coefficient of

pre-test which is in the minus manner suggests that it is the lower proficiency students who make greater improvement in the gain outcome. Regarding group allocation, the intervention group was coded as 1 and the control group as 2. Therefore, the minus beta coefficient suggests that it was the intervention group who made greater improvement in overall English achievement.

In Table 7.2 which reports the multiple regression analysis of the same variables via the hierarchical forward method, it could be noticed that the pre-test score is the strongest predictor of the outcome of the intervention and control participants with an incremental r^2 of 0.34, as suggested in model two. In the third model where the group variable was added on top of all other variables including the pre-test score, the r^2 increases very slightly from 0.36 to 0.39. However, the increase in r^2 in model three provides evidence of the positive effect of the intervention as it was the last variable added to the model and could still make an incremental r^2, net and above all other variables.

Table 7.13: Simple regression of listening gain scores explained by background variables: intervention and control groups

Predictors	r^2	Unstandardised Coefficients		Standardised beta Coefficients
		B (Constant)	B(Predictor)	
Gender	0.00	0.68	-0.24	-0.02
Year of Intake	0.04	3.43	-2.11	-0.19
Hometown	0.00	-0.09	0.29	0.03
School Type	0.00	0.42	-0.23	-0.02
First Language	0.00	0.43	-0.17	-0.02
Parent Income	0.00	-0.77	0.76	0.07
First Gen H. Ed.	0.01	1.68	-0.90	-0.08
Admission Score	0.02	-1.23	1.77	0.15
Pre-Listening	0.23	6.13	-0.59	-0.47
Group	0.10	6.01	-3.62	-0.32

Table 7.14: Multiple regression model predicting listening gain scores: intervention and control group

Models and Predictors	r^2	r^2 Change
1. Admission Score	0.03	0.03
2. Admission Score, Pre-Listening	0.42	0.39
3. Admission Score, Pre-Listening, Group	0.44	0.02

Similar to the results explaining overall gain scores presented earlier, the r^2 of each variable in Table 7.13 suggests that most variables have a very low association with the listening gain score of the intervention and control participants as they are all below 0.05. Only the pre-listening scores and the group allocation can potentially be explanatory factors for the listening gain score but such association is slim (r^2 0.23 and 0.10 respectively). The beta coefficient of pre-listening which is in the minus manner suggests that it is the lower proficiency students who make greater improvement in the listening outcome. For group allocation where the intervention group was coded as 1 and the control group as 2, the minus beta coefficient suggests that it was the intervention group who made greater improvement in listening.

In the multiple regression analysis in Table 7.14, it is noticeable that that the pre-listening score is the strongest predictor of the listening outcome of the intervention and control participants with an incremental r^2 of 0.39, as suggested in model two. In the third model where the group variable was added on top of all other variables including the pre-listening score, the r^2 increases very slightly from 0.41 to 0.43. Such increment provides evidence of the positive effect of the intervention on listening, net and above all other variables.

The next set of analyses was drawn from the data of the intervention and comparison groups. As some variables could not be obtained from the comparison group, the included background variables used as predictors are gender, hometown, school type, first language, parental education, pre-test and group allocation. The regression results of intervention and comparison groups are in Tables 7.15 - 7.18.

Table 7.15: Simple regression of overall gain scores explained by background variables: intervention and comparison groups

Predictors	r^2	Unstandardised Coefficients		Standardised beta Coefficients
		B (Constant)	B(Predictor)	
Gender	0.01	0.82	1.33	0.09
Hometown	0.00	1.28	0.96	0.06
School Type	0.00	2.33	0.15	0.01
First Language	0.00	2.22	0.36	0.04
First Gen H. Ed.	0.01	3.92	-1.03	-0.08
Pretest	0.05	5.50	-0.24	-0.23
Group	0.11	9.46	-4.26	-0.34

Table 7.16: Multiple regression model predicting overall gain scores: intervention and comparison groups

Models and Predictors	r^2	r^2 Change
1. Gender	0.02	0.02
2. Gender, Pretest	0.08	0.06
3. Gender, Pretest, Group	0.36	0.28

The r^2 of each variable in Table 7.15 suggests that most variables have a very low association with the overall gain score of the intervention and comparison students as they are all 0.5 or below. The only variable which has some meaningful association with the outcome is the group allocation but such association is small (r^2 0.11). The beta coefficient of group allocation which is in the minus manner suggests that it is the intervention students who make higher improvement in the gain outcome.

Unlike the data drawn from intervention and control students, the multiple regression analysis in Table 7.16 shows that the group allocation is the strongest predictor of the outcome of the intervention and comparison participants, net and above other biographical variables with an incremental r^2 of 0.28, as suggested in model three. This affirms the impact of the intervention. The pre-test scores added in the second model makes only a slight change from the previous step.

Table 7.17: Simple regression of listening gain scores explained by background variables: intervention and comparison groups

Predictors	r^2	Unstandardised Coefficients		Standardised beta Coefficients
		B (Constant)	B(Predictor)	
Gender	0.00	-0.21	0.67	0.07
Hometown	0.00	0.07	0.50	0.05
School Type	0.00	0.55	0.22	0.03
First Language	0.01	0.43	0.47	0.08
First Gen H. Ed.	0.01	1.81	-0.77	-0.09
Pre-Listening	0.11	3.21	-0.34	-0.33
Group	0.10	5.11	-2.73	-0.32

Table 7.18: Multiple regression model predicting listening gain scores: intervention and comparison groups

Models and Predictors	r^2	r^2 Change
1. Pre-Listening	0.13	0.13
2. Pre-Listening, Group	0.39	0.27

Different from the results using the overall gain score, Table 7.17 suggests that two variables, i.e., pre-listening scores and group allocation have some meaningful association with the listening outcome of the intervention and comparison students (r^2 0.11 and 0.10 respectively). For pre-listening, the beta coefficient in the minus manner suggests that it is the lower proficiency students who make greater improvement in the gain listening outcome. For group allocation, the minus beta coefficients suggest that the intervention students make higher improvement in the listening outcomes. All other variables have no meaningful association with the listening gain scores as they are all below 0.05 and cannot explain the change in the outcome.

The multiple regression analysis in Table 7.18 indicates that the pre-listening score in model one is a meaningful predictor of the outcome which makes an r^2 change of 0.13 but the group allocation in model two is the strongest predictor, above other biographical variables

with an incremental r^2 of 0.27. This, again, provides support for the association of the intervention with the improvement in the listening outcome.

To look into how the primary outcomes may be explained by the background variables, running an analysis with only the data from the intervention group is also helpful as this group received the intervention. Thus, in the regression analysis of the data from the intervention students, the same variables except the group variable were entered into the model. The results are presented in Tables 7.19-7.22.

Table 7.19: Simple regression of gain scores and variables of intervention group

Predictors	r^2	Unstandardised Coefficients		Standardised beta Coefficients
		B (Constant)	B(Predictor)	
Gender	0.01	3.64	1.24	0.08
Year of Intake	0.08	10.93	-3.98	-0.28
Hometown	0.02	7.56	-1.85	-0.13
School Type	0.01	5.60	-1.06	-0.09
First Language	0.03	6.18	-1.56	-0.16
Parent Income	0.00	5.72	-0.23	-0.02
First Gen H. Ed.	0.04	9.09	-2.64	-0.19
Admission Score	0.01	4.86	1.41	0.10
Pretest	0.18	12.98	-0.46	-0.43

Table 7.20: Multiple regression model predicting gain scores of intervention group

Models and Predictors	r^2	r^2 Change
1. Admission Score	0.01	0.01
2. Admission Score, Pretest	0.22	0.21

The r^2 of each variable in Table 7.19 suggests that the only variable which has some meaningful association with the overall gain score of the intervention students is the pre-test scores with a small r^2 of 0.18. The beta coefficient of pre-test which is in the minus manner suggests that it is the lower proficiency students who make higher improvement in the gain outcome. The year of intake with an r^2 of 0.8 may also not be negligible. As cohort 2019 was coded as 1 and cohort 2020 as 2, the minus beta coefficient suggests that it is the cohort 2019

students who have been in a higher year in higher education who made greater improvement. All other variables have no meaningful association with the outcome as they are all below 0.05 and cannot explain the outcome of the intervention group. Likewise, the multiple regression analysis in Table 7.20 suggests that the pre-test score is the strongest predictor of the gain score outcome, net and above other variables with an incremental r^2 of 0.21.

Table 7.21: Simple regression of gain listening scores and variables of intervention group

Predictors	r^2	Unstandardised Coefficients		Standardised beta Coefficients
		B (Constant)	B(Predictor)	
Gender	0.00	1.91	0.38	0.04
Year of Intake	0.03	4.77	-1.64	-0.17
Hometown	0.01	3.75	-1.08	-0.11
School Type	0.00	2.42	-0.30	-0.04
First Language	0.01	2.64	-0.51	-0.08
Parent Income	0.00	2.44	0.04	0.00
First Gen H. Ed.	0.04	5.03	-1.80	-0.20
Admission Score	0.03	0.72	1.57	0.16
Pre-Listening	0.24	7.58	-0.51	-0.49

Table 7.22: Multiple regression model predicting gain listening scores of intervention group

Models and Predictors	r^2	r^2 Change
1 First Gen Hi-Ed	0.05	0.05
2. First Gen Hi-Ed, Admission Score	0.10	0.04
3. First Gen Hi-Ed, Admission Score, Pre-Listening	0.37	0.27

Similar to the results from the overall gain score, Table 7.21 suggests that the only variable which has some meaningful association with the overall listening score of the intervention students is the pre-listening score with a small r^2 of 0.24. The beta coefficient in the minus manner suggests that it is the lower proficiency students who make higher improvement in the listening outcome. All other variables have no meaningful association with the outcome as they are all below 0.05 and cannot explain the outcome. The multiple regression

analysis in Table 7.22 suggests that the pre-listening score is the strongest predictor of the outcome, net and above other biographical variables with an incremental r^2 of 0.27.

7.3 Subgroup analysis: impact of the intervention by background variables

As the impact of the intervention on the primary outcome is confirmed in Chapter Five and reaffirmed in section 7.2, this section will investigate the intervention students' characteristics which might receive varied impacts from the intervention. This investigation was conducted by calculating effect sizes from the intervention group data to examine how the outcomes are different in relation to each characteristic of the participants. The outcomes in the following analyses are the overall English gain score and the listening gain score of the intervention group (IG). The results are presented in the following tables.

Table 7.23: Standardised difference of the outcome by gender

Outcomes	Gender	N	Mean	SD	Effect Size
IG overall gain	Female	145	4.89	6.76	-0.18
	Male	52	6.13	7.79	
IG listening gain	Female	145	2.29	4.65	-0.08
	Male	52	2.67	4.85	

As indicated in Table 7.23, male students have slightly higher gain scores in both overall gain scores and listening gain scores but both effect sizes are small. Therefore, there is no obvious evidence that the intervention is especially helpful for a particular sex. Overall, male students may have a slightly higher improvement but that difference is not substantial enough to conclude that the intervention is more suitable for male students than the female peers.

Table 7.24: Standardised difference of the outcome by year of intake

Outcomes	Year	N	Mean	SD	Effect Size
IG overall gain	2019	113	6.95	7.28	0.59
	2020	83	2.96	6.02	
IG listening gain	2019	113	3.12	4.93	0.35
	2020	83	1.48	4.14	

As Table 7.24 shows, the students in 2019 intake who are in the second year of their study have greater improvement with a medium effect size in the overall gain scores and a small effect size in the listening gain scores. These meaningful differences suggest that having been in higher education longer helps students apply metacognitive intervention better.

Table 7.25: Standardised difference of the outcome by hometown

Outcomes	Hometown	N	Mean	SD	Effect Size
IG overall gain	Deep South	102	5.72	7.04	0.27
	Others	60	3.87	6.75	
IG listening gain	Deep South	102	2.67	4.85	0.23
	Others	60	1.58	4.39	

Table 7.25 suggests the students whose hometowns are from the southernmost provinces or the 'deep south' have a slightly greater improvement than the students from other areas with small effect sizes of 0.27 in the overall gain scores and 0.23 in the listening gain scores. Intriguingly, this finding is contrary to the common belief that having student input from the deep south areas would hinder learning development. Despite small effect sizes favouring students from the deep south, the findings indicate that students' background from the southernmost areas does not necessarily have a negative effect on their academic development.

Table 7.26: Standardised difference of the outcome by school type

Group	School Type	N	Mean	SD	Effect Size
IG overall gain	Religious	82	5.79	7.52	0.24
	Non-Religious	73	4.11	6.25	
IG listening gain	Religious	82	2.54	5.40	0.14
	Non-Religious	73	1.88	3.88	

According to Table 7.26, the effect sizes slightly favour the students with backgrounds from religious schools in both overall and listening gain scores. However, the effect sizes are small. Despite small effect sizes, the results seem to contradict the common assumption that student backgrounds from religious schools will negatively affect their learning development. On the contrary, the students from religious schools can improve their learning outcomes if appropriate support is given.

Table 7.27: Standardised difference of the outcome by first language

Group	First Language	N	Mean	SD	Effect Size
IG overall gain	Dialect Malay	70	6.21	7.07	0.24
	Southern Thai	65	4.54	6.88	
IG listening gain	Dialect Malay	70	2.81	4.97	0.22
	Southern Thai	65	1.75	4.52	
IG overall gain	Dialect Malay	70	6.21	7.07	0.44
	Standard Thai	27	3.15	6.60	
IG listening gain	Dialect Malay	70	2.81	4.97	0.15
	Standard Thai	27	2.07	4.41	

Among the three mother tongues, namely, dialectal Malay, southern Thai and standard Thai, Table 7.27 suggests that the students with dialectal Malay as the first language have a slightly better improvement than the speakers of other mother tongues. However, the effect sizes are generally small except for the comparison of the overall gain score between Malay speakers and standard Thai speakers (ES 0.44). Overall, the Malay speakers tend to perform better than the speakers of other mother tongues in the overall English but the differences in listening comprehension are less evident.

Table 7.28: Standardised difference of the outcome by parent income

Group	Parent Income	N	Mean	SD	Effect Size
IG overall gain	Below minimum	90	5.49	6.74	0.03
	Meet minimum	105	5.26	7.08	
IG listening gain	Below minimum	90	2.48	4.67	-0.01
	Meet minimum	105	2.51	4.53	

Table 7.28 suggests that there is no meaningful difference in the improvement between students with lower and higher income backgrounds in both the overall gain score and the listening gain score. While this income variable is usually related to learning outcomes, it does not make a meaningful difference in the improvement of the students who received the intervention.

Table 7.29: Standardised difference of the outcome by parental education

Group	First Gen H. Ed.	N	Mean	SD	Effect Size
IG overall gain	Yes	76	6.47	6.79	0.39
	No	85	3.76	6.97	
IG listening gain	Yes	76	3.29	4.40	0.42
	No	85	1.33	4.81	

According to Table 7.29, being the first generation to attend higher education does not seem to negatively affect the improvement of the participants in the intervention group, contrary to the general assumption about this variable. Conversely, the effect sizes favour the first-generation-in-higher-education students in both the overall gain score (0.39) and the gain listening score (0.42). It could be said that students who are the first generation in higher education have more notable improvement and seem to gain slightly higher benefits from the intervention. Therefore, being the first generation in higher education does not negatively affect the development of the intervention students.

Table 7.30: Standardised difference of the outcome by admission scores

Group	Score range	N	Mean	SD	Effect Size
IG overall gain	Lower	45	6.27	7.78	-0.20
	Higher	65	7.68	6.34	
IG listening gain	Lower	45	2.29	5.32	-0.33
	Higher	65	3.86	4.24	

From Table 7.30, the intervention students with higher admission scores have a slightly better improvement than their peers with lower admission scores. However, such effect sizes are small. Overall, students with higher admission scores may have a slightly higher improvement but the differences in improvement are not remarkably clear.

Table 7.31: Standardised difference of the outcome by pre-test

Group	Pretest	N	Mean	SD	Effect Size
IG overall gain	Lower	115	7.68	7.09	0.92
	Higher	82	1.77	5.38	
IG listening gain	Lower	109	3.97	4.90	0.81
	Higher	88	0.43	3.57	

From Table 7.31, the students with lower pre-test scores have largely more remarkable improvement in both overall gain scores and listening gain scores. Such large differences indicate that the students with lower pre-existing proficiency have gained substantially higher benefits from the intervention. In other words, the intervention can be more beneficial to students with lower proficiency levels.

7.4 Discussion of the association of the primary outcomes and background variables

The results from the regression analysis in section 7.2 indicate that the change in the outcomes, namely, overall English gain score and listening gain score were not explained by most of the biographical backgrounds of the students. Some existing evidence indicates links between learning outcomes and student socioeconomic backgrounds such as parent incomes and parent education (e.g., Early et al., 2020; Gorard, 2018). Indeed, such links are also observed in the participants of this study. For instance, a large proportion of students are from the family with low-income backgrounds while a large proportion of students have low pre-existing English proficiency, as suggested by the pre-trial indicators. The findings from secondary data analysis to be presented in Chapter Nine also reveal the links between students' academic performance and their socioeconomic backgrounds. However, in terms of development, the results from this chapter suggest that these background variables are not obstacles for foreign language development as they had no meaningful association with the outcome changes. This accords with some recent evidence which suggests that cognitive ability has a stronger association with students' attainment than socioeconomic factors such as social class and income (O' Connell & Marks 2022). Therefore, a well-designed educational intervention which seeks to enhance students' cognitive ability can lead to improved attainment and overcome the corollary of socioeconomic factors.

Considering the research context, most participants are from the restive southernmost areas of Thailand, have their mother tongue other than the national language and went to Islamic schools in their secondary education. These have created a stereotyped assumption that students with these backgrounds will face difficulties to develop academically. Such assumption is not supported by the regression analysis results in the section 7.2 because no meaningful association was found between these features and the improvement in the language outcomes. Moreover, the results from the subgroup analysis of standardised difference in section 7.3. show that students with these features, i.e., being from the deep South, coming from religious schools, having a Malay dialect as a mother tongue, perform slightly better but

the effect sizes are small. Indeed, these characteristics should not be used to stereotype students because such assumptions can create a stereotype threat which can negatively affect the students' learning and motivation (Steele & Aronson, 1995). Students should be encouraged to strive to develop regardless of their socioeconomic and biographical backgrounds.

The results from the multiple regression analysis suggest that the pre-test score is the strongest predictor of the outcomes. In fact, this pre-test variable should not have been included as a predictor in the regression analysis. Because the target outcomes in the analysis were gain scores and gain listening scores, using pre-test scores as a predictor is not appropriate as they are already part of the target outcomes. This mistake should be eliminated in future research. However, whether with or without the pre-test scores as a predictor, the results still suggest that group allocation is a meaningful predictor of the primary outcomes, net and above other background variables.

The results from the subgroup analysis of standardised mean difference reveal that the intervention students who had lower pre-test scores improved largely higher than their more proficient peers. Such results showing more benefits of the intervention for the less proficient students for improving their language outcomes are in line with many other studies such as Cross (2011); Goh & Taib (2006); Graham & Macaro (2008) and Yang (2013). One of the trial teachers who taught in three intervention clusters also expressed that the two groups with generally lower pre-existing English proficiency appeared to show more remarkable improvement than the one with higher pre-existing proficiency. Why the less proficient (LP) students have higher improvement than their more proficient (MP) peers is an interesting question.

There are several reasons to explain why the intervention has a stronger impact on LP learners in the intervention group. One reason can be due to the components of metacognitive instruction which encompass various types of strategies beyond language skills (Chamot, 2008). The teaching and activities which are based on a metacognitive approach are usually aimed at encouraging learners to apply the strategies which effective learners report using. Metacognitive instruction unveils the processes and strategies effective listeners use to the LP learners who usually lack such awareness (White, 2008). By making the learning strategies explicit, the LP learners are facilitated and equipped with strategies to plan, monitor and evaluate their learning (Goh, 2008). In line with Vandergrift and Tafaghodtari (2010), an important difference between more-skilled and less-skilled listeners lies in the metacognitive

awareness. Therefore, guidance on metacognitive processes from teachers and more-able peers can provide more support for the less proficient listeners.

Another reasonable explanation is related to the self-affirmation and self-efficacy promoted in the approach. The less-able learners tend to believe that they are not good at a certain domain, which is English in this case, and such low self-belief hampers their development (Bandura, 1999). Some evidence suggests that the teaching of thinking skills can reduce the stereotype threat the low achievers have (Good et al., 2003). Metacognitive instruction can raise the low-proficient language learners' awareness of the available strategies which, in turn, enhances their self-efficacy to regulate their learning (Graham & Macaro, 2008). Even though the correlation analysis of different question categories suggests that many students may still perceive that English and English listening is difficult, the awareness-raising of the different strategies they can apply, including linguistic, cognitive and affective strategies, can help them maintain their attention and motivation to work with a task. Raising students' awareness that they can better cope with the task when they attempt relevant strategies and practise more can boost the self-belief that they are able to improve (Bandura, 1999).

The inclusion of retrieval practices in the PMER model used in the trial could be another contributing factor for the impact especially on the LP learners. At the end of each metacognitive session, a retrieval practice was given and the answers and reasons behind them were discussed. The retrieval practices which provide corrective feedbacks can enhance the retention ability of low-confidence learners (Butler et al., 2008). The immediate recall activities can facilitate the transfer of the newly learnt knowledge and skills to the long-term memory (Oxford, 2017). As Agarwal et al. (2017) found from their data, retrieval practices can be a particularly helpful learning strategy for low achieving students.

Another possible explanation for the higher gain score made by the LP students is the ceiling effect. In the APA's dictionary of psychology, the ceiling effect is defined as "a situation in which the majority of values obtained for a variable approach the upper limit of the scale used in its measurement" (VandenBos, 2015 p.166). It is reasonable to argue that students who scored higher in the pre-test will have less room for increasing their gain scores because their pre-test score are closer to the maximum score in the measurement. In other words, the more-proficient students may be affected by the ceiling effect which makes their average gain score lower than the lower-proficient students.

To investigate this effect, two methods were used. First, the scores of more-proficient (MP) students were explored to see how many students got the full score or close to full score (35 or 34) at the post-tests. This can show the tendency of how much they may be capped by the maximum score. It was found that no students got the full score or the score next to full score of the overall post-test and only six students (6.8%) among the MP students got the full score or the score next to the full score (19 or 20) of the post-test listening section. As a second method, the pre-test scores of the MP students were explored to compare with their post-test. It was found that in the overall pre-test scores, none of the MP students scored in the top range (34 or 35) nor in the next range (32 or 33) and only four students (5%) scored in the following range (30 or 31). This indicates that it is unlikely the MP students' overall gain scores are affected by the ceiling effect. On the listening section, six students (6.8%) scored in the top range (19 or 20) and 12 students (13.64%) scored in the following range (17 or 18). These students may be at risk of being affected by the ceiling effect to improve their gain listening score. However, these students count as 19 per cent of all MP students. Thus, their high pre-test listening scores may have a small effect on the average score of the whole MP group and cannot explain their lower improvement compared to the LP group.

From the discussion in the previous paragraph, it could be concluded that the ceiling effect should not be a valid explanation for the greater improvement by the LP group on the overall gain scores. It may have played a small role in the improvement of listening scores but more valid reasons are self-affirmation and self-efficacy principles in the intervention, retrieval activities and the intervention components with guidance and strategies drawn from effective learners which make the LP group gain more benefits from the metacognitive intervention. As the intervention group was predominated by low proficient students, as shown in Table 7.9, the benefits of the intervention which these students have obtained can contribute to the overall impact of the intervention on the primary outcomes.

7.5 Chapter summary

In brief, this section provides evidence that only the pre-test score can be the moderating factor for the primary outcomes while other biographical variables have no meaningful association with the outcomes. The subgroup analysis further explains that the intervention has more benefits for the lower-proficient students and students who have been in higher education longer. The reasons for more benefits on the LPS lie in the awareness-raising of available

strategies, the retrieval practices which support knowledge transfer and the promotion of self-affirmation and self-efficacy.

CHAPTER EIGHT

RESULTS: PROCESS EVALUATION

This chapter describes the process evaluation procedures and analysis from various instruments. These include classroom observation, descriptive statistics of additional post-trial questionnaire items and responses to the open-ended questionnaire which were coded and categorised into benefits, drawbacks and suggestions of the intervention. In addition, deeper investigations via interviews with students and teachers are presented. Then, the results are discussed and summarised at the end of the chapter.

8.1 Introduction

A process evaluation was included in the main study for several reasons. First, it was embedded to investigate the fidelity to the research design (Siddiqui et al., 2018). This process investigation can explore whether the outcomes might have been caused by other confounding factors. In addition, because the main deliverers of the intervention were the regular teachers, the fidelity to intervention protocol could also be investigated via this investigation. Moreover, the process evaluation can provide more in-depth data about the difficulty of implementation and the perceptions of the stakeholders. This can enhance the contextual understanding of the outcomes from the impact evaluation. The process evaluation can provide answers to the fourth research question:

4. What are the teachers' and students' perceptions towards the metacognitive instruction?

The process evaluation was conducted with the data drawn from classroom observation, additional close-ended items in the post-trial questionnaire, responses to the open-ended questionnaire and semi-structured interviews with teachers and volunteering intervention students. The classroom observation data was drawn from all condition types, i.e., intervention, control and comparison classes. The interviews were conducted with some intervention and control students and all teachers at the end of the trial. The other tools elicited responses from the intervention students only.

The process evaluation elicited both numerical and textual data. The numerical data from the responses to the additional close-ended questionnaire were tallied and used to compute mean and standard deviation which exhibit the participants' overall perception towards the intervention. The textual data from the classroom observation notes, interview transcripts and students' responses to the open-ended questionnaire were coded, counted for frequencies and categorised into main themes. They can shed light on the fidelity to the research design and the

intervention and explain the benefits, difficulties and drawbacks of the invention and potential implications for future implementations.

8.2 Classroom observation

The class observations took place in all condition groups with an aim to observe the teaching styles and material employment by each teacher and investigate the fidelity to the intervention, possible contamination and possible barriers or difficulties the teachers and the students encountered. The data from the observation protocol and observation notes are summarised below.

In the non-intervention clusters, two teachers were involved in the control group and one in the comparison group. The comparison class was observed once online because the teaching was arranged fully online at this university during the COVID 19 outbreak. The other two control classes were observed once each on-site. Each teacher had different teaching styles. One teacher emphasised translation of new vocabulary words; another teacher used a more communicative approach while the other teacher used a project-based approach. None of them, even the education teacher who taught both control and intervention clusters, discussed the ideas of metacognition in their control classes. The project-based teaching, if conducted appropriately, is supposed to promote metacognition and self-regulation but this concept was not explicitly addressed or made aware to the students. Therefore, there was no apparent sign of contamination in the control and comparison clusters.

In the intervention clusters, two teachers (excluding the researcher) participated. The observation took place once in one of each teacher' classes. Different aspects found during the observation are presented in Table 8.1.

Table 8.1: Observed features of the intervention classes

Features	Key observed elements
Class organisation	-The provided class materials were used in both classes. -Both classes were conducted in a well-organised manner and followed the sequence in the guideline. -The allotted time for the intervention seemed insufficient to cover the prepared materials extensively. Sometimes, the teacher had to rush summing up one part or lessen the discussion time to move to the next.

Fidelity	-There was a desirable degree of fidelity as both teachers tried to follow the given guideline and materials. However, due to time limitation, adjustments were needed occasionally.
Strategy Instruction	- Both teachers followed the suggested pedagogical sequences step by step and had frequent use of questions to elicit thinking and responses. -The Pol Sci teacher managed to find his own way to explain the concept in simple terms and gave relevant examples which were seen less in the other class. -Both teachers tried to link the strategies to the learning of other skills such as reading.
Interactions	-Teachers had good interactions with students and promoted group discussions among students. -With limited time, some discussions had to be abruptly brought to a summary to move on to the next stage. -In some online classes, the pair and small group discussions were quite limited.
Student reactions	-Most students well participated and responded to the prompts and tasks but some students seemed less responsive. - Some students were less confident at first but pair and small group discussions helped them to engage more.
Barriers, difficulties	- The provision of materials via an online channel may cause the class to run less smoothly as students needed to switch between different windows on their devices. -Group/pair discussion about the use of strategies was helpful but was time-consuming. The limited time and the contents waiting to be covered restricts the time for discussion. -Some terms may seem unfamiliar to students. This required teachers to find ways to simplify them for students.

8.3 Intervention students' responses to the additional post-trial questionnaire items

In the post-trial questionnaire, four close-ended items were added in the intervention group to elicit the intervention students' opinions about the intervention they have used.

Table 8.2: Intervention students' perceptions of the intervention

Perceptions	Strongly disagree (1) (percent)	Quite disagree (2) (percent)	Not sure (3) (percent)	Quite agree (4) (percent)	Strongly agree (5) (percent)	Mean / SD
Perceived that confidence is improved	8.5	16.5	31.1	38.4	5.5	3.16 / 1.04
Perceived that motivation is improved	4.9	17.1	25	39.6	13.4	3.40 / 1.07
Perceived that listening skill is improved	7.3	14.6	31.1	36.0	11.0	3.29 / 1.08
The intervention is helpful.	7.3	11.0	27.4	39.0	15.2	3.44 / 1.10

Table 8.3: Intervention students' perceptions of the intervention by faculty

Perceptions	Faculty	Strongly disagree (1) (percent)	Quite disagree (2) (percent)	Not sure (3) (percent)	Quite agree (4) (percent)	Strongly agree (5) (percent)	Mean / SD	Effect Size
Perceived that confidence is improved	Education	6	7	16	15	2	2.99 / 1.08	-0.31
	Political Science	2	9	15	23	4	3.31 / 0.99	
Perceived that motivation is improved	Education	4	7	15	16	4	3.22 / 1.07	-0.31
	Political Science	1	10	10	23	9	3.55 / 1.05	
Perceived that listening skill is improved	Education	5	7	16	16	3	3.09 / 1.10	-0.34
	Political Science	2	8	15	20	8	3.46 / 1.03	
The intervention is helpful.	Education	6	6	12	15	7	3.25 / 1.25	-0.33
	Political Science	1	5	15	24	8	3.61 / 0.93	

The results in Tables 8.2 and 8.3 suggest that most intervention students have positive attitudes towards the intervention. There were a number of students (approximately 18 %) having negative views about the intervention. When considering the responses in relation to

students' disciplines, the Pol Sci students have a higher impression with the intervention. As presented in the previous chapter, the Pol Sci students are second-year students while the EDU students are freshmen. This may have two implications. One is the teacher who taught all three Pol Sci clusters might cause the students to have a more positive perception of the intervention. Another interpretation may be that being in a higher year of study may make students have better attitudes toward the metacognitive intervention.

8.4 Intervention students' responses to the open-ended questionnaire

Open-ended questions were included in the intervention group's post-trial questionnaire. The questions asked about their overall opinions of the intervention, the good or bad things about the intervention they experienced and suggestions they might have. Overall, most students see the advantages of the intervention but some limitations and drawbacks were also expressed. The responses were coded and categorised to develop key themes, as presented below.

8.4.1 Advantages

Table 8.4: Codes of students' responses on the advantages of the intervention

Codes	Counts
improve strategic listening, skills	49
improve planning, preparation, goal setting	32
improve self-awareness, problem-solving, evaluation	19
better learning/systematic learning skills	17
understand English listening better	17
improve confidence, motivation in English	16
improve systematic, reflective thinking, higher thinking	16
improve English skills	15
organised content, sequence, easy to understand	15
improve focus, selective attention in listening	14
improve skills for making inference, dealing with unknown words	8
applicable for real use in other skills, wider situations	7
useful for taking real tests	6
schema activation	5
applicable for real use (real listening)	5

helping to decrease pressure, anxiety	2
changing mindset, attitude to learning English	2
good for learners with any proficiency level	2
Independent practice	2

From the codes above, main categories which manifest the benefits of the intervention were developed, based on their frequencies.

Improving ways of learning

Most students view the strategies introduced in the intervention as the most useful features of the intervention as they improve their ways of learning. The strategies mostly reported as helpful are planning, problem-solving and evaluation, directed attention, making inferences and schema activation respectively. Moreover, some students perceived that the invention guides them to have more organised and regulated learning.

Enhancing understanding

Another significant benefit of the intervention reported is it helps students to better understand the listening task. A similar number of students reported that the intervention does not only improve comprehension but also improves their overall English.

Improving thinking

The next benefit of the intervention is improving systematic thinking and self-awareness. The intervention is perceived by some students to guide them on what to do at each phase of listening which helps them to think and approach the task systematically. Some students reported they evaluate their understanding more frequently which improves the awareness of their own learning.

Improving motivation and confidence

A considerable number of students expressed that the intervention improves their motivation and confidence in English and English listening. It made them realise that not understanding the text fully is not uncommon and guessing or even skipping may be appropriate strategies in some cases. Such awareness could maintain their motivation to work with the tasks.

Organised content and sequence

A number of students described the contents and pedagogical stages as systematic and organised. This is one of the features which makes the instruction easy to follow.

Practical applications

Additionally, a few students believed that the intervention can be applied in other language skills and other English courses. Some thought it can be used for real English tests or real-life listening.

8.4.2 Drawbacks

Table 8.5: Codes of students' responses on the drawbacks of the intervention

Codes	Counts
I know little/not enough vocabulary	21
some terms used are difficult	11
seems difficult concept, principles	9
students have different backgrounds, less-able students	9
too many details	8
takes time, time consuming	8
a lot of theoretical concepts	5
cannot follow the audio, too fast	4
online leaning is not preferred	4
not effective, not usable in all situations	4
the audio, listening text is difficult	3
the intervention seems complicated	3
the class moves quick, can't catch sometimes	3
didn't pay enough attention, lose attention	3
not appropriate in limited time	3
can lead to being too focused on principles	3
repetitive pattern leads to boredom	2

Personal limitations

The most highly perceived drawback reported pertains to students' personal limitations. The first limitation often mentioned is a limited vocabulary. This affects their ability to process the materials which contain words unfamiliar to them. Another limitation is their different background knowledge especially their proficiency levels. The very low proficient students might find the intervention more difficult. This can suggest that the intervention may not have adequately supported this group of students.

Constructs of the intervention

Another major drawback which is about the intervention itself is some complicated elements in the intervention. Some students expressed that the intervention has many difficult terms while the underlying concept already seems complicated. They also thought that the intervention has quite a lot of details and theoretical concepts which make it look sophisticated to some students. A few students expressed that the design of the intervention look too repetitive and can become boring.

Time constraints

Related to the previous drawbacks about excessive details, some students perceived that each phase in the session takes a long time. Given that the period the intervention was introduced was constrained by the limited availability of time due to the COVID outbreak, the class time spent on the detailed principles of the intervention means that there is less time for engaging students in group discussions.

Additional limitations

As most parts or all parts of the intervention in some groups were delivered online, few students expressed that online learning may make them feel less engaged and found the intervention less enjoyable. In addition, few students thought that the intervention cannot be used in all situations such as when the accent of the speakers is totally unfamiliar or when lacking knowledge about the topic.

8.4.3 Suggestions

Table 8.6: Codes of students' responses on suggestions for the intervention

Codes	Counts
need more time	3
continue teaching it	3
more time for conversation, group discussion	2
fewer difficult terms	1
good and should be supplied in other subjects too	1
may be better if integrating with other skills	1

Suggestions for improvement

Some students suggested that the intervention should be continued. For improvement of the intervention, some students suggested that the intervention should be applied in the courses which can arrange more time for applying the intervention. Some students advised that more time should be given for peer discussions. Few students suggested applying it in other subjects or integrating it with other skills such as reading or speaking. In addition, it was repeated again that fewer difficult terms may make students find it more accessible.

8.5 Student interviews

The students were asked to voluntarily join in a semi-structured interview. Eight students from the intervention group and six from the control group agreed to take part. The questions were about their attitudes towards English, their opinions about the Exit Exam policy, opinions about English teaching and learning in Thailand. For the intervention students, additional questions about the metacognitive intervention were asked.

Students' attitudes towards English

All of the students in both intervention and control groups agreed that English is important and inevitable for them. Two main reasons they gave are the usefulness of English for future career paths and capacity for global communication. Some students added that English is helpful for learning, accessing more information and discovering new things. Additionally, some students believed competence in English would give them wider life opportunities.

Students' opinion on the Exit Exam policy

Almost all students agreed with having this policy while two students agreed in principles but disagreed with its actual implementations. They reported that the benefits of the policy are its encouragement for preparation and practice and the assessment results which allow students to realise their proficiency level. Even when asked whether the policy adds unnecessary extra burden to them on top of the regular graduation requirement, students viewed it as a sensible requirement rather than a burden. Another reason is that passing the university requirement assures students' basic English competence.

Nonetheless, the two students with inhibitions on their agreement with the policy suggested the assessment does not have to be a test and should take various forms. They perceived that the current university test does not actually assess all four language skills and does not account for their practical English ability. One student added the sole reliance on test results can be judgemental and discouraging for some students.

Experiences and perceptions of English education in Thailand

Most students have a shared perception that English education in Thailand has been excessively focusing on the rules and grammar of the language and there has been a lack or insufficiency of communication and opportunities for oral and aural skills. Some students add that the mistakes students make in classrooms are usually treated negatively. Some students feel that the prevailing classroom environment is a 'teach for test'. They perceive that such practices are discouraging and demotivating. They also attribute these practices as the major obstacles for English education in the country because such ways of teaching are parts of the reasons which cause the students' lack of motivation, confidence and positive attitudes to learn English. Additional factors inhibiting success are the learners' fixed mindset and the teachers' teaching techniques which do not adequately engage the learners. For solutions, students suggested that the teaching approach should be more communicative and the learners themselves need to seek more chances for practices. In addition, teachers should assess learners' motivation and attitudes and constantly try different techniques to support students to learn and develop.

Students' opinions on key factors for progress in learning English

Most students perceived that motivation, interest and attitude are the most important for successful learning. This should be accompanied with opportunities for practice and

authentic use of English. A few students added that learning in groups with friends is a practical way to provoke communication and support improvement because learning with more-capable peers encourages one to improve and allow an exchange of better ways to learn.

Opinions on metacognitive intervention

Overall, the interviewed students have positive perceptions of the intervention. They reasoned that it introduces and guides them with useful strategies for learning. The strategies they usually mentioned are planning, directed attention and problem-solving as they help them to panic less when dealing with the task. Another reason is it helps them become more aware of their own knowledge. Some students thought it increases their understanding from listening and helps them analyse and tackle the tests better. A few students said their application of the strategies in their recent English test was helpful and suggested using them in other English courses.

> "In real tests, we can't expect to know everything, so we need to know what to focus first and what to do next".
>
> Nin (pseudonym)

> "It helps me to have good planning, have aims in listening and tackle the tests better".
>
> Kate (pseudonym)

However, a number of flaws of the intervention have been reported. The two most mutually agreed reasons are the time-consuming structure and the use of unfamiliar terms in the method. As the intervention has a lot of details, covering all of them in the background phase and especially in the practice phase requires time, while the actual time available was rather limited. Regarding the unfamiliar terms in the intervention, the use of simpler terms may have let them learn the technique quicker. In addition, a few students thought that when the problem is about an unfamiliar or strange accent, the technique is less helpful.

> "This approach doesn't help much with the accent we are not familiar. Also, it takes time".
>
> Pong (pseudonym)

> "It has many technical terms which I don't know their meaning".
>
> Nur (pseudonym)

8.6 Teacher interviews

Three teachers who were responsible for all intervention and control clusters were recruited for a semi-structured interview after the end of the intervention. The main questions were about their views on English teaching and learning in Thailand, the Exit Exam policy and additional questions on the intervention for the intervention teachers.

Teachers' views on the policy and English teaching and learning

On the exit exam policy, two teachers are fully aware of it but the other teacher, who is not a full-time staff, does not know much about the policy. Both full-time teachers agreed with the policy in principle but similarly felt that sufficient support system is lacking. Moreover, one teacher suggested that assessment should take various forms and should be competency-based, rather than merely testing students' knowledge. One teacher suggested more practical supports would be needed such as intensive English-medium opportunities or improving internationalisation on campus life by having more exchange students. The other suggested non-credit preparation English course for students and development supports for teachers such as in-house training or overseas.

> "I can see the rationale of the policy but if it only requires students to take tests without appropriate practical supports, it is not very helpful".
>
> Teacher A

On English teaching and learning in Thailand, what was agreed by all teachers is that students generally have weak English backgrounds. They pointed out that such weakness might have been partially caused by the teaching styles of Thai teachers of English since the primary level which primarily focus on rules instead of the real use of English for communication. The rule-focused teaching also discourages and demotivates students as they may find the class boring or irrelevant and the 'grammar police' culture makes learning English even intimidating. Despite competence being endorsed in the curriculum, competency-based practice does not frequently occur in the actual classrooms. The interviewed teachers believe that Thai teachers, as non-native speakers of English, need development supports to become more confident to 'teach English in English' more fluently.

> "Teachers' mindset on language teaching should be clear. Language is for communication, not mastering rules".
>
> Teacher C

214

On the aspect of learning, two of the teachers believe learners' motivation and desire to learn are essential for successful learning. They observed both low and high motivations in their students and observed that motivation usually aligns with their performance. In addition, two of them think the teaching methodologies are a key factor for learning; therefore, continued teacher development is necessary. Teaching materials and activities also have a crucial role as they can affect learners' interests and attention.

Views on the metacognitive intervention

Two teachers who delivered the intervention described it as a good approach for teaching. One of them has already been interested in applying knowledge from cognitive science in teaching and was willing to apply the intervention. The other teacher found the intervention completely new but thought that it is a systematic and helpful approach. Planning and task analysis are the most apparent beneficial skills which both teachers mutually reported. The strategies helped students to feel more ready and better capable of focusing and maintaining attention in the listening tasks. Both teachers similarly suggested that the approach can be applicable to reading skills and can be used in other English courses. Teacher B has urged students to really use the strategies in future English tests and English courses. Teacher B stated that she had also applied the technique in her recent English test required for teachers.

> "At least, the planning strategies are obviously applicable in actual use. I have recommended the students apply them in real tests and other courses"

> Teacher B

> "Planning, skipping, task analysis has the most obvious benefits. It's applicable to reading".

> Teacher A

However, there were some dominant drawbacks the teachers experienced. First, both teachers agreed that the intervention has quite excessive details. It takes much time for the teachers to follow the guidelines fully while the time is limited due to the reduced teaching time caused by the COVID 19 outbreak. Teacher B thought the principle phase of the intervention has a number of terms students find difficult. Teacher A observed that the contents in the sessions seem to follow the same pattern and can be deemed repetitive, causing students who were attentive in the first few sessions to pay less attention in the later sessions. Teacher B also suggested that the online learning platform is not very convenient for use because

students have to change between windows and screens very often, making the class run less smoothly.

> "I can notice students feel less engaged in later sessions. It could be because the design of the intervention seems repetitive".

<div align="right">Teacher A</div>

> "It has some difficult terms and a lot of details. It is quite time-consuming within the time limitation"

<div align="right">Teacher B</div>

With such problems, Teacher A suggested that the detailed principles about metacognitive awareness should be reduced to allow more time for practice especially for students with low proficiency because they need more time to process and complete the task. He also suggested that the future application of the approach should be adapted by using tasks with contents directly related to students' subject disciplines, hoping that it will maintain students' interest better. Teacher B recommended reducing details and difficult terms. She also suggested the approach could be helpful to other English teachers and the training of this approach should be provided to them as well. She believes the intervention will work better in the normal teaching time and environment.

8.7 Discussion of the process evaluation findings

The classroom observation suggests that there was no sign of contamination of the intervention in the non-intervention groups. The components of the intervention were not exposed to the non-intervention groups during the trial. In the intervention clusters, a fairly reasonable degree of fidelity was observed among the teachers as they tried to comply with the guidelines in the teachers' manual, with some adjustment for practical purposes. No notable fidelity threat was observed because a teacher guidebook and class materials were provided and both teachers chose to use and follow them. As the provided materials were quite detailed, teachers could not comply with them fully. This excessive detail was expressed during the teacher interview as a drawback of the intervention. Even the researcher, who also joined as a teacher in one cluster, also had to make adjustments and skip some parts to meet the time availability.

Regarding students' behaviours and reactions in the classroom, most students participated well in the metacognitive tasks. However, some students seemed less involved.

They could be students who lack sufficient English knowledge and skills to actively involve in the metacognitive tasks. Without sufficient relevant knowledge and cognitive skills, it is unfeasible to apply metacognitive strategies effectively (Quigley et al., 2018). For successful comprehension, sufficient knowledge of vocabulary and word recognition skills are crucial for decoding and making meaning because they can ease the cognitive load in the short-term memory during the comprehension processes (Pressley, 2002). During the tasks, some of these students were less responsive at first but started to gradually get involved more actively in the pair work and small group discussions. This echoes the role of feedback from interactions with others which is one of the primary feedback sources in the social cognitive perspective to self-regulated learning (Zimmerman, 2000).

As the classroom observations occurred both on-line and on-site, some potential differences could be noticed. In the on-line observation, the researcher presence may have less influence on the participants' attention and behaviours. However, it seemed to be more difficult than the on-site observation because students appeared on their own screen in MS TEAM, making it difficult for the observer to see all of them at once, especially in the big class like the comparison group. Moreover, because the classroom interactions were quite limited on-line, the observations on students' reactions and interactions were quite limited accordingly.

For the on-site observations, the teacher-student and peer-to-peer interactions seemed to occur more frequently. Of course, this was partially due to the styles of the teachers. However, being in the real classroom together in person makes interactions more instant and more convenient. In the researcher's opinion, the on-site observation is still preferred but the observer's presence should not be too intrusive that it affects the participants' behaviours or intervenes with the normal classroom practices.

From the students' perspectives, most students had positive attitudes and were satisfied with the intervention, as suggested by the questionnaire responses in Table 8.2. The open-ended questionnaire responses in section 8.4 also reveal that students reported application of several strategies and found them helpful such as being more strategic in listening, setting goals, planning and improving confidence. Such reported benefits support the findings of the positive impact of the intervention discussed in Chapter Five. Indeed, it is one of the main aims of metacognitive instruction to guide students with different strategies so that they can regulate their learning more effectively (Anderson, 2008; Goh, 2008). When students' awareness about various available strategies is raised, it is more likely that they will use them and use them in a

more effective manner. In line with Pintrich (2002), self-efficacy, which is usually embedded in metacognitive instruction, is a primary source of motivation. Self-regulatory activities affect one's motivation, choices of learning strategies, effort and persistence (Zimmerman & Moylan, 2009). In addition, the systematic way which the PMER model was sequenced and presented makes it easier for students to follow. Explicit and clear instruction can ease the cognitive load of students (Sweller, 1994). As suggested in the systematic review in Chapter Three, the clear pedagogical sequence is a fundamental feature in most successful metacognition-based studies.

Despite positive perceptions and reported benefits, some major drawbacks of the intervention have been expressed repeatedly in most process evaluation means. From the additional post-trial questionnaire responses asking students' opinions about the intervention, around 18 per cent voiced negatively about it. From class observations, some students seemed less engaged than others. The open-ended questionnaires and interviews can provide insights into these manners. The first reason is the difficult terms used in the intervention. This was expressed by both teachers and students. They can make the intervention look difficult to some students. Such difficulties can increase students' cognitive load which could hinder their learning (Ashman, 2015). Secondly, students expressed that they do not know enough vocabulary. This makes them find the intervention difficult or unhelpful. In fact, it is impossible or ineffective to be metacognitive when the relevant cognitive knowledge is lacking (Quigley et al., 2018). Different sources of knowledge such as linguistic knowledge and prior knowledge are required and different types of strategies interact in the metacognitive and self-regulatory processes (Vandergrift & Goh, 2012). To build comprehension, vocabulary knowledge is necessary for the word-level processes of comprehension (Pressley, 2002). It seems that some students lack sufficient relevant basic knowledge to metacognitively regulate their listening and that the intervention did not adequately address the bottom-up strategies which are also essential in listening comprehension (Graham & Macaro, 2008) which may support this group of learners better. When students do not see the relevance of the approach, it is more difficult for them to develop metacognitively (Veenman et al., 2006). These drawbacks may in part contribute to the ambiguous and mixed impact of the intervention on the secondary outcome of metacognitive awareness discussed in Chapter Six. Thus, a higher proportion in the intervention dealing with the bottom-up strategies for perception such as features of sounds in natural connected speech should be considered.

From the pre-test scores shown in Chapter Seven, it was found that the majority of the intervention students are low proficiency learners. The researcher realised since the beginning

of the trial that the scenario where the low proficiency students may think that they are not good enough or the intervention is too difficult could occur and that may result in the intervention being ineffective. However, as one of the main aims of the intervention is for students to be aware of their own strengths and weaknesses and try different strategies to improve their learning, it was hoped that the metacognitive strategies and guidance on self-regulation will help these low achievers to regain their self-belief and take better control of their own learning, regardless of whether the intervention will be effective or not. Even though the fixed mindset still remains for some students, the fact that more students viewed the intervention positively and that the low-proficient group gains more benefits from the intervention demonstrates that the intervention has done more good than harm to the students, especially the less capable learners.

From the classroom observation, the two intervention teachers tried to comply with the intervention guidebook. From the interviews, they expressed support for the intervention and found it useful. The systematic nature and planning strategies are frequently mentioned as the positive features of the intervention because they are easy to follow and help students to feel more ready and maintain their focus better. Indeed, strategic thinking is the key construct in metacognition which promotes meaningful learning (Moseley et al., 2005). Moreover, both teachers agreed that the principles are applicable to reading classes and can be used in other English courses. Both listening and reading are comprehension skills and share many processes such as word recognition, decoding, bottom-up and top-down processing (Pressley, 2002). Obviously, the approach can be applied to reading and in fact, more studies on metacognition have been done on reading comprehension (e.g., Macaro & Lynn, 2008; Tavakoli & Koosha, 2016; Teng & Reynolds, 2019). Furthermore, there are other features in the intervention such as self-efficacy and reflective thinking which can be useful for students in wider contexts because they are positive dispositions which are transferable across different situations (Perkins & Salomon, 2012).

From the class observations and discussions with the intervention teachers, it was noticed that one intervention teacher was particularly interested in applying the metacognitive instruction with his students. During the pre-intervention orientation, he expressed enthusiasm about the approach because he has developed an interest in applying knowledge from cognitive science in teaching. During the class observation, he delivered the intervention with eagerness and was able to provide his own explanations and examples to make it easier for students to understand the concepts and terms. Interestingly, the outcomes from his clusters are high and

even higher than the group taught by the researcher. This emphasises the vital role of the teachers in self-regulation classrooms. As the sociocultural theory postulates, expert advice from the teachers or more able peers can well support learning. Teachers' modelling and scaffolding are essential for the learners to develop self-regulatory skills (Quigley et al., 2018). From the social cognitive perspective, social interaction with the environment such as teachers is a fundamental source of the feedback utilised in the self-regulatory processes (Zimmerman & Moylan, 2009) and contributes to more substantial self-regulatory competence (Bandura, 1986).

Nevertheless, some challenges of the intervention were faced by the teacher and student participants. The major difficulty the teachers experienced while delivering the model which was expressed during the interview was the excessive details in the teacher's guidebook. In line with students' voices, the teachers suggested that a lot of details seem to cause students to pay less attention in the later sessions. While the concept itself sounds complicated, the excessive details in the intervention make it appear to be exhaustive (Wongdaeng, 2021). As expressed in some studies in the systematic review in Chapter Three, the lengthy and excessive lectures about an intervention can make students feel bored (Yang, 2013). Therefore, more time should be devoted to opportunities for students to actually regulate their learning (Chou, 2017; Rahimirad & Shams, 2014). From the interview, a teacher suggested that the principle phase should be decreased to spare more time for practice and discussion. This suggestion is logical because the opportunities to actually perform the tasks and learn from the more-able learners or teachers enhance self-reflection and make the self-regulatory processes develop more effectively (Bandura, 1986). Moreover, the newly learned language knowledge and skills are reinforced by practical application in the tasks (Field, 2008).

Another drawback is related to the repetitive nature of the pedagogical sequence. A teacher suggested that many students were attentive in the first few sessions but seemed to pay less attention in the later sessions because the metacognitive sessions appeared to follow the same pattern. During the implementation, each metacognitive session was not continually taught next to each other but spanned over the semester in combination with the normal syllabus. Thus, this repetitive structure was not deemed negative at the time the sessions' materials were designed. In fact, the teacher's reflection is right because repeating the same pedagogical routine for a long time can diminish students' attention (Goh, 2008). For the future implementation of the approach, the contents and materials which are more directly relevant to

students' subject discipline and varied instructional contexts may be used (Vandergrift & Tafaghodtari, 2010) to better cater for students' needs and diversify the learning environments.

It should also be noted that the student interviews were conducted with volunteering students from the intervention and control groups. It was possible that the volunteering intervention students might be students who were quite satisfied with the intervention. Even though they reported both benefits and drawbacks of the intervention, more or different views on the limitations or weaknesses of the intervention could have been obtained if the students who had less satisfaction with the intervention had been included in the interviews. Nonetheless, the interviews with a few students were not the sole method for process evaluation. The close-ended and open-ended questionnaires which elicited responses from all intervention participants could have obtained the perceptions, reactions and suggestions from the intervention students, inclusive of those satisfied or dissatisfied with the intervention.

In addition, one indirect but significant lesson learnt from the process evaluation is the increasingly influential role of blended or technology-enhanced pedagogy for education of today and the future. The worldwide chaos caused by an outbreak of the COVID 19 pandemic made the role of technology even more dominant as the availability of technological support allowed teaching and learning in many institutions to be maintained (Başal & Eryılmaz, 2021). However, online teaching has certain limitations, as expressed by some trial students and teachers and also in the researcher's own observation. Some students suggested they feel less engaged in the online sessions while some teachers expressed that it is less convenient to move between tasks and devices during the online instruction. This implies that it is important for teachers in this age to improve themselves on online or blended learning pedagogy in order to optimise the functionalities which the available tools can offer to engage learners in the online mode of teaching because it is uncertain about when the pandemic is going to end. Even after the pandemic, being able to effectively apply the available technology to enhance teaching and maximise students' learning experiences is advantageous. Unfortunately, there are areas where access to technology to support education is limited and this limitation aggravates inequality (Hadjeris, 2021). Technology and infrastructure supports are certainly needed to minimise this inequality gap along with pedagogical support to maintain teaching quality despite encountering challenges.

8.8 Chapter summary

The results in this section answer the fourth research question. From the questionnaire and interviews, most students have positive views of the intervention and the teachers see that the benefits of the intervention outweighs the experienced drawbacks. The awareness-raising and regulation of available strategies, self-belief, strategic problem-solving and reflective thinking are reported as useful elements of the intervention. Meanwhile, the difficult terms and exhaustive details can overwhelm some EFL learners. This demonstrates that the metacognitive intervention can be applicable to EFL contexts. However, the drawbacks expressed by the participants should be considered in future implementation.

CHAPTER NINE

RESULTS: SECONDARY DATA ANALYSIS

This chapter presents the analysis of secondary data from larger samples of the same population. The obtained secondary data involves the full cohort data of university English tests, parental income and admission English scores. The results from these data are discussed in comparison to the same data from the intervention students. How these data are related and how they may shed light on the impact of the intervention are discussed, followed by the summary of the chapter.

9.1 Introduction

The use of available official data is useful for public policy evaluation because secondary analysis of large-scale data can provide empirical evidence and predict long term education trajectories (Smith, 2008). As one main aim of the trial is to investigate whether the intervention can help the students to have a higher potential to pass a standardised English test to satisfy the graduation requirement demanded by the Exit Exam Policy, the available results of larger non-trial students' test scores can provide more insights into the impact of the intervention as well as the impact of the policy on the stakeholders. The available data which could be obtained from the full cohort 2019 and 2020 are students' recent University English test scores, their parental income data and their admission English scores. These data are from one of the participating universities because such data from the other university where the comparison participants attend was not accessible.

9.2 University English Test (TMM) scores

The university English test, which runs on the TellMeMore (TMM) software, is one of the acceptable tests which students can take to qualify for graduation. Below, the test scores of cohort 2019, where the Pol Sci trial participants are from, and the cohort 2020, where the EDU trial participants are from, are presented in order by pre-trial scores and post-trial scores. Normally, the TMM test is arranged for two rounds per academic year at the beginning and towards the end of the year for students to take as a placement test and a progress test. Due to the COVID-19 outbreak which caused a cancellation of on-site activities for a few months, the TMM test administration in that year took place only once after the trial. Therefore, for cohort 2020 students who were new first-year students and had just started their semester during the COVID-19 outbreak, the TMM test had not been administered before the implementation of the trial. Meanwhile, the pre-trial data of cohort 2019 students is available because it was the

test results administered at the end of the previous academic year. Thus, both pre/post-trial data of TMM test are available for cohort 2019 but only post-trial test data are available for cohort 2020.

Figure 9.1: Cohort 2019 pre-trial TMM English scores
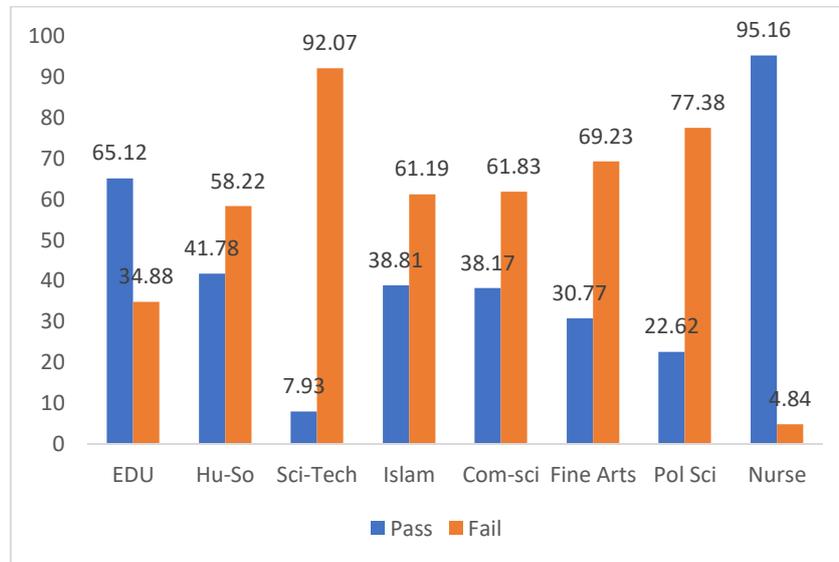
Figure 9.2: Cohort 2019 post-trial TMM English scores
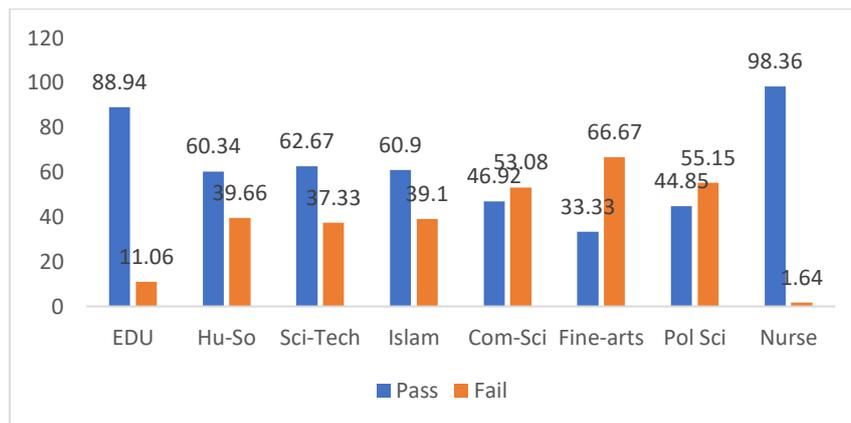
Figure 9.3: Cohort 2020 post-trial TMM English scores
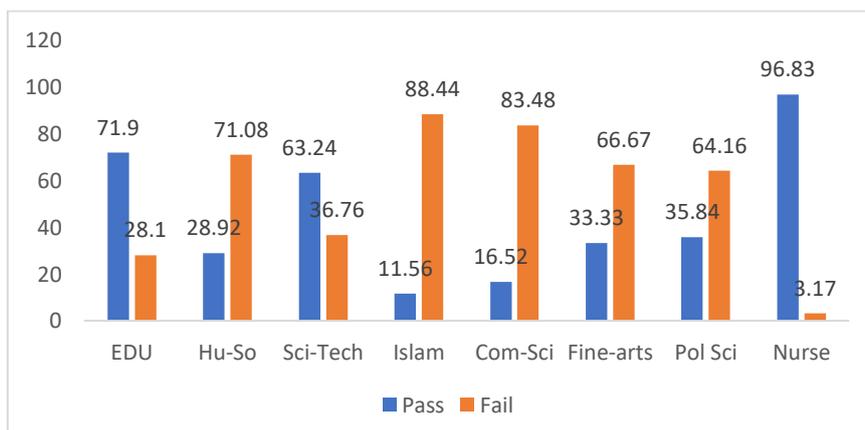
From the full-cohort data, most students cannot satisfy the English Exit Exam policy's requirement in their early years in higher education. For cohort 2019, which are second-year students, the pre-trial data drawn when they were freshmen shows most students failed to qualify the criteria. The post-trial data, which were drawn recently in the second year of their study, sees more students satisfying the requirement. However, a sizable number of students still need to develop their English competence further to meet the graduation criteria.

Next, the above full-cohort data were used to compare with the data of the trial participants who managed to pass the English requirement before and after the trial. This can provide an indication of how much the intervention may have an impact on helping students to satisfy the graduation requirement.

Table 9.1: Increased passing rates of University English Test by each faculty of cohort 2019 and 2020 students

Faculty	Passing Rate: Cohort 2019 (%)			Passing Rate: Cohort 2020 (%)	
	Pre-Trial	Post-Trial	Change	Pre-Trial	Post-Trial
EDU	65.12	88.94	23.82	N/A	71.90
Hu-So	41.78	60.34	18.56	N/A	28.92
Sci-Tech	7.93	62.67	54.74	N/A	63.24
Islam	38.81	60.9	22.09	N/A	11.56
Com-sci	38.17	46.92	8.75	N/A	16.52
Fine Arts	30.77	33.33	2.56	N/A	33.33
Pol Sci	22.62	44.85	22.23	N/A	35.84
Nursing	95.16	98.36	3.2	N/A	96.83
Whole Cohort Average	41.81	64.40	22.59	N/A	42.72
Intervention	14.50	65.65	51.15	N/A	84.71
Control	6.30	35.43	29.13	N/A	63.64

Among the cohort 2019 students, the trial participants from this cohort are mainly from the Faculty of Political Science (Pol Sci). Considering the increased pass rate between the trial participants from this cohort and the all students from the same faculty (Pol Sci), the improvement rate is highest in the intervention group (51.15%), followed by the control group (29.13%) and the rate of all Pol Sci students (22.23%) respectively. This indicates that the intervention group has made a remarkably higher increase in the pass rate than their control and non-trial peers. Compared to the average improvement of the whole cohort 2019 from all

faculties, the intervention students also have a substantially greater increase in the pass rate. This once again affirms the impact of the intervention. Considering the data by each faculty, the increased pass rate in the intervention group is higher than all faculties except the Science and Technology which has a slightly higher improvement rate than the intervention group. The data shows that many cohort 2019 students still cannot pass the tests. This means that they do need more support to help them qualify for the English proficiency requirement. The significant increase in the pass rate made by the intervention group indicates that promoting metacognitive and self-regulatory skills can be an effective measure to support students in this context.

Regarding cohort 2020, the trial participants from this cohort are from the Faculty of Education. Table 9.1 suggests that the students in this cohort who received the intervention have a higher pass rate (84.71%) than the control group (63.64%) which has a lower pass rate than the average rate of education students (71.90%). The intervention group in this cohort also has a remarkably higher pass rate than the average of the whole cohort 2020. Moreover, the intervention group has a higher pass rate than the rates in all faculties except the nursing students who generally have higher English proficiency than other faculties, as the data from cohort 2019 also suggests. This implies that the majority of cohort 2020 students need to develop their English to meet the requirement. The results also reaffirm that promoting students' metacognitive and self-regulatory competence can be an effective measure to help students meet the university' English requirement for graduation.

9.3 Parental income of students from cohorts 2019 and 2020

For parental income, data from both cohorts 2019 and 2020 are available and have sufficient information for identifying which data belongs to the trial participants to allow comparisons. Thus, the full-cohort data from both 2019 and 2020 students can be used to compare with the intervention group data. The income data were classified into two groups based on the minimum wage announced in the government gazette in 2020 of approximately 9,000 baht per month as a benchmark.

Table 9.2: Parental income of cohort 2019/ 2020 students and the intervention students (IG)

Parental Income	Full Cohort Number	Full Cohort Percent	IG Number	IG Percent
Below minimum wage	1,873	51.24	90	45.69
Meet minimum wage	1,774	48.54	105	53.30
Missing	8	0.22	2	1.02

Total	3,655	100	197	100

As can be seen from Table 9.2, a large proportion of students, both overall and the intervention participants, are from the family which has a lower income than the minimum wage set by the government. The percentage of students with low parental income in the intervention group is slightly lower than the average of the two cohorts. This large volume of students with low-income backgrounds suggests a form of segregation by income (Gorard & See, 2013) which can have some association with the overall poor performance, as reported in many studies (e.g., Early et al., 2020; Gorard, 2018; Siddiqui, 2017) and can be linked to the TMM test performance in the data set presented earlier in section 9.2.

Nevertheless, appropriate support can promote student development regardless of their income background. From Table 9.2, the income data is not much different between the intervention students and the full cohort data while the intervention students have a notably greater improvement, as shown in the previous section. Similar to the findings from the subgroup analysis and regression analysis of the trial participants in Chapter Seven, the parental incomes which may have some link to the low pre-trial achievement of these students do not impede the improvement of the students who were exposed to the metacognitive intervention as a support for more effective learning. An intervention which develops students' cognitive ability can have stronger effects on students' attainment than some socioeconomic variables (O' Connell & Marks 2022).

9.4 Admission English score of cohort 2019 students

The admission English data of both cohorts 2019 and 2020 were requested from the two participating universities but only one university allowed some access to these data. Only cohort 2019 data could be obtained because it was requested early during the pilot study which was conducted in academic year 2019. This data of cohort 2020 was also requested later at the beginning of the main trial which recruited students from both cohorts. However, there was a change in the policy from the national agency in charge of administering the test which disallowed access to test takers' identity from cohort 2020. Thus, only the admission English score of cohort 2019 has sufficient information to identify which data belongs to the trial participants to allow comparisons. Therefore, only cohort 2019 data was included and could be used to compare with the same kind of data from the intervention students.

Table 9.3: Admission English scores, cohort 2019

Admission English (Full score:80)	Full Cohort Number	Full Cohort Percent	IG Number	IG Percent
25 and lower	762	44.02	45	39.47
26-39	720	41.59	49	42.98
40 and higher	246	14.21	16	14.04
Missing	3	0.17	4	3.51
Total	1,731	100	114	100

From Table 9.3, similar patterns of score ranges can be seen from the whole cohort data and the data of intervention students. That is to say, the percentages between the low scorers (below 25) and moderate scorers (26-39) are not much different in the whole cohort data (44.02/ 41.59) and the intervention data (39.47/ 42.98). Likewise, the percentage of high scorers (40 and over) are similar in both groups (14.21 and 14.04). This shows that the student intake at this institution is segregated by low English proficiency. This academic segregation can hinder the disadvantaged students from equal access to success opportunities (Siddiqui, 2017). The new graduation requirement can be an additional challenge for them for graduation. Thus, appropriate support from the university, the government or other parties is required for this disadvantaged group of learners to abate the negative consequence from this new challenge. The promotion of self-regulated learning can be one of the efficient supports.

9.5 Discussions of the findings

The secondary data on admission English scores and the university test score, in line with the primary data from the pre-test, consistently illustrate that the majority of students in this disadvantaged area have low proficiency in English. This indicates that there is academic segregation in tertiary institutions in this area because most of the population in the same intake shares similar negative backgrounds (Gorard, 2018). Regarding the parent income data, a similar trend is seen as the low incomers constitute a large proportion of the students in the same intake. The income variable is one of the most relevant socioeconomic factors associated with learning outcomes (Siddiqui, 2017). Students under these segregations are considered disadvantaged because the issues can affect their success opportunities in life. This emphasises the need for support for these groups of students to have equitable access to quality opportunities for development in learning outcomes as well as wider outcomes (Gorard & See,

2013). Indeed, this issue should be given more attention in education policy and practice and this applies to English education policy, especially in the contexts where there is high inequality to access opportunities, resources and other support for developing practical English language competencies.

As a form of support for these advantaged students, a pedagogical approach based on metacognition and self-regulated learning was proposed in this study. The intervention students, despite sharing similar characteristics of being dominated by low income, low admission English score and low pre-test score, can improve significantly. The improvement rate of the intervention students is clearly higher than the control students and the average of students in the whole cohorts including those who did not attend the trial. This shows that appropriate pedagogical support can be helpful for students regardless of their backgrounds. Thus, it can enhance equity in education as it can help students to achieve their educational goals without having their socioeconomic backgrounds as an obstacle (Field, Kuczera & Pont, 2007). The provision of appropriate support for these disadvantaged students can also respond to the United Nations' Sustainable Development Goal (SDG) 4 of promoting inclusive and equitable education for all.

9.6 Chapter summary

The secondary data in this section provides a broader picture of the make-up of the population from whom the trial participants were drawn. Segregations by income background and academic performance were observed in the 2019 and 2020 intakes of the population. This seems to support some existing evidence that these two variables are associated. The intervention students share these characteristics of having large students with low income, low admission English scores and low pre-test scores. However, their pass rate of the university English test can improve significantly after the trial, compared to non-intervention students. Such findings reiterate the impact of the intervention and demonstrate that their biographical backgrounds do not disallow them to develop and improve their learning outcomes if appropriate support is given. Such findings are well in tandem with the results from the subgroup analysis and regression analysis of the trial participant data. This echoes the significant role of educational support which should reduce the inequality gap and increase opportunities for success for students with less privileged backgrounds.

CHAPTER TEN

LIMITATIONS AND IMPLICATIONS

This chapter discusses the limitations of the study which might need to be considered along with the findings. Next, the implications for students, teachers and policy makers are suggested, followed by specific suggestions on the relevant policy. It ends with a summary of the chapter.

10.1 Limitations

In an attempt to synthesise the current evidence of metacognitive interventions in tertiary EFL contexts, a systematic review, which appears to be the most exhaustive one on the topic at the time, was conducted. However, the evidence on effectiveness is not secure due to the limitations of the included studies in terms of sample size and validity assurance. This could be because the search was not exhaustive enough or the inclusion criteria were not sufficiently rigid. Therefore, the large synthesised effect size in the meta-analysis is not secure, making the evidence on effectiveness inconclusive. Nevertheless, it indicates the high potential of the approach and the findings from the thematic analysis can be applicable to the main trial.

For the most reliable results of the main trial, attempts have been made to maintain the internal validity of the study. Examples of such attempts are the multi-method design with all five key research elements, random allocation operated by an independent third party, stratification by faculty, standardised measure of the primary outcomes, back-translation of the questionnaire, delivery of the intervention by the regular teachers in most clusters and provision of manuals for teachers for fidelity. In addition, the analyses of headline results included a sensitivity analysis to account for attrition and investigation of researcher's effect on the primary outcomes. However, some limitations remained which can affect the study. Such limitations are acknowledged and discussed below.

The first limitation is related to sample size which is an essential research element. Overall, more than 800 participants partook and remained until the end of the trial. Nevertheless, as the trial is a cluster randomised trial, having 14 clusters in total is considered a small trial (Torgerson & Torgerson, 2008). Ideally, a higher number of clusters with a small cluster size is preferable to a small number of clusters with a big cluster size (Gorard, 2013). In practice, this trial has a small number of clusters with unequal sizes of randomised clusters and immensely larger non-randomised clusters which were included for comparison. With this

awareness, there was no attempt in the interpretations to generalise the findings. Above all, when the overall participants are considered, the trial was the biggest on the topic at the time it was implemented.

Another major limitation caused by the teacher shortage at the trial institution is that one teacher in the trial had to teach both intervention and control clusters. This provoked a risk of dilution bias and made it impossible to make the teacher blind to the group condition (Torgerson, 2003). To minimise this issue, the teacher was requested that she adhere to the intervention guidelines with the intervention group and not to introduce the approach to the control group until the trial was over. In addition, the researcher took responsibility for delivering the intervention in one of the intervention clusters. This was done to allow comparisons of the results. However, this attempt was inevitably at risk of another kind of bias by the experimenter's effect (Gorard et al., 2017). In the analysis, this threat was accounted for by comparing the results between the researcher-taught intervention cluster and the non-researcher intervention clusters. No evidence of the experimenter's effect was found, as discussed in Chapter Five.

As the main trial was implemented during the period when there was a COVID-19 outbreak, the regular teaching time was reduced and partially moved to online platforms. The reduced time caused the time for the intervention to be limited. As the trial teachers suggested, the time for students' practice in the class where they could learn from each other and from the teacher was inadequate. Moreover, the interactions via online teaching and learning were limited. The online tools are very helpful and can bridge many gaps which mere traditional teaching cannot address, especially if they are used in a balanced way. However, over-reliance on online teaching limits the actual interactions which play a vital role in learning, according to the social cognitive perspective (Zimmerman, 2000).

Another limitation was the low response rate of the questionnaire of the non-randomised comparison group. The questionnaire was used as a tool for assessing the secondary outcome of the study which is students' metacognitive awareness. The response rate from the intervention and control groups was well above 80% of the number of students recruited into their groups. The percentage of comparison participants who completed both pre and post questionnaire was below 70% of the recruited students. This makes the results from this group weak and vulnerable to the missing cases which undermine the security of the results (Gorard et al., 2017). This low-response rate could be caused by the fact that the questionnaire

was arranged to be completed online to allow time for students to think and answer at their own pace. Inevitably, this low response rate affected the security of the findings from this group regarding the secondary outcome. Thus, the interpretations of the findings from the comparison group questionnaire need to be made with caution.

Overall, there were some limitations which can be threats to the validity of the study or affect the optimal plan for the implementation of the intervention. These threats need to be taken into consideration when interpreting the results. The limitations related to the practical implementation can represent the issues which should be improved in the application of the approach.

10.2 Implications for learning, pedagogy and policy

From the findings discussed so far in the previous chapters, this section proposes implications for learning and future pedagogical practices of metacognitive instruction, suggestions for promoting it as a policy and recommendations for the betterment of the Exit English Exam policy.

10.2.1 Metacognitive instruction: pedagogy and policy implications

As the evidence suggests, the metacognitive intervention is the principal factor in improving the intervention students' achievement with substantial effect sizes compared to the control and the comparison students. The intervention also led to the intervention students improving obviously higher than the control group as well as the non-trial students in satisfying the university's English requirement. The awareness-raising of different types of strategies students can use to deal with the task makes them more competent to maintain and accomplish the task (Graham& Macaro 2008; Pintrich, 2002). The underlying principles and theories, such as self-efficacy and strategic and reflective thinking, are positive dispositions propelling such improvement (Wongdaeng, 2021). The approach is, then, beneficial for improving second language learning (Chuo, 2017; Graham & Macaro, 2008; Vandergrift & Tafaghodtari, 2010). Potentially, it can be advantageous for wider outcomes, as indicated in numerous studies (e.g., Dignath & Büttner, 2008; Higgins et al., 2005; Wang et al., 1990).

The results also affirm that the metacognitive and self-regulatory instruction is applicable in contexts outside Western cultures. The positive effect sizes favouring the intervention group and the positive attitudes in most responses from the process evaluation validate the high potential of implementing the approach with this group of Thai EFL learners.

Moreover, the regression analysis results show that the participants' backgrounds such as their backgrounds in private Islamic schools do not hamper the effect of the metacognitive intervention. Indeed, reflective thinking, which is the very crucial feature of metacognition, is essentially infused in Islamic values (Kraince, 2007). Most of the studies in the systematic reviews presented in Chapter Three also support the applicability of this approach in EFL contexts (e.g., Habibian 2015; Farahian & Avarzamani 2018). Therefore, being in non-Western cultures should not be an objection to promoting students' metacognitive and self-regulatory competence.

Although the metacognitive intervention appeared to be effective, in the researcher' belief about teaching methods, there is no one single method which is comprehensively better than the others. The classroom observation results suggest that when students were prompted to work collaboratively in groups, some students who appeared reluctant at first engaged more actively in the metacognitive tasks. This accords with the rationale of cooperative teaching and learning which has long had its own body of research (Brown, 2001; Richards & Rodgers, 2001). In addition, the technology-enhanced approach has been undeniably influential in teaching and learning today. Despite such influence, the sole reliance on online modes of teaching suffers from some limitations on the hands-on interactions and experiences. That means the combination of online and face-to-face teaching or blended learning, which is also another increasingly emerging area of research, is a practical and appropriate option in today's world as it maximises opportunities for knowledge acquisition and skill development (Whittaker, 2013). Thus, despite the metacognitive approach being advocated, it should not be exclusively treated as the best method for teaching English but as another practical and effective way in combination with other methods for improving teaching and learning.

With the empirical evidence from this study and substantial existing evidence, it is reasonable to believe that promoting metacognition and self-regulated learning is an appropriate approach for EFL pedagogical practice in higher education and can be promoted in higher education policy at an institutional level or wider. On pedagogy, the practice of applying the metacognitive and self-regulatory approach in teaching has been widely employed by tertiary EFL practitioners. However, in the Thai EFL context in particular, the approach has not yet been a common practice. One of the main reasons for the hindrance seems to be the complexity associated with the concept, as the trial teachers expressed. Indeed, many researchers view the notion of metacognition as one of the complex concepts regarding learning development. In fact, it is not the ultimate aim for teachers to convey complex concepts or

terms to the students. What should be emphasised is the strategies and the positive dispositions such as reflective and strategic thinking and self-efficacy which play a pivotal role in improving learning (Moseley et al., 2005) and can be transferred across contexts (Perkins & Salomon, 2012). The awareness-raising and regulation of the available strategies, cognitive, affective and metacognitive, make the approach helpful particularly for less-proficient EFL learners (Cross, 2011; Goh & Taib, 2006; Vandergrift & Tafaghodtari, 2010).

Therefore, for students to apply the approach in the future, there are implications for them to consider. It is important for students to realise that not knowing something and some skills is not a problem for learning. Rather than being discouraged from that deficiency, it is better for them to be aware of what they know and do not know and try to improve by capitalising on what they know, trying different strategies and practicing more (Quigley et al., 2018). Skipping, for example, is not always a bad learning habit if it is used strategically and appropriately. To fulfil their learning goals, self-questioning and self-reflection play an important part (CAIE, n.d.). They help students to realise what they know, what they do not know, what strategies may be relevant and what to do to deal with the encountering difficulties.

As discussed as a drawback of the PMER intervention, the use of unfamiliar technical terms in the guidebook makes students experience difficulty using it. A way to compensate for this drawback is for the learners to focus more on the processes and strategies such as self-efficacy, strategic planning and monitoring and reflection. Interacting with peers during the metacognitive tasks which may be more fully optimised in the in-person or blended settings can help them understand and apply the approach better. Moreover, the use of retrieval practices at the end of the task can help students realise what they have learnt and what remains lacking so that they can improve in the provided independent practices or in their own independent practices.

For teachers, some pedagogical considerations should be noted for an effective and successful implementation of the approach. First, the instructional design should be based on an explicit and clear pedagogical sequence such as the PMER model proposed in this study or the CALLA, MPS, TWA (see Chapter Two) or any other sequence the teachers view as appropriate. An explicit and clear instruction of strategies minimises unnecessarily extra load for students' short-term memory (Ashman, 2015). Sophisticated terms should be kept to a minimum and the amount of content should not be too exhaustive in order not to overwhelm students' cognitive load. In addition, it is inadequate to merely teach metacognitive knowledge

and skills but the opportunities for students to regulate their listening and discuss their strategic application with peers should be adequately provided. Peer discussion and teacher support are helpful especially for less-able learners (Lantolf et al., 2015). These can be promoted via activities such as think-pair-share, group discussions and other collaborative activities.

As teachers, apart from providing sources of knowledge and opportunities for practice, teacher support in the task is also vital. These can range from clear and relevant instruction and explanations, provision of worked examples, modelling of how things work before the task, monitoring and guidance during the task and post-task explanations of how things should have been done. Moreover, another important role for teachers is to support students' agentic role in their own learning by developing their self-efficacy, self-reflection, strategic thinking and self-regulation (Bandura, 1999). Therefore, being aware of the learners' self-regulation when designing a lesson can make the classroom to be a place to promote self-regulated learning and learner agency.

At a policy level, there are multiple valid reasons to advocate the metacognitive and self-regulatory promotion in higher education policy. From the evidence from this study and other sources (e.g., Chou, 2017; Lui & Li, 2015), the metacognitive approach is an effective methodology for improving English learning capacity. On top of that, the metacognitive intervention is economical to implement at a public policy level (EEF, 2018). Moreover, as the approach seems to be more beneficial to lower-capable learners, promoting it as a policy may help reduce the achievement gap. If the approach is to be promoted as a policy, teacher support such as training is required to equip them with better understanding and applicable techniques for instruction. With the reported effectiveness, the approach can be promoted as a policy at the institutional level and is also promising for wider implementation.

No matter at what level the approach is to be advocated as a policy, it is crucial that the application of this approach is robustly evaluated. The inadequacy in the accountability system and evaluation is one of the deficiencies in the education policy in Thailand (UNESCO Bangkok, 2017). Therefore, apart from training on metacognitive instruction, training on evidence-based evaluation should be provided to guide teachers to be more earnest in using evidence to inform their practice. In their attempt to implement the metacognitive instruction or any other interventions, teachers should be able to investigate the effect of their implementation on students' learning or other outcomes, rather than simply following any given models or guidelines. The evidence from their own implementation will help them gain

a better understanding of the approach in their own context which guides further development and can add to the body of research and overall picture of metacognition in EFL settings. In fact, this is essentially how the external validity of social science research is enhanced (Cohen, 1994). Moreover, robust evaluations of an education policy are essential for assuring the effectiveness of the policy in terms of impact and cost-worthiness (Gorard & Siddiqui, 2018).

10.2.2 Exit English Exam Policy

From the interviews with teachers and students, they all agreed with the Exit English policy in principles. Most students viewed the policy as an external force for them to improve their English which is perceived as a helpful asset for their career paths. However, there are some major drawbacks perceived by some of the student and teacher participants which should be considered for the betterment of the policy. Firstly, the forms of assessment which qualify for the graduation requirement should be more varied and competence-based. Relying only on standardised tests can cause extra anxiety to students. Contextualised concession suggested by Boliver et al. (2017) for widening participation in HE among disadvantaged students can be considered in this policy. Apart from standardised English tests, other measures such as completing immersion or intensive English activities and passing extra English training should be acceptable, especially for students with poor academic backgrounds. As the trial was set in the southernmost areas where there are students who are competent in a second language other than English such as Malay, Arabic and Chinese, allowing competence in other second languages as an alternative can be a reasonable option. Whether it is English or other languages, foreign language competence is desirable intercultural citizenship in the current globalised age (Byram, 2008). Moreover, varying the forms of acceptable assessment and including more competence-based assessment such as accomplishment in English-medium projects, competitions, activities, etc., can allow students to involve more actively in competence development.

Secondly, the support system is highly crucial for policy implementation. As expressed by some trial teachers, sufficient support for teachers and students was lacking. Instead of launching a policy as a set of standards one should follow, the policy should be accompanied with mechanisms to support its implementation (Trowler, 2003). Teacher development support such as in-house or overseas training for teachers to teach English in English effectively is suggested by some of the interviewed teachers. Governmental and university support for students for authentic English use is also vital. One teacher suggested that students need

opportunities to involve in situations where they really have to use English for practical, functional purposes. The trial participants live in a disadvantaged context where such opportunities are extremely rare, compared to many parts of the country. The government policy which is imposed across the country has different effects on students in different contexts and those in such disadvantaged areas seem to be under a higher challenge. Therefore, need-based support for the different stakeholders is important to enhance equitable opportunities for development (Unterhalter, 2009).

In addition, another teacher suggested that metacognitive intervention is useful and should be applied by more teachers and students. In the disadvantaged contexts like the one where this study was conducted, this kind of intervention which promotes self-belief is particularly appropriate (Good et al., 2003). The results of this study and many others (e.g., Goh & Taib, 2006; Vandergrift & Tafaghodtari, 2010) suggested that the metacognitive and self-regulatory interventions are more beneficial to less-able learners. As the secondary data shows that most of the population at one of the higher education institutions in this region have low proficiency in English, an intervention raising their awareness of self-regulatory strategies, self-efficacy, reflective and strategic thinking can be helpful support for students to develop their competence to satisfy the Exit English Exam policy requirement.

Indeed, the use of evidence in policy planning and evaluation is essential for the Exit English Exam policy or any other education policy. Since the launch of the policy in 2016, no robust evidence-based evaluation of the policy has been published to date. The lack of a rigorous accountability system for education policy has been one critical threat to the country's educational development (UNESCO Bangkok, 2017). Therefore, apart from the more comprehensive policy development and appropriate support system, the policy makers must seriously take policy evaluation into account. Robust evaluations which rigorously examine the reported findings and evidence are crucial for assuring the effectiveness of education policy and practice as they provide credible evidence for further policy planning, research and practice (Gorard & Siddiqui, 2018).

10.3 Chapter summary

Despite multiple attempts to maintain the validity of the study, this chapter acknowledges and discusses limitations related to the small number of clusters in the cluster RCT, possible bias from having one teacher teaching in two conditions, the limited time of the intervention for students to actually involve in the metacognitive and self-regulatory tasks and

the low response rate of the questionnaire, especially from the non-randomised comparison group.

From the discussed findings in previous chapters and the acknowledged limitations, implications were proposed for students, teachers and policy makers about the application of metacognitive intervention in EFL contexts. Students are encouraged to focus on self-questioning, reflection and application of various strategies to address the difficulties they experience while performing a task. Teachers should be reminded to have a clear pedagogical sequence and relevant examples and focus on regulating the metacognitive competence such as self-efficacy and strategic and reflective thinking, rather than being obsessed with the fancy terms. At the policy levels, training and support are recommended to equip the teacher with the fundamental understanding and applicable strategies regarding metacognitive instruction. Training on using evidence to inform their practice is also vital so that the teachers can build up evidence-based capacity for their professional development.

CHAPTER ELEVEN

CONCLUSION

This chapter provides a summary of the concept of metacognition, empirical study and key research findings. It is followed by a succinct synopsis of the academic contributions of the study. Next, suggestions for future research regarding research design, intervention design and policy development are provided. The chapter ends with concluding thoughts which encapsulate the highlight findings and direction for research on policy and practice in English language teaching.

11.1 Summary of the study

Among the attempts to improve learning, the metacognitive and self-regulatory approach has been widely advocated. It has also been widely applied in English language education. The increasingly important role of English in education policy worldwide makes proficiency in English a desirable asset or requirement for university graduates. To support Thai learners of English, a metacognitive and self-regulatory approach to teaching and learning can be an appropriate measure. With the recent announcement of the Exit English Exam policy for universities in Thailand, students in the southernmost areas who are segregated by low English proficiency and socioeconomic backgrounds need support for coping with the new challenge. Thus, a metacognitive intervention can be helpful support for this group of learners.

To examine the existing evidence of metacognitive interventions in tertiary EFL contexts, a systematic review was conducted. The total records of 2,942 were identified from seven online data sources. The PICO criteria, i.e., population, intervention, comparison and outcome were used to screen the studies. Finally, 29 studies were included in the analysis which is based on the mixed methods involving meta-analysis and thematic synthesis. The findings suggest that the intervention has a high potential for improving learning among EFL students. It also found a need for studies on the topic to pay more attention to key research elements, namely, design, sample size and validity measures. The results from the thematic synthesis provide practical suggestions for the revision and implementation of the intervention in the main trial.

Simply defined as 'thinking about thinking', metacognition in this study is operationally defined as *the learners' awareness of their own knowledge, experiences and emotions and the ability to apply that awareness to regulate their thinking while learning or accomplishing a task*. Two main components of metacognition are metacognitive knowledge

and metacognitive regulation (Veenman et al., 2006). Metacognitive knowledge includes knowledge of oneself, the encountered tasks and available relevant strategies (Pintrich 2002) while metacognitive strategies are the processes the learners operate to regulate their learning through planning, monitoring and evaluating their own learning (CAIE, n.d.). Metacognitive awareness interacts with the cognitive and affective strategies during the learning processes (O'Malley & Chamot, 1990).

In this study, the Plan, Monitor, Evaluate and Retrieve (PMER) model was designed to use with the participants in tertiary institutions in southernmost Thailand. The first three processes in the model are the most common metacognitive strategies and strategy sets under each process were designed with language learners in mind and expressed in a mnemonic manner to aid memory. Metacognitive Pedagogical Sequence, in line with Vandergrift and Goh (2012) was adapted to make the model resonate with classroom practices. Retrieval practice is a key distinctive feature of the model to enhance the transfer of learning. Lessons from the pilot study and systematic reviews, such as providing class materials and practice kits for students outside the classroom and promoting collaborative learning, were considered and integrated into the main trial.

The first research question examined the impact of the metacognitive intervention on the participants' listening comprehension and overall English proficiency. This was investigated by a cluster randomised controlled trial with standardised English tests as a pre/post outcome measure. The results found medium effect sizes, supporting the impact of the intervention on both listening comprehension and overall English proficiency. The multi-strategy components, self-reflection and retrieval practices are attributed as the features which contribute to the impact of the intervention.

The third research question provides further explanations of the impact of the intervention on the primary outcomes as it considers the relationship between the impact and the participants' background characteristics. This was explored by a regression analysis and a subgroup analysis of standardised mean differences to determine if their characteristics contribute to the change in the outcome variables. It was found that the impact on the primary outcomes was not explained by the participants' backgrounds except the pre-test score. In addition, it was found that it is the lower proficiency students who make greater improvement in the gain outcomes. In the multiple regression analysis, the group variable which was added lastly in the regression model still gives a slight improvement, indicating the impact of the

intervention above other variables. This reaffirms the impact of the intervention confirmed in the first research question.

In the subgroup analysis of the standardised mean difference among the intervention students, it was found that the intervention students who had lower pre-test scores made a higher improvement than their more proficient peers. This is in line with the results from regression analysis. This seems to suggest that the intervention has more impact on the lower proficiency students at intermediate and pre-intermediate levels. This could be due to the design and the components of the PMER intervention which incorporates features reported to be particularly helpful for less-ability learners, i.e., awareness-raising of different types of strategies, the promotion of self-affirmation and self-efficacy and retrieval practices. As the pre-test results (Table 7.9) suggest, the intervention group was predominated by low-proficient students. Therefore, the intervention design which offers features directly relevant to low-proficient learners who are the majority of the intervention group can be another contribution to the overall impact.

The secondary outcome of the study enquired in the second research question is the learners' metacognitive awareness in listening. A questionnaire based on Vandergrift et al. (2006) was used as an outcome measure. The results turned out to differ from most studies on the topic because evidence of the impact of the intervention on the students' metacognitive awareness appeared to be rather mixed and ambiguous. The intervention even had a negative result on improving negative listening behaviours. The accuracy of judgement analysis shows that the participants in all groups have low accuracy of their metacognitive judgement. It also suggests that the post-trial judgements are slightly more accurate, providing a justification for examining the impact from the post-trial responses. From the post-response effect sizes, only the comparison between the intervention and control groups has the results secure from the missing data. The results indicate a small impact of the intervention on positive metacognitive behaviours, namely, planning and evaluation, directed attention and problem solving but it is ineffective to minimise the negative behaviour of mental translation and the perceived difficulty of English and English listening.

The process evaluation which primarily seeks to answer research question 4 provides further insights into the impact evaluation through multiple methods to obtain more in-depth data. From classroom observations, no apparent fidelity threat was observed. Overall, most students hold positive attitudes towards the intervention. Prevalent benefits of the intervention

perceived by the students are strategic listening, goal setting, planning and improving confidence. Similarly, the teachers viewed planning strategies helpful for students to feel more ready and maintain their focus better. However, the difficult terms, lengthy details of the principle phase and the repetitive patterns of the sessions were pointed as the drawbacks of the intervention which may have hindered some anticipated outcomes. Thus, it is advisable that the regulation of metacognitive skills should be given more time and emphasis to allow more opportunities for students to work with peers to actually apply the learnt strategies in the learning tasks.

The secondary data analysis was drawn from larger non-trial participants from the same population. The whole cohort data suggest that there are segregations by low income and low academic performance. The comparison of university test pass rates before and after the trial between the intervention students and the full cohort participants shows that a higher percentage of the intervention students satisfy the English requirement for graduation. The findings reaffirm the impact of the PMER intervention and indicate that metacognitive and self-regulatory promotion programmes can be an efficient way to support EFL learners, especially those with intermediate proficiency levels or below. As a higher percentage of the intervention students can satisfy the English requirement, it could be said that metacognitive development programmes can promote equity in education by helping students with less privileged backgrounds to achieve their education goal, complying with the United Nations' Sustainable Development Goal (SDG) 4, aiming at more inclusive and equitable education for all.

The study also examined the consequences of the Exit English Exam policy and found interesting findings about the implementation of the policy in the marginalised areas. All interviewed teachers and students agreed with the policy in principle, seeing that it encourages students to improve their English skills which will be useful for their study and future career paths. This accords with the policy aim proposed by the policy makers. However, some of the participants voiced criticisms about the way the policy was implemented. Firstly, the policy lacks adequate support mechanism for its efficient implementation especially among the stakeholders in disadvantaged areas. Indeed, the conflicting interests and value systems between policy makers and those responsible for taking action need to be considered in the policy development process (Trowler, 2003). The different needs of the stakeholders should be considered to provide equitable opportunities which can bridge the inequality gap caused by their socio-economic and geographical backgrounds (Ulterhalter, 2009). Teacher development

support was recommended by the interviewed teachers. Moreover, some teachers and students suggested that the accepted assessment forms be more varied to cater for different groups of students. As Boliver et al. (2017) suggested, a contextualised concession may help enhance opportunities to remain in higher education for those with less privileged backgrounds. In line with this suggestion, competence in other second languages such as Malay, Arabic or Chinese should also be eligible for satisfying the requirement of language competence.

11.2 Academic contributions of the study

The results of the systematic review in this study affirm the benefits of metacognition and self-regulation in EFL contexts. The important features which metacognitive interventions should consider are explicit and clear instructional procedure, focus on the regulation of skills and provision of materials for students' practice. These are practical pedagogical implications for practitioners and researchers. However, the evidence of its effectiveness is not secure. This emphasises the need for studies on this topic to pay greater attention to essential research elements to derive robust evidence on effectiveness. It also calls for studies in the field which concerns how effective a programme or an intervention is to be earnest about the internal validity of their studies, extending its external validity via replication and reviewing the exiting evidence via a systematic approach.

From the main trial which was the largest trial on the topic at the time it was conducted, the metacognitive and self-regulatory approach was found to be effective for improving the language learning outcomes of EFL learners in the southernmost areas of Thailand. The results support the mixed-method approach to metacognitive instruction which combines general discussion of the concept and clear links of the idea to English language listening. The subgroup analysis provided further information that the approach can be more beneficial to lower proficiency students who have pre-intermediate and intermediate levels. Thus, the metacognitive and self-regulatory approach may also be a potential measure to decrease the achievement gap. Moreover, the data from the process evaluation reveals a particular set of strategies which the participants found more obviously beneficial than some others, i.e., strategic listening, goal setting and planning, self-efficacy, focusing attention and making inferences.

Moreover, the main study proposes some theoretical and pedagogical information for the teaching of listening comprehension which seems to be underresearched in second language settings (Chou, 2017; Yabukoshi, 2021) and is weakly principled (Graham & Santos, 2015).

The metacognitive intervention seeks to provide a holistic way to develop listening comprehension which considers both linguistic and non-linguistic competence (Vandergrift and Goh, 2012). It raises students' awareness of various types of strategies, namely, listening strategies, cognitive strategies, affective strategies and metacognitive strategies. These strategies can be more supportive to the learners in their attempts to approach the listening tasks than the common three-stage teaching which seems to primarily practise and test comprehension.

The study also shines more light on some debates in teaching thinking. First, it affirms the benefit of explicit teaching. The explicit approach can reduce unnecessary cognitive load so that the working memory can be more fully utilised for the learning task itself (Ashman, 2015). Explicit teaching also helps students to view the instructional sequence as a systematic procedure. Second, the study provides further evidence that a thinking approach such as metacognition can be taught to learners beyond Western cultures. This is also affirmed in other studies (e.g., Habibian, 2015; Farahain & Avarzamani, 2018).

11.3 Suggestions for future research

The suggestions in this section will be different from what has been discussed in the implications in the previous chapter. It proposes suggestions for future research. The suggestions are on the design and implementations of studies on the topic, the design of interventions and suggestions for evaluation research of the Exit English Exam policy.

As suggested by the systematic review which is a part of this thesis, research on this topic needs to concern more about research elements. To provide reliable evidence, studies should rely on a suitable research design for the research question. For effectiveness evaluation, the designs which allow counterfactual evidence such as randomised controlled trials, regression discontinuity designs or comparative designs should be considered. Moreover, studies should work on a larger scale and take appropriate care of validity issues. In terms of analysis, the report of effect size which is encouraged by the APA should be included to estimate the effect of the intervention.

As metacognition is a complex concept, it is quite challenging to design an intervention which adequately encompasses fundamental concepts and remains friendly for the general users. The complexity and interpretations of the concept highly align with the discipline they are used in (Veenman et al., 2006). As observed from the literature review, metacognition research from psychology seems more complicated than from educational studies. To apply the

concept in educational contexts, this study adopted a mixed approach to metacognitive instruction by introducing a general discussion of metacognition and self-regulation and linking it to English listening. The pedagogical sequences were designed to accommodate the metacognitive processes, namely, planning, monitoring and evaluating. The intervention guidebook has substantial details with the aim to provide the users with adequate background information. However, such details were perceived by the participants as exhaustive. Moreover, the technical terms related to metacognitive processes caused some negative impression to some students. In short, the design of metacognitive and self-regulatory interventions should consider an appropriate combination of general domain and subject specific aspects of metacognition, supportive pedagogical sequence, appropriate amount of detail and suitable levels of the terms used in the intervention.

If the intervention is to be used with learners with different proficiency levels including those of low proficiency, an adequate proportion of bottom-up strategies for perception of what is heard from listening should be considered in the design of the intervention. As the results from the questionnaire and interviews in the process evaluation suggested, some intervention students, despite being a minority, did not have a positive perception about the intervention. Some of them reported lacking sufficient knowledge and skills to employ the strategies suggested in the intervention. Therefore, more training on object-level knowledge and skills such as features of sounds in natural connected speech could be helpful as both bottom-up and top-down processing are essential in listening comprehension (Graham & Macaro, 2008). This can be particularly supportive for low-proficiency learners who might lack such competence and tend to find metacognitive strategies more difficult or irrelevant (Quigley et al., 2018).

Another interesting point to consider regarding the design of the intervention is to integrate multiple language skills. In this study, the intervention primarily targeted listening comprehension which relatively has less research on, compared to other skills (Vandergrift, 2007; Yabukoshi, 2021) and has less clear methodology which really undertakes teaching rather than testing listening (Field, 2008). As suggested by the trial teachers, the ideas from the PMER intervention can be applied to reading lessons. This is because both are comprehension skills which involve several similar processes (Pressley, 2002). Moreover, productive skills such as speaking or writing could be integrated to provide a more comprehensive approach which employs metacognitive and self-regulatory principles to enhance language development. Even though the skill sets in each language area may be dissimilar, the metacognitive processes, i.e., planning, monitoring and reflection can be applied in all language skills as

reported in many studies such as in reading (e.g., Macaro & Lynn, 2008; Mason, 2013), writing (e.g., Teng, 2016) and speaking (e.g., Seifoori, 2016).

Assessment measures are also an important area for consideration for future studies which will include metacognitive awareness as one of the target outcomes. Among the common methods used to assess metacognition such has questionnaires, interviews, think-aloud protocols, observations and stimulated recalls, this study employed a questionnaire which is the most widely used tool. Each assessment method has its advantages and drawbacks (Veenman et al., 2006). The results drawn from the questionnaire tool in this study experienced ambiguity regarding how accurate the participants' responses were. Although the study included open-ended questions, interviews and observations which provided more information about the participants' metacognitive awareness, these methods were not specifically designed for assessing the outcome. Therefore, it might be reasonable for future studies to include multiple fine-tuned tools for metacognitive assessment with both off-line and online formats if possible (Saraç & KaraKelle, 2012).

The increasing necessity to employ technology in teaching influenced by the restrictions during the COVID 19 pandemic and the vast varieties of technology on offer made teachers' capacity for blended or online teaching more essential. Therefore, future research on metacognition and self-regulation may consider optimal approaches to incorporate technology in developing self-regulatory competence. Alternatively, the use of technology for supporting self-regulation in relation to different factors such as age, gender, school types or other socio-economic backgrounds may be investigated.

A part of the study looked at how the intervention helps students cope with the new language requirement demanded by the Exit English Exam policy. Via process evaluation, the participants were asked about their opinions toward the policy. The findings suggested that participants supported the policy in principle but voiced some criticisms about how it was implemented. However, the interviews were conducted with a small number of participants. Hence, studies to rigorously evaluate the impact of this policy are needed. Such studies should include essential research elements discussed several times earlier, such as design, scale and standardised measurement, to obtain credible evidence of how effective the policy has been in fulfilling its aim of improving the English proficiency of university graduates and how much it affects the relevant stakeholders.

11.4 Concluding thoughts

To end this final chapter, the researcher would like to encapsulate some of the headline findings in brief as a takeaway as the following.

- Metacognition is a beneficial intervention for learners of English as a foreign language for improving their language outcomes. It can be particularly useful for non-advanced learners such as those with intermediate and pre-intermediate proficiency. To apply this approach, positive dispositions such as self-efficacy and reflective thinking should be emphasised.

- The effect of the intervention on metacognitive awareness is unclear but some findings which withstood the missing data show a slight effect on positive behaviours but no effect for reducing negative behaviours. Mixed methods of assessment tools may be included in future research for clearer evidence.

- No clear evidence demonstrates a negative association between the intervention students' language outcomes and their socioeconomic backgrounds which are usually viewed as inferior such as low parental income, background in private Islamic schools or foreign mother tongues.

- Teacher training about the approach and how to implement it is required. It can be before commencing the intervention or as a way of Continued Professional Development (CPD). Teacher training should also be on using evidence to inform their pedagogical practices.

- The Exit Exam Policy is based on a reasonable aim but should be accompanied with an appropriate support system. Such supports are especially crucial for participants in less privileged areas. In addition, more serious attention should be given to policy evaluation.

The metacognitive intervention in this study is found to be effective for improving language outcomes. As emphasised throughout the thesis, the study has paid enormous attention to the internal validity but does not aim to generalise the findings. Indeed, a proper way to enhance the external validity of social research is through replication. Metacognitive interventions have been widely applied earlier in EFL contexts but the evidence remains unclear due to some major flaws in those studies. With careful attention to the design, scale, attrition, outcome measure and validity threats, the findings on the primary outcome of this study are considerably secure. However, to have a more comprehensive understanding of the

role of the approach in an overall EFL realm, replications of metacognition-based research based on all key research elements in wider EFL contexts are still required. More high-quality primary studies would allow evidence syntheses to have more reliable results which would provide a more comprehensive picture of the evidence in the field. Hopefully, advocation of these research practices will minimise the 'anything works' culture in ELT research and provide evidence which would bring about the real impact and development in English language education in Thailand and the EFL contexts at large.

REFERENCES

Note: The references with an asterisk are studies included in the systematic review.

*Abdelhafez, A. 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*. Doctoral Dissertation., University of Exeter.

Abrami, P., Bernard, R., Borokhovski, E., Wade, A., Surkes, M., Tamim, R. and Zhang, D. (2008). Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis. *Review of Educational Research*, 78(4), 1102-1134.

Agarwal, P. K., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, a principal, and a scientist. *Educational Psychology Review, 24*(3), 437-448.

Agarwal, P. K., Finley, J. R., Nathan S. Rose, N. S. & Roediger III, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory, 25*(6), 764-771.

*Al-Ghazo, A. (2016). The effect of explicit instruction of metacognitive learning strategies on promoting Jordanian language learners' reading competence. *Journal of Education and Practice,* 7(10), 170-177.

*Altay, I. F. & Altay, A. (2017). The impact of online reading tasks and reading strategies on EFL learners' reading test scores. *Journal of Language and Linguistic Studies, 13*(2), 136-152.

Ameri, M. (2020). Criticism of the sociocultural theory. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal), 3*, 1530-1540.

An, X. and Shi, Z. (2013) Does Metacognitive Instruction Improve Listening Comprehension. *Theory and Practice in Language Studies,3* (4), 632-636.

Anderson, J.R. (1995). *Cognitive psychology and its implications* (2nd ed). New York: Freeman.

Anderson, J.R. (2005). *Cognitive psychology and its implications* (6th ed). New York: Worth Publishers.

Anderson, L. W. & Krathwohl, D. R. (ed.). (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Anderson, N. J. (2002). The role of metacognition in second language teaching and learning. *ERIC Digest*. Washington, DC: Center for Applied Linguistics.

Anderson, N. J. (2008). Metacognition and good language learners. In C. Griffiths (ed), *Lessons from Good Language Learners* (pp.99-109). Cambridge: Cambridge University Press.

Ashman, G. (2015). Why Students Make Silly Mistakes in Class (and what can be done). Retrieved September, 10, 2019 from: https://theconversation.com/why-students-make-silly-mistakes-in-class-and-what-can-be-done-48826

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (eds.), *The psychology of learning and motivation: Advances in research and theory* (vol. 2, pp. 89–195). New York: Academic Press.

Aubrey, K. & Riley, A. (2019). *Understanding and using educational theories* (2nd ed). London: Sage.

Baker, W. (2012). English as a lingua franca in Thailand: Characterisations and implications. *Englishes in Practice, 1*, 18–27.

Baker, W., & Jarunthawatchai, W. (2017). English language policy in Thailand. *European Journal of Language Policy, 9*(1): 27–44.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice- Hall, Inc.

Bandura, A. (1991a). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In R. A. Dienstbier (Ed.), *Nebraska Symposium on Motivation, 1990: Perspectives on motivation* (pp. 69–164). University of Nebraska Press.

Bandura, A. (1991b). Social cognitive theory of self-regulation. *Organizational Behaviour and Human Decision Processes, 50*, 248–287.

Bandura, A. (1999). Social cognitive theory: An agentic perspective. *Asian Journal of Social Psychology, 2*(1), 21-41.

Başal, A., & Eryılmaz, A. (2021). Engagement and affection of pre-service teachers in online learning in the context of COVID 19: engagement-based instruction with web 2.0 technologies vs direct transmission instruction. *Journal of Education for Teaching, 47*(1), 131-133.

Bell, L., & Stevenson, H. (2006). *Education policy: Process, themes and impact*. New York: Routledge.

Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. *International Journal of Exercise Science, 8*(3), 297-302.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263.

Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. European psychologist, 1(2), 100-112.

Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research*, *31*(6), 445-457.

Boliver, V., Gorard, S., & Siddiqui, N. (2017). A more radical approach to contextualised admissions. In *Where next for widening participation and fair access? New insights from leading thinkers*. Oxford: Higher Education Policy Institute.

Boliver, V., Gorard, S., & Siddiqui, N. (2021). Using contextual data to widen access to higher education. *Perspectives: Policy and Practice in Higher Education*, 25(1), 7-13.

Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chicester: Wiley.

Bown, J., 2009. Self-regulatory strategies and agency in self-instructed language learning: a situated view. *The Modern Language Journal, 93* (4), 570-583.

Bozorgian, H. (2015). Less-skilled learners benefit more from metacognitive instruction to develop listening comprehension. *International Journal of Research Studies in Language Learning, 4*(1), 3-12.

*Bozorgian, H. & Alamdari, E. F. (2017). Multimedia listening comprehension: Metacognitive instruction or metacognitive instruction through dialogic interaction. *ReCALL, 30*(1), 131–152.

Bråten, I. (1991). Vygotsky as Precursor to Metacognitive Theory: I. The Concept of Metacognition and Its Roots. *Scandinavian Journal of Educational Research, 35*(3), 179-192.

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

British Educational Research Association (BERA) (2018). Ethical guidelines for educational research (4th ed). Retrieved August 10, 2019, from: https://www.bera.ac.uk/ researchers-resources/publications/ethical-guidelines-for-educational-research-2018

Brooks, M. C., & Sungtong, E. (2015). The challenge of multicultural education during insurgency: The case of Southernmost Thailand. In G. Jean-Marie, S. Sider, and C. Desir (Eds), *Comparative International Perspectives on Education and Social Change in Developing Countries and Indigenous Peoples in Developed Countries* (pp.187–197). Charlotte, NC: Information Age Publishing.

Brown, A. L. (1977). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 77–165). Hillsdale: Erlbaum.

Brown, D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed). NJ: Prentice Hall Regents.

Bryman, A. (2012). *Social Research Methods* (4th ed). Oxford: Oxford University Press.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 918.

Byram, M. (2008). *From foreign language education to education for intercultural citizenship: Essays and reflections*. Clevedon: Multilingual Matters Ltd.

*Cabrera-Solano, P., Gonzalez-Torres, P., Solano, L. & Castillo-Cuesta, L. (2019). Using visual vocabulary app as a metacognitive strategy in EFL learning: A case of pre-service teachers. *International Journal of Learning, Teaching and Educational Research,* 18(12), 302–315.

Callan, G. L., Marchant, G. J., Finch, W. H., & German, R. L. (2016). Metacognition, strategies, achievement, and demographics: Relationships across countries. *Educational Sciences: Theory & Practice, 16*(5).

Cambridge Assessment International Education (n.d.). *Getting started with metacognition*. Retrieved June 20, 2020, from https://cambridge-community.org.uk/professional-development/gswmeta/index.html

Campbell, M., Mollison, J., Steen, N., Grimshaw, J., & Eccles, M. (2000). Analysis of cluster randomized trials in primary care: a practical approach. *Family Practice*, *17*(2), 192-196.

Centre for Education Statistics and Evaluation (CESE) (2017). *Cognitive load theory: Research that teachers really need to understand*. New South Wales Department of Education, Centre for Education Statistics and Evaluation. Retrieved July 19, 2020, from https://education.nsw.gov.au/content/dam/main-education/about-us/educational-data/cese/2017-cognitive-load-theory.pdf

Chamot, A. U. (2008). Strategy instruction and good language learners. In C. Griffiths (ed*), Lesson from Good Language Learners* (pp.266-281). Cambridge: Cambridge University Press.

Chamot, A. U., & O'malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. New York: Addison-Wesley Publishing Company.

*Chang, M. M. (2010). Effects of self-monitoring on web-based language learner's performance and motivation. *CALICO Journal,* 27(2), 298–310.

*Chou, M.H. (2017). A Task-based Language Teaching Approach to Developing Metacognitive Strategies for Listening Comprehension. *International Journal of Listening*, 31(1), 51-70.

Cleary, T. J., & Zimmerman, B. (2004). Self-regulation empowerment programme: A school-based program to enhance self-regulated and self-motivated cycles of student learning. *Psychology in the Schools, 41*(5), 537–50.

Coe, R. (2002). It's the effect size, stupid: what effect size is and why it is important. Paper presented at the Annual Conference of the British Educational Research Association,

Exeter, 2002. Retrieved December 21, 2019. http://www.leeds.ac.uk/educol/ documents/00002182.htm.

Coe, R. (2004). Issues arising the use of effect sizes in analysing and reporting research. In I. Schagen and K. Elliot (eds), *But what does it mean? The use of effect sizes in educational research* (pp.80-100). Berks: The National Foundation for Educational Research.

Cohen, A. D. (2014). *Strategies in learning and using a second language* (2nd ed). New York: Addison-Wesley Longman.

Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324*(5925), 400-403.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cohen, J. (1994). The Earth is Round (p<.05). *American Psychologist, 49,*(12), 997-1003.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8th ed.) London: Routledge.

Crandall, J. J., & Bailey, K. M. (2018). Introduction. In J. J. Crandall & K. M. Bailey (eds.), *Global perspectives on language education policies* (pp.1-20). Oxon: Routledge.

Cross, J. (2011). Metacognitive instruction for helping less-skilled learners. *ELT Journal, 65*(4), 408-416.

Crystal, D. (2003). *English as a Global Language* (2nd ed.). Cambridge: Cambridge University Press.

Dabarera, C., Renandya W., & Zhang, L. (2014). The impact of metacognitive scaffolding and monitoring on reading comprehension. *System, 42*, 462–473.

Darasawang, P., & Watson Todd, R. (2012). The effect of policy on English language teaching at secondary schools in Thailand. *English in Southeast Asia: Features, policy and language in use*, *42*(207-220).

de Bruijn-Smolders, M., Timmers, C. F., Gawke, J. C., Schoonman, W., & Born, M. P. (2016). Effective self-regulatory processes in higher education: research findings and future directions. A systematic review. *Studies in Higher Education, 41*(1), 139-158.

De Vaus, D. A. (2001). *Research Design in Social Research*. London: Sage.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*(3), 231-264.

Dong, N. & Lipsey, M. (2011). Biases in estimating treatment effects due to attrition in randomised controlled trials: A Simulation study. SREE Conference, 2011.

Dörnyei, Z. (2001). *Motivational Strategies in Language Classroom*. Cambridge: Cambridge University Press.

Dörnyei, Z. (2005). *The Psychology of Language Teaching*. Mulwah, NJ: Erlbaum.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J. and Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14* (1), pp. 4–58.

Durham University (2017). Research Integrity Policy and Code of Good Practice. Retrieved August 20, 2019from https://www.dur.ac.uk/research.innovation/governance /policy/integrity/

Durham University (2018). Ethics policy. Retrieved August 20, 2019 from https://www.dur. ac.uk/research.innovation/governance/policy/ethics/

Dweck, C. S. (2009). *Mindset: The new psychology of success*. Random House Digital, Inc.

Early, E., Miller, S., Dunne, L., Thurston, A., & Filiz, M. (2020). The influence of socio-economic background and gender on school attainment in the United Kingdom: A systematic review. *Review of Education, 8*(1), 120-152.

Education Endowment Foundation (EEF). (2018). *Teaching and Learning Toolkit*. Retrieved July 15, 2019, from https://educationendowmentfoundation.org.uk/public/ files/Toolkit/complete/EEF-Teaching-Learning-Toolkit-October-2018.pdf

Education Endowment Foundation (EEF). (2019). *Classification of the Security of Findings from EEF Evaluations*. Retrieved July 15, 2020, from https://educationendowment foundation.org.uk/public/files/Evaluation/Carrying_out_a_Peer_Review/ Classifying_the_security_of_EEF_findings_2019.pdf

Education in Thailand's Deep South: not a Secular Matter. (2017, April 20). Prachathai. https://prachatai.com/english/node/7087

EF English Proficiency Index. (2019). *Thailand*. Retrieved November 20, 2019, from https:// www.ef.com/wwen/epi/regions/asia/thailand/

Efklides, A. (2008). Metacognition: Defining Its Facets and Levels of Functioning in Relation to Self-Regulation and Co-regulation. *European Psychologist, 13*(4), 277-287.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL Model. *Educational Psychologist, 46*, 6-25.

El Soufia, N. and See, B. H. (2019). Does explicit teaching of critical thinking improve critical thinking skills of English language learners in higher education? A critical review of causal evidence. *Studies in Educational Evaluation, 60*, 140-162.

Elliot, K. and Sammons, P. (2004). Exploring the use of effect sizes to evaluate the impact of different influences on child outcomes: Possibilities and limitations. In I. Schagen and K. Elliot (eds), *But what does it mean? The use of effect sizes in educational research* (pp.6-24). Berks: The National Foundation for Educational Research.

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press.

Ellis, R. (2003). *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

English skills drop again (2020, Nov 27). The Bangkok Post. https://www.bangkokpost.com /thailand/general/2026031/english-skills-drop-again

Ennis, R. H. (1962). A concept of critical thinking: A proposed basis for research in the teaching and evaluation of critical thinking ability. *Harvard Educational Review*, *32*(1), 81–111.

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18*(3), 4–10.

EPPI-Centre. (2003). *Review Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research Version 0.9.7*. London: EPPI-Centre, Social Science Research Unit.

*Farahian, M., & Avarzamani, F. (2018). The impact of portfolio on EFL learners' metacognition and writing performance. *Cogent Education, 5*: 1450918. https://doi.org/10.1080/2331186X.2018.1450918

Feyten, C. M. (1991). The power of listening ability: An overlooked dimension in language acquisition. *Modern Language Journal 75*, 173-80.

Field, J. (2008). *Listening in the language classroom*. Cambridge, England: Cambridge University Press.

Field, S., Kuczera, M. & Pont, B. (2007). *No More Failures: Ten Steps to Equity in Education*. OECD, Paris.

Flavell, J. H. (1976). Metacognitive Aspects of Problem Solving. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231-235). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.

Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive Developmental Inquiry. *American Psychologist*, *34* (10), 906–11.

Flavell, J. H. (1996). Piaget's legacy. *Psychological Science, 7*(4), 200-203.

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge: Cambridge University Press.

Franz, J., & Teo, A. (2018). 'A2 is Normal'–Thai Secondary School English Teachers' Encounters with the CEFR. *RELC Journal, 49*(3), 322-338.

Fry, G. W., Bi, H., and Apahung, R. (2018). Regional Educational Disparities in Thailand. In G. W. Fry (ed.), *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp.373-391). Singapore: Springer.

Gao, X., (2010). *Strategic Language Learning: The Roles of Agency and Context*. Multilingual Matters: Bristol.

Gascoine, L., Higgins, S. and Wall, K. (2017). The assessment of metacognition in children aged 4-16 years: a systematic review. *Review of education.*, *5* (1). pp. 3-57.

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

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

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

Goh, C. C. & Hu, G. (2014). Exploring the relationship between metacognitive awareness and listening performance with questionnaire data. *Language awareness, 23*(3), 255-274.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24*, 645–662.

Gorard, S. & See, B.H. (2013). *Overcoming disadvantage in education*. London: Routledge.

Gorard, S. & Siddiqui, N. (2018). There is Only Research: The Liberating Impact of Just Doing Research. *International Journal of Multiple Research Approaches, 10*(1), 328–333.

Gorard, S. (2010). All evidence is equal: the flaw in statistical reasoning. *Oxford Review of Education, 36*(1), 63-77

Gorard, S. (2012). Statistical and correlational techniques. In Arthur, J., Waring, M., Coe, R., & Hedges, L.V. (Eds), *Research Methods & Methodologies in Education* (pp.23-30). London: Sage.

Gorard, S. (2013). *Research Design: Creating Robust Approaches for the Social Sciences*. London: Sage.

Gorard, S. (2014). A proposal for judging the trustworthiness of research findings. *Radical Statistics, 110*, 47–59.

Gorard, S. (2018). *Education Policy: Evidence of Equity and Effectiveness*. Bristol: Bristol University Press.

Gorard, S. (2021). *How to Make Sense of Statistics*. London: Sage

Gorard, S., & Gorard, J. (2016). What to do instead of significance testing? Calculating the 'number of counterfactual cases needed to disturb a finding'. *International Journal of Social Research Methodology*, *19*(4), 481-490.

Gorard, S., See, B.H. & Siddiqui, N. (2017). *The Trials of Evidence-Based Education*. Abingdon, Oxon: Routledge.

Gough, D., Oliver, S. & Thomas, J. (2012). Introducing systematic reviews. In D., Gough, S., Oliver and J, Thomas (Eds), *An Introduction to Systematic Reviews* (pp.1-16). London: SAGE.

Graham, S. & Santos, D. (2015). *Strategies for Second Language Listening: Current Scenarios and Improved Pedagogy*. Basingstoke: Palgrave Macmillan.

Graham, S. (2007). Learner Strategies and Self-Efficacy: Making the Connection. *Language Learning Journal 35*, (1), 81–93.

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

Graham, S., Santos, D., & Vanderplank, R. (2011). Exploring the relationship between listening development and strategy use. *Language Teaching Research, 15*(4), 435-456.

Griffiths, C. (2008). Strategies and good language learners. In C. Griffiths (Ed), *Lessons from Good Language Learners* (pp.83-98). Cambridge: Cambridge University Press.

*Habibian, M. (2015). The impact of training metacognitive strategies on reading comprehension among ESL learners. *Journal of Education and Practice, 6*(28), 61-69.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (2009). A growing sense of "agency". In D. J. Hacker., J. Dunlosky. and A. C. Graesser (eds.), *Handbok of Metacognition in Education* (pp.1-4). Oxon: Routledge.

Haddad, W. D., & Demsky, T. (1995). Education Policy-Planning Process: An Applied Framework. *Fundamentals of Educational Planning 51*. UNESCO, 7 Place de Fontenoy, 75700, Paris France.

Hadjeris, F. (2021). Revisiting sustainable development Goal 4 in the context of COVID-19 Pandemic: A case study of online teaching in Algerian higher education institutions. *Human Behavior and Emerging Technologies, 3*(1), 160-168.

Hakim, C. (2000). *Research design: Successful designs for social and economic research.* London: Routledge.

Hammond, J. (2001). *Scaffolding: Teaching and learning in language and literacy education. Primary English Teaching Association.* Marrickville: New South Wales, 2204, Australia.

Hansen, A. (2012). Trainees and teachers as reflective learners. In A. Hansen (ed.), *Reflective Learning and Teaching in Primary Schools* (pp. 32–48). Thousand Oaks, CA: Sage Publications.

Harputlu, L. & Ceylan, E. (2014). The effects of motivation and metacognitive strategy use on EFL listening proficiency. *Procedia - Social and Behavioral Sciences, 158*, 124-131.

Harris, K. R., & Graham, S. (1999). Programmatic intervention research: Illustrations from the evolution of self-regulated strategy development. *Learning Disabilities Quarterly, 22*, 251–262.

Hattie, J. (2009). *Visible Learning*. London: Routledge.

Haukås, A., Bjørke, C. & Dypedahl, M. (2018). *Metacognition in Language Learning and Teaching*. New York: Routledge.

Hayes, D. (2010). Language learning, teaching and educational reform in rural Thailand: an English teacher's perspective. *Asia Pacific Journal of Education, 30*(3), 305-319.

Hayes, D. (2016). The value of learning English in Thailand and its impact on Thai: perspectives from university students. *Asia Pacific Journal of Education, 36*(1), 73-91.

Hedges, L.V. (2012). Design of Empirical Research. In Arthur, J., Waring, M., Coe, R., & Hedges, L.V. (eds), *Research Methods & Methodologies in Education* (pp.23-30). London: Sage.

Higgins, S. (2013). Self-regulation and learning: evidence from meta-analysis and from classrooms. In British Journal of Educational Psychology Monograph Series II:

Psychological Aspects of Education - Current Trends: No. 10. Self-regulation and dialogue in primary classrooms. Whitebread, D., Mercer, N., Howe, C. & Tolmie, A. Leicester: British Psychological Society. 111-126.

Higgins, S. E. (2014). Critical thinking for a 21st century education: A cyber-tooth curriculum? *Prospects, 44*(4), 559–574.

Higgins, S. E. (2015). A Recent History of Teaching Thinking. In R. Wegerif., L. Li., and J. C. Kaufman, *The Routledge International Handbook of Research on Teaching Thinking* (pp.19-28). Oxon: Routledge.

Higgins, S. E. (2018). *Improving Learning: Meta-analysis of Intervention Research in Education*. Cambridge: Cambridge University Press.

Higgins, S., Hall, E., Baumfield, V. & Moseley, D. (2005). *A meta-analysis of the impact of the implementation of thinking skills approaches on pupils, Project Report.* EPPI-Centre, Social Science Research Unit, Institute of Education, University of London, London.

HM Treasury. (2011). *The Magenta Book: Guidance for Evaluation*. London: HM Treasury.

Huebner, T. (2019). Language Policy and Bilingual Education in Thailand: Reconciling the Past, Anticipating the Future. *LEARN Journal: Language Education and Acquisition Research Network, 12*(1), 19-29.

Hutchison, D. & Styles, B. (2010). *A Guide to Running Randomised Controlled Trials for Educational Researchers*. Slough: NFER.

Karpicke, J. D. (2009). Metacognitive Control and Strategy Selection: Deciding to Practice Retrieval During Learning. *Journal of Experimental Psychology, 138*(4), 469-486.

Kaur, A., Young, D., & Kirkpatrick, R. (2016). English education policy in Thailand: Why the poor results? In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 345-361). Cham, Switzerland: Springer.

Keyuravong, S. (2010). Insights from Thailand. In R. Johnstone (ed.), *Learning through English: Policies, challenges and prospects: Insights from East Asia*. British Council.

Khonamri, F. & Ahmadi, F. (2015). The effect of metacognitive strategy training and raising EFL learners' metacognitive knowledge on listening performance. *Indonesian Journal of Applied Linguistics, 5*(1), 18–27.

Killip, S., Mahfoud, Z., and Pearce, K. (2004). What Is an Intracluster Correlation Coefficient? Crucial Concepts for Primary Care Researchers. *Annals of Family Medicine*, *2*(3), 204–208.

Kirkpatrick, A. & Liddicoat, A. J. (2019). Language education policy in Asia: An overview. In A. Kirkpatrick and A., J., Liddicoat (eds.), the *Routledge international handbook of language education policy in Asia* (pp.3-13). Abingdon, Oxon: Routledge.

Kluwe, R. (1987). Executive decisions and regulation of problem solving behavior. In F. E. Weinert & R. H. Kluwe (eds.), *Metacognition, motivation and understanding* (pp.31-64). L. Erlbaum Associates.

Kobayashi, A. (2018). Investigating the effects of metacognitive instruction in listening for EFL learners. *Journal of Asia TEFL, 15*(2), 310–328.

Kraince, R. G. (2007). Islamic higher education and social cohesion in Indonesia. *Prospects, 37*(3), 345–356.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. University of Southern California.

Kuhn, D. (2000). Metacognitive development. Current Directions in Psychological Science, 9(5), 178–181.

Lan, Y. C., Lo, Y. L., & Hsu, Y.S. (2014). The Effects of Meta-Cognitive Instruction on Students' Reading Comprehension in Computerized Reading Contexts: A Quantitative Meta-Analysis. *Educational Technology & Society, 17* (4), 186–202.

Lantolf, J., Thorne, S. L., & Poehner, M. (2015). Sociocultural Theory and Second Language Development. In B. van Patten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 207-226). New York: Routledge.

Larson-Freeman, D., Driver, P., Gao, X., & Mercer, S. (2021). Learner agency: Maximizing learner potential. Oxford University Press. Retrieved from www.oup.com/elt/expert.

Li, L. (2015). A Confucian perspective on teaching thinking in China. In R. Wegerif, L. Li and J. C. Kaufman (Eds), *The Routledge International Handbook of Research on Teaching Thinking. (pp.45-57).* Oxon: Routledge.

Liow, J. C. (2010). *Islam, Education and Reform in Southern Thailand*. ISEAS Publishing.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: SAGE.

Lipsey, M. W. (1992). Juvenile Delinquency Treatment: A Meta-Analytic Inquiry into the Variability of Effects. In T. D. Cook, H. Cooper, D. S. Cordray, H, Hartmann, L. V. Hedges, R. J. Lights, T. A., Louis, and F. Mosteller (Eds.), *Meta-Analysis for Explanation* (pp. 83-128). New York: Russell Sage Foundation.

Livingston, J. (2003). Metacognition: An Overview. *ERIC Database*. Retrieved from https://files.eric.ed.gov/fulltext/ED474273.pdf

Lui, P. & Li, L. (2015). An overview of metacognitive awareness and L2 reading strategies. In R. Wegerif., L. Li., and J. C. Kaufman, *The Routledge International Handbook of Research on Teaching Thinking* (pp.266-279). Oxon: Routledge

Macaro, E. & Lynn E. (2008). Raising the Achievement of Young- Beginner Readers of French Through Strategy Instruction. *Applied Linguistics, 29* (1), 90–119.

Mackenzie, A. S. (2002, May). EFL curriculum reform in Thailand. In *1st Annual JALT Pan-SIG Conference, Kyoto, Japan*. Available at https://hosted.jalt.org/pansig/2002/ HTML/Mackenzie1.htm

Magno, C. (2010). The role of metacognitive skills in developing critical thinking. *Metacognition and Learning, 5*(2), 137-156.

Manyukhina, Y. & Wyse, D. (2019). Learner agency and the curriculum: a critical realist perspective. *The Curriculum Journal, 30*(3), 223-243, DOI: 10.1080/09585176.2019.1599973

Marin, L. & Halpern, D. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity, 6*(1), 1-13.

Mason, L. H. (2013). Teaching Students Who Struggle With Learning to Think Before, While, and After Reading: Effects of Self-Regulated Strategy Development Instruction, *Reading & Writing Quarterly, 29*:2, 124-144.

Mateos, M., Martín, E., Villalón, R., & Luna, M. (2008). Reading and writing to learn in secondary education: Online processing activity and written products in summarizing and synthesizing tasks. *Reading and Writing, 21*(7), 675-697.

Mayer, R. E. (1996). Learners as information processors: Legacies and limitations of educational psychology's second. *Educational psychologist, 31*(3-4), 151-161.

McCawley P. F. (2001). The logic model for program planning and evaluation. University of Idaho. Retrieved October, 2019 from www.uiweb.uidaho.edu/extension/LogicModel.pdf

McLaughlin, J. A., & Jordan, G. B. (1999). Logic models: a tool for telling your program's performance story. *Evaluation and program planning, 22*(1), 65-72.

McLaughlin, J. A., & Jordan, G. B. (2015). Using Logic Models. In K. E. Newcomer., H. P. Hatry. & J. S. Wholey, (eds.), Handbook of Practical Program Evaluation (4th ed pp.62-87). New Jersey: Jossey-Bass.

Mendolia, S., Paloyo, A. R., & Walker, I. (2018). Heterogeneous effects of high school peers on educational outcomes. *Oxford Economic Papers, 70*(3), 613-634.

Mercer, S. (2011). Understanding learner agency as a complex dynamic system. *System, 39*(4), 427-436.

Mertens, D. M. (2010*). Research and Evaluation in Education and Psychology: Integrating Diversity with Quantitative, Qualitative, and Mixed methods* (3rd ed.). Thousand Oaks: SAGE.

Ministry of Education (MoE). (2008). *The Basic Education Core Curriculum B.E. 2551 (A.D. 2008)*. Bangkok: Office of the Basic Education Commission.

Mitchell, R., Myles, S. & Marsden, E. (2013). *Second Language Learning Theories* (3rd ed). Oxon: Routledge.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology, 63*(8), 1-28.

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 6 (7): e1000097.

Mokhtari, K., & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94* (2), 249-259.

Moore, T. (2011). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research and Development, 30*(3), 261–74.

Moseley, D., Elliott, J., Gregson, M., & Higgins, S. (2005). Thinking skills frameworks for use in education and training. *British Educational Research Journal, 31*(3), 81–101.

Msaddek, M. (2016). The effect of metacognitive strategy instruction on Moroccan EFL learners' strategy use and reading achievement. *Arab World English Journal,* 7(3), 287-301.

Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *Modern Language Journal, 89*(1), 76–91.

National Research Council of Thailand (NRCT). (2015). National Policy and Guidelines for Human Research. Retrieved August 20, 2019, from https://www2.nrct.go.th /Portals/0/data/2559/01/National_Policy_Guidelines_for_Human_Research2015.pdf

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*(2), 102–116.

Nonthaisong, K. (2015). *The impact of the English language policy in the basic education core curriculum BE 2551(AD 2008) on English instruction of Thai secondary EFL teachers: a multiple case study in northeast Thailand*. Doctoral thesis, The University of Alabama, Alabama, USA.

Noom-Ura, S. (2013). English-Teaching Problems in Thailand and Thai Teachers' Professional Development Needs. *English Language Teaching, 6*(11), 139-147.

Norris, J. M., & Ortega, L. (2001). Does Type of Instruction Make a Difference? Substantive Findings from a Meta-analytic Review. *Language learning, 51*(1), 157-213.

Norwawi, N. M., Sempo, M. W., Islam, Y. M., & Ismail, N. (2019). Inquiry learning model inspired by Quranic Hamzah Istifham verses. *Proceedings of the 1st International Conference on Universal Wellbeing* (pp. 67-69.). UTM Kuala Lumpur, Malaysia.

Nunan, D. (2002). Learner strategy training in the classroom: An action research study. In J. C. Richards & W. A. Renandya (eds.), *Methodology in Language Teaching: An Anthology of Current Practice*, (pp.133-144). Cambridge: Cambridge University Press.

O' Connell, M., & Marks, G. N. (2022). Cognitive ability and conscientiousness are more important than SES for educational attainment: An analysis of the UK Millennium Cohort Study. *Personality and Individual Differences, 188*, 111471.

O'Malley, J.M., and Chamot, A.U. (1990). *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press.

O'Malley, J.M., Chamot, A.U., Stewner-Manzanares, G., Kupper, L., & Russo, R. (1985). Learning strategies used by beginning and intermediate ESL students. *Language Learning, 35*(1), 21–46.

OECD. (2005). *The Definition and Selection of Key Competencies: Executive Summary*. Retrieved October 5, 2019, from www.oecd.org/edu/skills-beyond school/ definitionandselectionofcompetenciesdeseco.htm

OECD. (2012). *Equity and quality in education: Supporting disadvantage students and schools*. Paris: OECD Publishing.

OECD. (2019). *Future of education and skills 2030: OECD learning compass 2030*. Retrieved November 10, 2020, from https://www.oecd.org/education/2030-project/teaching-and-learning/learning/learning-compass-2030/OECD_Learning_Compass_2030_concept_note.pdf

Office of the Education Council (OEC) Thailand. (2017). *Education in Thailand.* Bangkok: OEC. http://backoffice.onec.go.th/uploads/Book/1532-file.pdf

Ogle, D. M. (1989). The know, want to know, learn strategy. In K. D. Muth (Ed.), *Children's Comprehension of Text: Research into Practice* (pp.205–223). Newark, DE: International Reading Association.

O'Malley J.M., Chamot, A.U. & Küpper, L. (1989). Listening Comprehension Strategies in Second Language Acquisition. *Applied Linguistics, 10*, 418-37.

Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics, 30*, 85-110.

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. New York: Newbury House.

Oxford, R. (2002). Language learning strategies in a nutshell: Update and ESL suggestions. In J. C. Richards & W. A. Renandya (eds.), *Methodology in Language Teaching: An Anthology of Current Practice*, (pp.124-132). Cambridge: Cambridge University Press.

Oxford, R. (2011). Strategies for learning a second or foreign language. *Language Teaching 44*,(2), 167–180.

Oxford, R. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed). Oxon: Routledge.

*Panahandeh, E. & Asl, S. E. (2014). The effect of planning and monitoring as metacognitive strategies on Iranian EFL learners' argumentative writing accuracy. *Procedia - Social and Behavioral Sciences, 98*, 1409–1416.

Paris, S. G., & Winograd, P. (1990). Promoting metacognition and motivation of exceptional children. *Remedial and special Education, 11*(6), 7-15.

Perkins, D. (1992). *Smart Schools: Better Thinking and Learning for Every Child*. New York: Free Press.

Perkins, D. N., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional

Perry, J., Lundie, D., & Golder, G. (2019). Metacognition in schools: what does the literature suggest about the effectiveness of teaching metacognition in schools?, *Educational Review, 71*(4), 483-500.

Pherali, T. (2021). Social justice, education and peacebuilding: conflict transformation in Southern Thailand. *Compare: A Journal of Comparative and International Education*.

Pintrich, P. R. (2000). The role of Goal Orientation in Self-Regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 451–502). Academic Press.

Pintrich, P. R. (2002). The Role of Metacognitive Knowledge in Learning, Teaching and Assessing. *Theory Into Practice, 41*(4), 219-225.

Pintrich, P. R., & Zusho, A. (2007). Student motivation and self-regulated learning in the college classroom. In R. Perry, & J. Smart (Eds.), *The scholarship of teaching and*

learning in higher education: An evidence-based perspective (pp. 731–813). New York: Springer.

Pintrich, P.R., Smith, D. Garcia, T. & MaKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire* (MSLQ). Ann Arbor, MI: University of Michigan.

Plonsky, L. (2011). The Effectiveness of Second Language Strategy Instruction: A Meta-analysis. *Language learning, 61*(4), 993-1038.

Plonsky, L. (2014). Study quality in quantitative L2 research (1990-2010): A methodological synthesis and call for reform. *The Modern Language Journal, 98*(1), 450–70.

Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E., Farstrup and S. J., Samuels (Eds.), *What research has to say about reading instruction*, (pp. 291-309). Newark, DE: International Reading Association.
view of transfer. *Educational Psychologist, 47*(3), 248-258.

Quigley, A., D. Muijs and E Stringer. (2018). *Metacognition and Self-regulated Learning*. London: Education Endowment Foundation. Available at https://education endowmentfoundation.org.uk/public/files/Publications/ Metacognition /EEF_Metacognition_and_self-regulated_learning.pdf

*Rahimirad, M. & Shams, M. R. (2014). The Effect of Activating Metacognitive Strategies on the Listening Performance and Metacognitive Awareness of EFL Students. *International Journal of Listening, 28*(3), 162-176.

Ramanathan, V. & Kaplan, R. B. (1996). Some problematic "channels" in the teaching of L2 critical thinking in current L1 composition textbooks: Implications for L2 student writers. *Issues in Applied Linguistics,* 7, (2), 225–249.

Raoofi, S., Chan, S. H., Mukundan, J. and Rashid, S. (2013). Metacognition and Second/Foreign Language Learning. *English Language Teaching, 7* (1); 36-49.

*Rasekh, Z. E., & Ranjbary, R. (2003). Metacognitive strategy training for vocabulary learning. *Teaching English as a Second or Foreign Language,* 7(2).

Rattan, A., Good, C., & Dweck, C. (2012). "It's ok — Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology, 48*(3), 731-737.

*Razı, S. & Çubukçu, F. (2014). Metacognition and reading: Investigating intervention and comprehension of EFL freshmen in Turkey. *Procedia - Social and Behavioral Sciences,* 158, 288–295.

Renger, R., & Titcomb, A. (2002). A three-step approach to teaching logic models. *The American Journal of Evaluation*, 23(4), 493-503.

Richards, J. C. (2001). *Curriculum Development in Language Teaching*. Cambridge: Cambridge University Press.

Richards, J., C. & Rodgers, T., S. (2001). *Approaches and Methods in Language Teaching* (2nd ed.). Cambridge: Cambridge University Press.

*Roohani, A. & Asiabani, S. (2015). Effects of self-regulated strategy development on EFL learners' reading comprehension and metacognition. *GEMA Online Journal of Language Studies* 15, no.3: 31–49.

Rosenshine, B. (2012). Principles of Instruction: Research based principles that all teachers should know. *American Educator*, Spring 2012. http://www.aft.org/pdfs/ americaneducator/spring2012/Rosenshine.pdf

Rost, M. (2001). *Teaching and Researching Listening*. NY: Longman.

Rost, M. (2011). *Teaching and Researching Listening* (2nd ed.). Harlow UK: Longman.

Roulston, K. (2014). Analysing Interviews. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis* (pp.297-312). London: Sage.

Rubin, J (1975). What the 'good language learner' can teach us. *TESOL Quarterly, 9*, 41-51.

Saraç, S., & KaraKelle, S. (2012). On-line and Off-line Assessment of Metacognition. *International Electronic Journal of Elementary Education*, 4(2), 301-315.

Schraw, G. (2009). Measuring metacognitive judgments. In Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). *Handbook of metacognition in education*. (pp.415-429) Routledge.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460–475.

Schunk, D. H. (2012). *Learning theories an educational perspective* (6th ed). Pearson.

See, B.H., Gorard, S. & White, P. (2004). Teacher demand: Crisis, what crisis? *Cambridge Journal of Education*, *34* (1), 103-123.

Seiff, A. (2016). We Have No Freedom: Losing hearts and minds in Thailand's deep south. *World Policy Journal, 33*(4), 48-54.

*Seifoori, Z. (2016). Metacognitive awareness and the fluency of task-based oral output across planning conditions: The case of Iranian TEFL students. *Iranian Journal of Language Teaching Research, 4*(1), 11–26.

Shadish, W.R., Cook T. D. & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Shibli, D., & West, R. (2018). Cognitive load theory and its application in the classroom. *Journal of the Chartered College of Teaching*, 2. From https://impact.chartered. college/article/shibli-cognitive-load-theory-classroom/

*Shirvan, M. E., Ghonsooly, B. & Fatemi, A. H. (2016). The effectiveness of strategy-based instruction in teaching English as a second or foreign language: A meta-analysis of experimental studies. *US-China Foreign Language 14*(3), 163-181.

Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. Abingdon: Routledge.

Siddiqui, N. (2017). Socio-economic segregation of disadvantaged children between schools in Pakistan: comparing the state and private sector. *Educational Studies. 43*(4)391-409.

Siddiqui, N., Gorard, S. & See, B. H. (2018). The importance of process evaluation for randomised control trials in education. *Educational Research, 60*(3), 357-370.

Smith, E. (2008). *Using Secondary Data in Educational and Social Research*. New York: McGraw-Hill Education.

Smith, E. (2012). Secondary Data. In J. Arthur, M., Waring, R., Coe, & L.V. Hedges (Eds), *Research Methods & Methodologies in Education* (pp.125-130). London: SAGE.

Sodachan, M., & Chayanuvat, A. (2018). The Effectiveness of Metacognitive, Cognitive and Social Listening Strategies on Private University Freshman Students' Listening Achievement, Self-Awareness and Social Interaction: A Case Study. *Rangsit Journal of Educational Studies, 5*(1), 57-70.

Spada, N. (2019). Classroom Observation Research. In J. W Schwieter & A. Benati (Eds.), *The Cambridge Handbook of Language Learning* (pp.186-207). Cambridge: Cambridge University Press.

Spolsky, B. (2004). *Language Policy*. Cambridge: Cambridge University Press.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

Sternberg, R. (1998). Abilities Are Forms of Developing Expertise. *Educational Researcher, 27*(3), 11.

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges and J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (pp.435-452). New York: Russell Sage Foundation.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction, 4*(4), 295-312.

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*(2) 165–169.

Sukamolson, S. (1998). English Language Education Policy in Thailand. *Asian Englishes, 1*(1), 68-91.

*Tanewong, S. (2019). Metacognitive pedagogical sequence for less-proficient Thai EFL listeners: A comparative investigation. *RELC Journal, 50*(1), 86–103.

Tashakkori A., & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks: Sage.

*Tavakoli, H. & M. Koosha. (2016). The effect of explicit metacognitive strategy instruction on reading comprehension and self-efficacy Beliefs: The case of Iranian university EFL students. *Porta Linguarum* 25: 119–133.

Taylor, A., Stevens, J., & Asher, J. W. (2006). The effects of explicit reading strategy training on L2 reading comprehension. In J. M. Norris and L. Ortega (eds), *Synthesizing research on language learning and teaching* (pp. 213-244). Amsterdam: John Benjamins Publishing.

*Teng, F. (2016). Immediate and delayed effects of embedded metacognitive instruction on Chinese EFL students' English writing and regulation of cognition. *Thinking Skills and Creativity, 22*, 289–302.

Teng, F. (2020). The role of metacognitive knowledge and regulation in mediating university EFL learners' writing performance. *Innovation in Language Learning and Teaching, 14*, (5), 436-450.

*Teng, F. & Reynolds, B. L. (2019). Effects of individual and group metacognitive prompts on EFL reading comprehension and incidental vocabulary learning. *PLoS ONE, 14*(5), 1–24.

Thomas, J., Harden, A. & Newman, M. (2012). Synthesis: Combining results systematically and appropriately. In D., Gough, S., Oliver and J, Thomas (Eds.), *An Introduction to Systematic Reviews* (pp. 179-226). London: SAGE.

Tobias, S. & Everson, H. T. (2000). Assessing metacognitive knowledge monitoring. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 147–222). Lincoln, NE:Buros.

Tobias, S. & Everson, H. T. (2009). The Importance of Knowing What You Know: A Knowledge Monitoring Framework for Studying Metacognition in Education. In D. J. Hacker., J. Dunlosky. and A. C. Graesser, *Handbok of Metacognition in Education* (pp.107-128). Oxon: Routledge.

Torgerson, C. (2003). *Systematic Reviews*. London: Continuum.

Torgerson, C., Brooks, G., Gascoine, L. & S. Higgins, S. (2019). Phonics: Reading Policy and the Evidence of Effectiveness from a Systematic 'Tertiary' Review. *Research Papers in Education* 34, (2), 208–38.

Torgerson, C., Hall, J. & Light, K. (2012). Systematic reviews. In J. Arthur, M. Waring, R. Coe and L.V. Hedges (Eds.), *Research Methods and Methodologies in Education*, (pp. 217-230). London: SAGE.

Torgerson, C.J., Torgerson, D.J., & Taylor, C.A. (2015). Randomized controlled trials. In K. E. Newcomer., H. P. Hatry. & J. S. Wholey, (eds.), *Handbook of Practical Program Evaluation* (4th ed pp.158-176). New Jersey: Jossey-Bass.

Torgerson, D.J. & Torgerson, C.J. (2003). Avoiding bias in randomised controlled trials in educational research. *British Journal of Educational Studies, 51*(1), 36-45.

Torgerson, D.J., & Torgerson, C.J. (2008). *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*. Houndsmills: Palgrave Macmillan.

Trowler, P. (2003). *Education Policy* (2nd ed). London: Routledge.

*Tsai, Y. R. & Talley, P. C. (2014). The effect of a course management system (CMS)-supported strategy instruction on EFL reading comprehension and strategy use. *Computer Assisted Language Learning,* 27(5), 422-438.

Tuntivivat, S. (2016). The inter-relationship between violence and education amidst armed conflict in Southern Thailand. *Journal of aggression, conflict and peace research, 8* (4), 269-278

UNESCO Bangkok. (2017). *2017/2018 Global Education Monitoring Report: Thailand highlights*. Retrieved Aug 5, 2019, from https://bangkok.unesco.org/content/20172018-global-education-monitoring-report-thailand-highlights

United Nations (UN). (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development. New York: United Nations*. Retrieved April 16, 2019, from https://sustainabledevelopment.un.org/post2015/transformingourworld

Unterhalter, E. (2009). What is Equity in Education? Reflections from the Capability Approach. *Studies in Philosophy and Education, 28*(5), 415-424.

VandenBos, G. R. (ed). (2015). APA Dictionary of Psychology (2nd ed.). Washington,DC: American Psychological Association.

Vandergrift, L. (2004). Learning to listen or listening to learn? *Annual Review of Applied Linguistics*, *24*, 3–25.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191-210.

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

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.

Vandergrift, L., Goh, C., Mareschal, C., & Tafaghodatari, M. H. (2006). The Metacognitive Awareness Listening Questionnaire (MALQ): Development and validation. *Language Learning 56*, 431–62.

Veenman, M. V. J., Van Hout-Wolters, B. H. A. M. & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1*, 3-14.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research, 84*(1), 30-43.

Wang, W. (2016). Learning to listen: the impact of a metacognitive approach to listening instruction. *Asia-Pacific Edu Res, 25*(1),79–88

*Wang, Z. (2017). Keeping journals as a means of metacognitive strategies training in learner autonomy. In L. Sun, L. Hale, Q. Fan, and J. Zhang (Eds.), *Proceedings of the Sixth Northeast Asia International Symposium on Language, Literature and Translation* (pp.742-747). Georgia: The American Scholars Press.

Wegerif, R., Li, L. & Kaufman, J. C. (2015). *Introduction.* In R. Wegerif., L. Li., and J. C. Kaufman (eds.), *The Routledge International Handbook of Research on Teaching Thinking* (pp.1-7). Oxon: Routledge.

Wenden, A. (1987). Metacognition: An Expanded View on the Cognitive Abilities of L2 Learners. *Language Learning, 37*(4), 573-597.

Wenden, A. (1998). Metacognitive Knowledge and Language Learning. *Applied Linguistics 19*, 515-37.

White, G. (2008). Listening and good language learners. In C. Griffiths (ed*), Lesson from Good Language Learners* (pp.208-217). Cambridge: Cambridge University Press.

White, P. (2009). *Developing research questions*. London: Palgrave.

Whittaker, C. (2013). Introduction. In B., Tomlinson and C., Whittaker (Eds.), *Blended Learning in English Language Teaching: Course Design and Implementation* (pp.9-23). London: British Council.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist 54*(8), 594–604.

Wiliam, D. (2017). I've come to the conclusion Sweller's Cognitive Load Theory is the single most important thing for teachers to know <http://bit.ly/2kouLOq>, tweet, viewed 27 November 2021, <https://twitter.com/dylanwiliam/status/824682504602943489>.

Williams, E. & Williams, A. (2007). *ESOL and EFL: An Unhelpful Distinction?* Berkshire: CfBT Education Trust.

Wilson, N. S., & Bai, H. (2010). The Relationships and Impact of Teachers' Metacognitive Knowledge and Pedagogical Understandings of Metacognition. *Metacognition and Learning 5* (3), 269–88.

Wongdaeng, M. & Hajihama, S. (2018). Perceptions of project-based learning on promoting 21st century skills and learning motivation in a Thai EFL setting. *Journal of Studies in the English Language 13*,(2), 158–90.

Wongdaeng, M. (2020). Task-based Language Teaching in Thai Context: A Call for Robust Evidence, in A. Harriet & R. Sharon (eds), *Imagining Better Education 2019* (pp.101-120). Durham: Durham University, School of Education.

Wongdaeng, M. (2011). Improving learning through a metacognitive intervention for tertiary Thai EFL learners: Evidence from a cluster randomised controlled trial. In *Conference Proceedings of the 14th International Conference Innovation in Language Learning* (pp.93-98). https://www.filodiritto.com/proceedings

Wongdaeng, M. (2022). The role of metacognition in the learning of English as a foreign language. In N., Siddiqui and S. Gorard (eds), *Making your doctoral research project ambitious: Developing Large-Scale Studies with Real-World Impact.* Abingdon: Routledge.

Wudthayagorn, J. (2019, March 9-12). English exit examinations for Thai university students: A closer look at the policy implementation [Paper presentation]. *Conference of the American Association for Applied Linguistics* (AAAL), Atlanta, GA.

Yabukoshi, T. (2021). Self-regulation and self-efficacy for the improvement of listening proficiency outside the classroom. *Language Learning Journal, 49*(1), 27–40.

*Yang, C. (2013). The effects of metacognitive strategy training on the graduates listening. *International Academic Workshop on Social Science*, 918–923.

Yokakul, K. (2017, April 8). Inequality in Thai education still a major handicap. *The Bangkok Post*. https://www.bangkokpost.com/thailand/general /1229176/urban-rural-gap-a-big-issue-in-education

*Younsi, A. (2017). *Investigating the Role of Explicit Metacognitive Strategy Instruction (CALLA) in Enhancing Vocabulary Acquisition Among EFL Learners: The Case of First-Year Students of English at Larbi Ben M'Hidi University, Oum El Bouaghi*. Master's thesis., Larbi Ben M'hidi University.

Zeng, Y., & Goh, C. (2018). A self-regulated learning approach to extensive listening and its impact on listening achievement and metacognitive awareness. *Studies in Second Language Learning and Teaching, 8*(2), 193-218.

*Zenotz, V. (2012). Awareness development for online reading. *Language Awareness, 21*(1–2), 85–100.

Zimmerman, B. J. (1989). Models of self-regulated learning and academic achievement. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 1-25). New York: Springer.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation (pp. 13–39). Academic Press.

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice, 41*(2), 64-70.

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American educational research journal, 31*(4), 845-862.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). *Handbook of metacognition in education* (pp.299-316). Routledge.

APPENDICES

Appendix 1: Metacognitive awareness questionnaire

This questionnaire is part of a study implementing a pedagogical approach to improve
English listening and metacognition. All the information that you provide
will be strictly confidential. This questionnaire will not be used to evaluate any students and
will not affect the course grade.
แบบสอบถามนี้เป็นส่วนหนึ่งของงานวิจัยเพื่อทดลองโมเดลการสอนเพื่อพัฒนาทักษะการฟังภาษาอังกฤษ

และการความคุมกระบวนการคิด (metacognition) ข้อมูลทั้งหมดที่ผู้ตอบแบบสอบถามให้ไว้ จะถูกรักษา

เป็นความลับ แบบสอบถามนี้จะไม่ถูกใช้เพื่อประเมินนักศึกษา และจะไม่ส่งผลใดๆ ต่อผลการเรียนทั้งสิ้น

Part 1: General Information (ตอนที่ 1 ข้อมูลทั่วไป)
Instructions: Please make a tick in front of the information which applies to you or supply
information where appropriate.
(**คำชี้แจง** โปรดทำเครื่องหมายถูกหน้าข้อความที่สอดคล้องกับข้อมูลของท่าน และเติมข้อมูล
ที่เกี่ยวข้อง)

Student ID (รหัส นศ.): _____

Gender (เพศ): _____Male   _____Female

Subject Major (วิชาเอก): _____

Hometown (ภูมิลำเนา): _____Pattani   ___Yala   _____Narathiwat
                     _____Songkhla   ___ Satun   _____ Phatthalung
                     _____Others: Please specify (อื่นๆ ระบุ)_____

Your High School (โรงเรียนมัธยมปลาย):
        _____ A government's public school in your hometown.
        (โรงเรียนของรัฐในภูมิลำเนาของตนเอง)

        _____ A private religious school (ex. Buddhist, Christian, Islamic) in your hometown.
         (โรงเรียนสอนศาสนา (เช่น โรงเรียนพุธ โรงเรียนคริสต์ โรงเรียนเอกชนสอนศาสนาอิสลาม) ในภูมิลำเนา

         ของตน)

        _____ A government's public school in another province. Please specify _____
        (โรงเรียนของรัฐในจังหวัดอื่นซึ่งไม่ใช่ภูมิลำเนา โปรดระบุจังหวัด)

        _____ A private religious school (ex. Buddhist, Christian, Islamic) in another
         province. Please specify_____.
        (โรงเรียนสอนศาสนา (เช่น โรงเรียนพุธ โรงเรียนคริสต์ โรงเรียนเอกชนสอนศาสนาอิสลาม) ในจังหวัดอื่นซึ่ง

        ไม่ใช่ภูมิลำเนา โปรดระบุจังหวัด)

_____ Other types of school: Please specify in Thai language _____.

(โรงเรียนประเภทอื่น โปรดระบุ เป็นภาษาไทย)

What is the score of your most recent standardised English test? Please answer all that apply.

(คะแนนการทดสอบภาษาอังกฤษล่าสุด กรุณาระบุทุกการทดสอบที่เคยสอบ)

___ TellMeMore Placement: score…….   ___PSU GET: score ……..

___TellMeMore Progress: score ……..

___Other tests (ex. TOEIC, TOEFL), please specify_____: score ……..

___Never take tests (ไม่เคยสอบวัดระดับภาษาอังกฤษ)

Are you the first generation in your family entering university (i.e. parents, older generations, and older brother or sisters never attended university)?

(ท่านเป็นคนรุ่นแรกในครอบครัวที่ได้รับการศึกษาระดับมหาวิทยาลัยหรือไม่ กล่าวคือ ผู้ปกครองได้รับการศึกษาสูงสุดต่ำ

กว่าระดับปริญญาใช่หรือไม่)

_____ Yes   ____No   _____ Don't know.

What language do you use as your mother tongue? (ท่านใช้ภาษาใดเป็นภาษาแม่)

_____ A southern Thai dialect (ภาษาไทยพื้นถิ่นภาคใต้)

_____ A Malay dialect of Pattani, Yala and Narathiwat (ภาษามลายูพื้นถิ่นปัตตานี ยะลา

นราธิวาส)

_____ A standard Thai dialect (ภาษาไทยมาตรฐาน)

_____ A Chinese dialect (ภาษาจีน สำเนียงใดสำเนียงหนึ่ง)

_____ Any other language or dialect. Please specify _____. (ภาษาอื่นๆ

โปรดระบุ)

What is the status of English in your language learning? (สถานะของภาษาอังกฤษในการเรียนรู้ทาง

ภาษาของคุณเป็นเช่นใด)

_____ First Language/Mother Tongue (ภาษาที่หนึ่ง/ภาษาแม่)

_____ Second Language (ภาษาที่ที่สอง)

_____ Third Language (ภาษาที่ที่สาม)

Part 2: Metacognitive Awareness Listening Questionnaire (MALQ)

(**ตอนที่ 2** แบบสอบถามด้านการฟัง)

Instructions: Read each statement and circle the number of your choice to indicate how much you agree with each one. (**คำชี้แจง** อ่านข้อความแต่ละข้อความและวงกลมล้อมรอบตัวเลขที่บ่ง บอกว่าท่านเห็นด้วยกับข้อความดังกล่าวมากน้อยเพียงใด)

1 = strongly disagree (ไม่เห็นด้วยอย่างยิ่ง)　2 = somewhat disagree (ค่อนข้างไม่เห็นด้วย)

3 = not sure (ไม่มั่นใจ)　　　　　4 = somewhat agree (ค่อนข้างเห็นด้วย)

5 = strongly agree (เห็นด้วยอย่างยิ่ง)

How much do you agree with the followings? ท่านเห็นด้วยกับข้อความต่อไปนี้ มากเพียงใด	strongly disagree (1)	somewhat disagree (2)	not sure (3)	somewhat agree (4)	strongly agree (5)
1. Before I started to listen, I had a plan in my head for how I was going to listen. ก่อนที่จะเริ่มฟัง ฉันมีการวางแผนไว้ในหัวว่าจะฟังอย่างไร	1	2	3	4	5
2. Before listening, I thought of similar texts that I might have listened to or have known. ก่อนฟัง ฉันนึกถึงบทความที่มีความคล้ายคลึงกันที่ฉันอาจเคยฟัง มาหรือเคยรู้มาก่อน	1	2	3	4	5
3. I had a goal in mind as I listened. ฉันมีเป้าหมายในขณะที่ฉันฟัง	1	2	3	4	5
4. As I listened, I periodically asked myself if I was satisfied with my level of comprehension. ในขณะฟัง ฉันถามตัวเองเป็นระยะว่าฉันพอใจกับระดับความ เข้าใจของตนหรือไม่	1	2	3	4	5
5. After listening, I thought about how I listened, and thought what I might do differently if I was to listen again. หลังฟัง ฉันคิดถึงกระบวนการฟังของฉัน และคิดถึงสิ่งที่ฉันอาจทำ ต่างไปจากเดิม ถ้าฉันได้ฟังอีกครั้ง	1	2	3	4	5
6. I focused harder on the listening text when I had trouble understanding. ฉันเพ่งความสนใจกับบทความที่ฟังมากขึ้น เมื่อฉันพบอุปสรรคใน การทำความเข้าใจ	1	2	3	4	5
7. When my mind wandered, I recovered my concentration right away. เมื่อสมาธิฉันหลุด ฉันจะดึงสมาธิกลับมาได้ทันที	1	2	3	4	5
8. When I had difficulty understanding what I heard, I gave up and stopped listening attentively. เมื่อฉันพบความยากในการเข้าใจสิ่งที่ฟัง ฉันถอดใจและหยุดฟัง อย่างตั้งใจ	1	2	3	4	5
9. I tried to get back on track when I lost concentration. ฉันพยายามดึงสมาธิกลับมา เมื่อฉันเสียสมาธิ	1	2	3	4	5
10. I felt that listening in English is difficult. ฉันรู้สึกว่าการฟังภาษาอังกฤษนั้นยาก	1	2	3	4	5
11. I found that listening in English was more difficult than reading, speaking, or writing in English.	1	2	3	4	5

ฉันพบว่าการฟังภาษาอังกฤษ ยากกว่าการอ่าน พูด หรือ เขียน ภาษาอังกฤษ					
12. I didn't feel nervous when I listened to English. ฉันไม่รู้สึกประหม่า เมื่อฉันฟังภาษาอังกฤษ	1	2	3	4	5
13. I translated in my head as I listened. ฉันแปลอยู่ในหัว ในขณะฟัง	1	2	3	4	5
14. I mainly translated key words as I listened. ฉันแปลคำสำคัญเป็นส่วนใหญ่ ในขณะฟัง	1	2	3	4	5
15. I translated word by word, as I listened. ฉันแปลแบบคำต่อคำ ในขณะฟัง	1	2	3	4	5
16. I used the words I understood to guess the meaning of the words I didn't understand. ฉันใช้คำที่ฉันเข้าใจ มาเดาความหมายของคำที่ฉันไม่เข้าใจ	1	2	3	4	5
17. As I listened, I used my prior knowledge of the topic for my understanding. ในขณะฟัง ฉันใช้ความรู้ที่มีอยู่เดิมของฉันที่เกี่ยวกับหัวข้อนั้น เพื่อให้เกิดความเข้าใจ	1	2	3	4	5
18. I realize that my relevant experience about the topic can help me understand. ฉันตระหนักว่าประสบการณ์ของฉันที่เกี่ยวข้องกับหัวข้อนั้น ช่วย ให้ฉันเข้าใจได้	1	2	3	4	5
19. As I listened, I quickly adjusted my interpretation if I realized that it was not correct. ในขณะฟัง ฉันจะปรับการตีความความหมายได้อย่างรวดเร็ว เมื่อ ฉันรู้สึกว่าการตีความนั้นไม่ถูกต้อง	1	2	3	4	5
20. I used the general idea of the listening text to help me guess the meaning of the words that I didn't understand. ฉันใช้ความคิดรวบยอดของบทความที่ฟัง เพื่อช่วยให้ฉันเดา ความหมายของคำที่ฉันไม่เข้าใจ	1	2	3	4	5
21. When I guessed the meaning of a word, I thought back to everything else that I had heard to see if my guess made sense. เมื่อฉันเดาความหมายของศัพท์คำหนึ่ง ฉันคิดย้อนไปถึงสิ่งต่างๆ ที่ฉันได้ยิน เพื่อดูว่าการเดาของฉันสมเหตุสมผลหรือไม่	1	2	3	4	5

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

Thank you very much for your cooperation.

Appendix 2: Data extraction and quality rating of the included studies for the review

Notes: IG = Intervention Group, CG = Control Group, MI = Metacognitive Intervention/Instruction, BAU = Business as Usual, ES = Effect Size

CI= Confidence Interval, underg =undergraduates, sem= semester, uni = university, prep = preparation, prof = proficiency

Authors, year, country	Design	Participant	Intervention, length	Compared condition	IG	CG	Smallest Cell	Outcomes/ Measures	Validity assurance	Results as reported by authors	ES (g)	CI (lower, upper)	Rating
Abdelhafez, 2006, Egypt	Pre/post Control trial with random allocation	80 underg	MI for listening, 12 weeks	BAU	40	40	40	Listening & reading / Researcher adapted test	Average	Effective for both skills	1.62	1.12, 2.13	★★
Al-Ghazo, 2016, Jordan	Pre/post Control trial with sampling	60 underg	MI using CALLA model by Chamot and Rubin (1994), 1sem	BAU	30	30	30	Reading / Researcher-made reading test	Average, Risk of researcher effect	Effective	3.77	2.93, 4.61	★
Altay, 2017, Turkey	Pre/post Control trial with baseline observation	51 B1 Turkish speaking underg	Online MI, 6 weeks	BAU	25	26	25	Reading/ Adapted Uni Test	Average	no effect	0.22	-0.33, 0.77	★↗
Bozorgain, 2018, Iran	Pre/post Control trial	180 advance underg	MI for listening (MPS) / MI +dialogic instruction, 12 sessions	BAU	60	60	60	Listening / Listening Test adapted from Summit TV 2 Activity Worksheets	Average	Effective	1.17	0.78, 1.56	★↗

Study	Design	Sample	Intervention	Control	N	N	N	Outcome / test	Quality	Effectiveness	ES	CI	Rating
Cabrera-Solano, 2019, Ecuador	Pre/post Control trial	42 A1 underg	MI via vocab app with didactic technique, 5 months	BAU	20	22	20	Vocabulary /researcher vocab test	Average	Effective	0.79	0.16, 1.41	★
Chang, 2010, Taiwan	Pre/post Control trial	90 underg	MI with focus on self-monitoring, 1 sem	BAU	45	45	45	English & reading / English Prof test	Average	Effective for English, no effect for reading	0.28	-0.14, 0.69	★ ⌐
Chou, 2017, Taiwan	Pre/post Control trial	88 B1 underg	Task-based MI for listening, 18 weeks	BAU	44	44	44	Listening / IELTS Listening	Average	Effective	0.51	0.09, 0.94	★ ⌐
Farahain, 2018 Iran	Pre/post Control trial with sampling	69 pre intermediate underg	Portfolio as MI tool, 12 weeks	BAU	31	38	31	Writing/ Writing test	Average, Risk of researcher effect	Effective	2.71	2.06, 3.37	★
Habibian, 2015, Malaysia	Pre/post Control trial with criterion sampling	48 underg	MI for reading, 12 weeks	BAU	24	24	24	Reading / Reading test from TOEIC prep	Average	Effective	1.32	0.70, 1.95	★
Khonamri, 2015, Iran	Pre/post Control trial with baseline observation	40 intermediate underg	MI based on CALLA model, 7 sessions	BAU	20	20	20	Listening / Listening part from TOEFL	Average, Risk of researcher effect	Effective	2.28	1.49, 3.08	★

282

Kobayashi, 2018, Japan	Pre/post Control trial	48 underg	MI for listening (MPS), 1 sem	BAU	26	22	22	Listening / Test taken from TOEIC prep	Average, Risk of researcher effect	Effective	0.62	0.04, 1.20	★
Msaddek, 2016, Morocco	Pre/post Control trial	113 underg	MI for reading based on RSI model, 12 sessions	BAU	63	50	50	Reading/ reading test	Average	Effective	1.77	1.34, 2.21	★⌐
Nakatani, 2005, Japan	Pre/post Control trial with baseline observation	62 underg	MI with focus on awareness-raising, 12 weeks	BAU	28	34	28	Speaking / oral test with rubric for assessment	Average	Effective	0.64	0.13, 1.15	★⌐
Panahandeh, 2014, Iran	Pre/post Control trial with baseline observation	60 underg	MI based on CALLA model with focus on planning & monitoring, 8 weeks	BAU	30	30	30	Writing / Writing test	Average	Effective	0.66	0.14, 1.18	★⌐
Rahimirad, 2014, Iran	Pre/post Control trial with random allocation & baseline observation	50 upper intermediate underg	MI for listening (MPS), 8 sessions	BAU	25	25	25	Listening/ IELTS Listening	Average	Effective	1.16	0.56, 1.76	★⌐
Rasekh, 2003, Iran	Pre/post Control trial	53 intermediate underg	MI for vocabulary, 10 weeks	BAU	27	26	26	Vocabulary /Researcher -made vocab test	Average	Effective	0.96	0.39, 1.53	★

Razi, 2014, Turkey	Pre/post Control trial	93 underg	MI for reading, 6 weeks	BAU	46	47	46	Reading / Researcher-made reading test	Average	Effective	0.46	0.04, 0.87	★⌐
Roohani, 2015, Iran	Pre/post Control trial	70 underg	MI based on SRSD model, 6 weeks	BAU	35	35	35	Reading / Reading summary marked by 2 markers using rubric from Hoyt 2010	Average	Effective	1.42	0.90, 1.95	★⌐
Seifoori, 2016, Iran	Pre/post Control trial with random allocation	114 underg	MI with focus on planning strategies, unclear	BAU	58	56	56	Speaking / Preliminary English Test/ Picture Description Test	Average	Effective	0.47	0.11, 0.86	★★
Shirvan, 2016	Review method with non-robust criteria	meta-analysis of 26 studies	SBI with metacognitive strategies as one main variable, N/A	Mixture			N/A	N/A	Average, lacking quality appraisal	large effect size	0.82	0.54, 1.09	★⌐
Tanewong, 2019, Thailand	Pre/post Control trial	64 low prof underg	MI for listening (MPS), 20 sess, 10 wks	BAU	31	33	31	Listening / List of IELTS prepare	Average	no effect at achievement test	0.52	0.02, 1.02	★⌐
Tavakoli & Koosha, 2016, Iran	Pre/post Control trial with random allocation	100 undeg	MI for reading, 12 weeks	BAU	50	50	50	Reading/ Researcher made reading test	Average	Effective	1.50	1.05, 1.94	★★

284

Teng, 2016, China	Pre/post Control trial	120 chinsese underg	Cooperative MI, 36 sessions	BAU, cooperative learning	40	40	40	Writing / Writing test with rubric marked by outsider	Fairly good with independent markers	Effective	2.32	1.75, 2.88	★ⲻ
Teng, F., 2019, China	Pre/post Control trial with criterion sampling	171 underg	Collaborative MI, 120-minute session but unclear how many session	BAU, Collaborative	41	44	41	Reading / Researcher reading test	Average	Effective	6.15	5.13, 7.16	★ⲻ
			MI, 120-minute session but unclear how many session	BAU, Collaborative	42	44	42	Reading / as above	Average	Effective	1.35	0.88, 1.82	
Tsai, 2014, Taiwan	Pre/post Control trial	114 Chinese speaking underg	Tech-supported MI for reading, unclear	BAU	56	58	56	Reading / author adapted test	Average	Effective	2.14	1.68, 2.60	★ⲻ
Wang, 2017, China	Pre/post Control trial	66 underg	MI using journal as a mean, 3 sem	BAU	33	33	33	English / CET 4	Average	no effect	0.19	-0.30, 0.67	★ⲻ
Yang, 2013, China	Pre/post Control trial	150 mix-prof graduates	MI in high prof learners, 40 sessions	BAU high	15	12	12	Listening / Test	Average	not specified	0.36	-0.41, 1.12	★
			MI in mid prof learners, 40 sessions	BAU mid	48	34	37	Listening / Test		effective	0.85	0.40, 1.30	
			MI in low prof learners, 40 sessions	BAU low	21	17	17	Listening / Test		Effective	0.85	0.19, 1.52	

| Younsi, 2017, Algeria | Pre/post Control trial with criterion sampling | 44 underg | MI based on CALLA model, 6 sessions | BAU | 21 | 23 | 21 | Vocabulary / Test (unclear) | Average | Effective | 2.61 | 1.81, 3.42 | ★ |
| Zenots, 2012, Spain | Pre/post Control trial with baseline observation | 143 underg | MI for reading, 5 weeks | BAU | 95 | 48 | 48 | Reading / researcher made reading test | Average | Effective | 1.17 | 0.80, 1.54 | ★↰ |

Appendix 3: Samples of materials in students' practice book
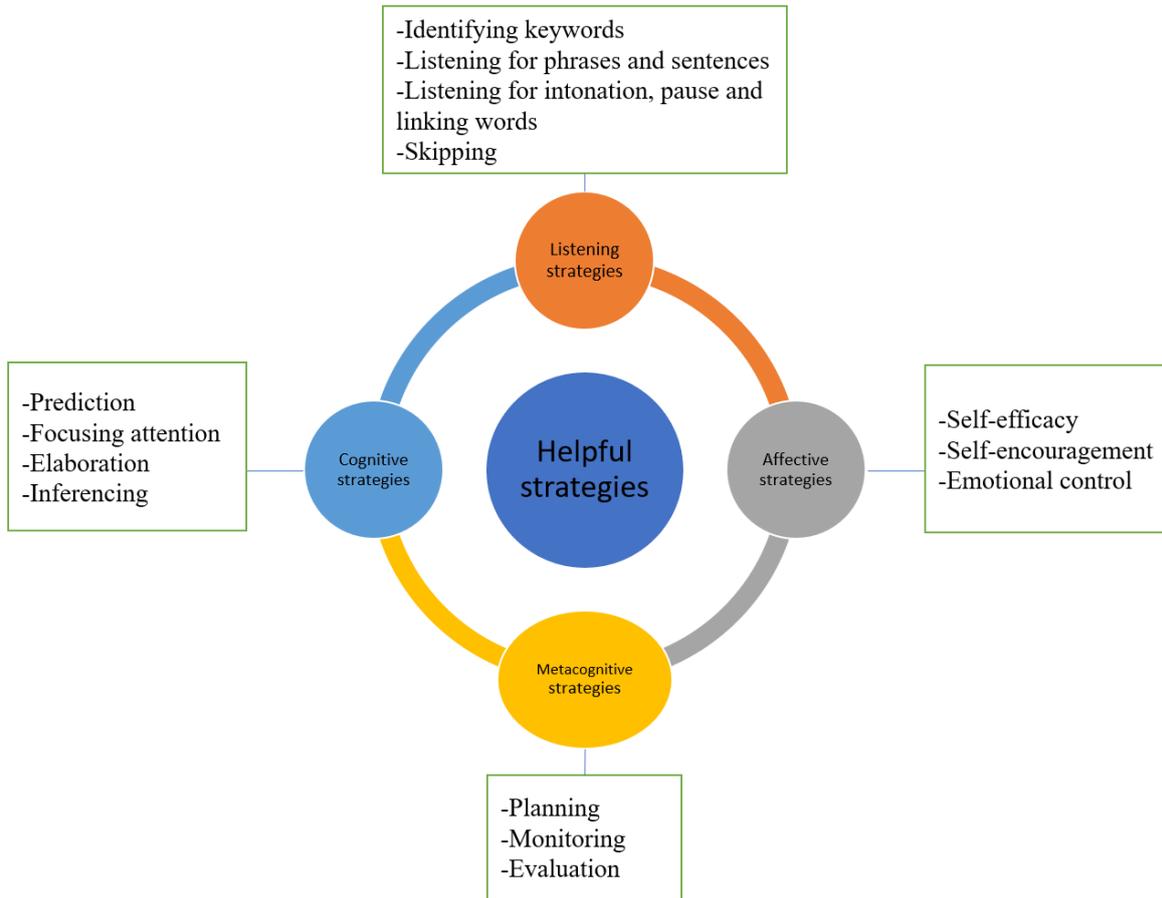
Figure A 3.1: Helpful learning strategies

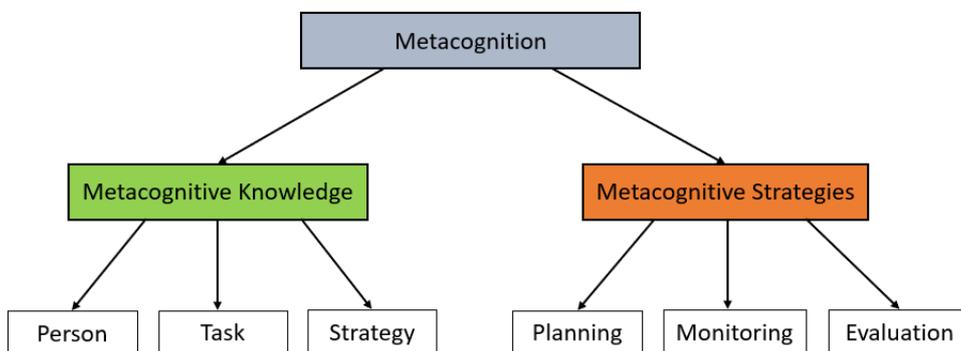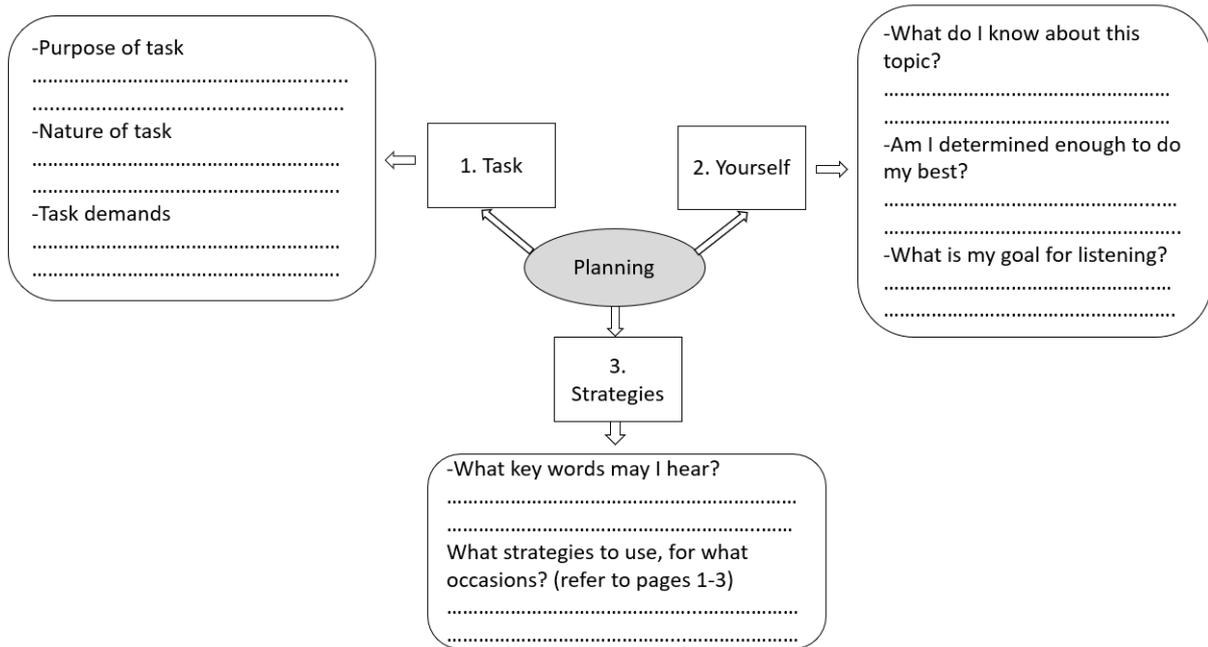**Figure A 3.2:** Components of metacognition

Figure A 3.3: Samples of a metacognitive task for planning and a worked example

-Purpose of task
....
....
-Nature of task
....
....
-Task demands
....
....

⟸ **1. Task**

2. Yourself ⟹

-What do I know about this topic?
....
....
-Am I determined enough to do my best?
....
....
-What is my goal for listening?
....
....

Planning

3. Strategies

-What key words may I hear?
....
....
What strategies to use, for what occasions? (refer to pages 1-3)
....
....

Example

-Purpose of task
…understanding people talking about global issues……………
-Nature of task
…circle answers of 6 m/c questions……………………………
-Task demands
…not quite familiar, so need to be highly focused and use a lot of inferences ………………..

⟸ **1. Task**

2. Yourself ⟹

-What do I know about this topic?
…global warming because people cut trees and use too much energy ………………
-Am I determined enough to do my best?
…I need to start with the belief that I can do it………………….
-What is my goal for listening?
…control myself well and get at least 4 correct ……………

Planning

3. Strategies

-What key words may I hear?
.. global warming, air pollution (from photo), climate change…………………
-What strategies to use, for what occasions? (refer to pages 1-3)
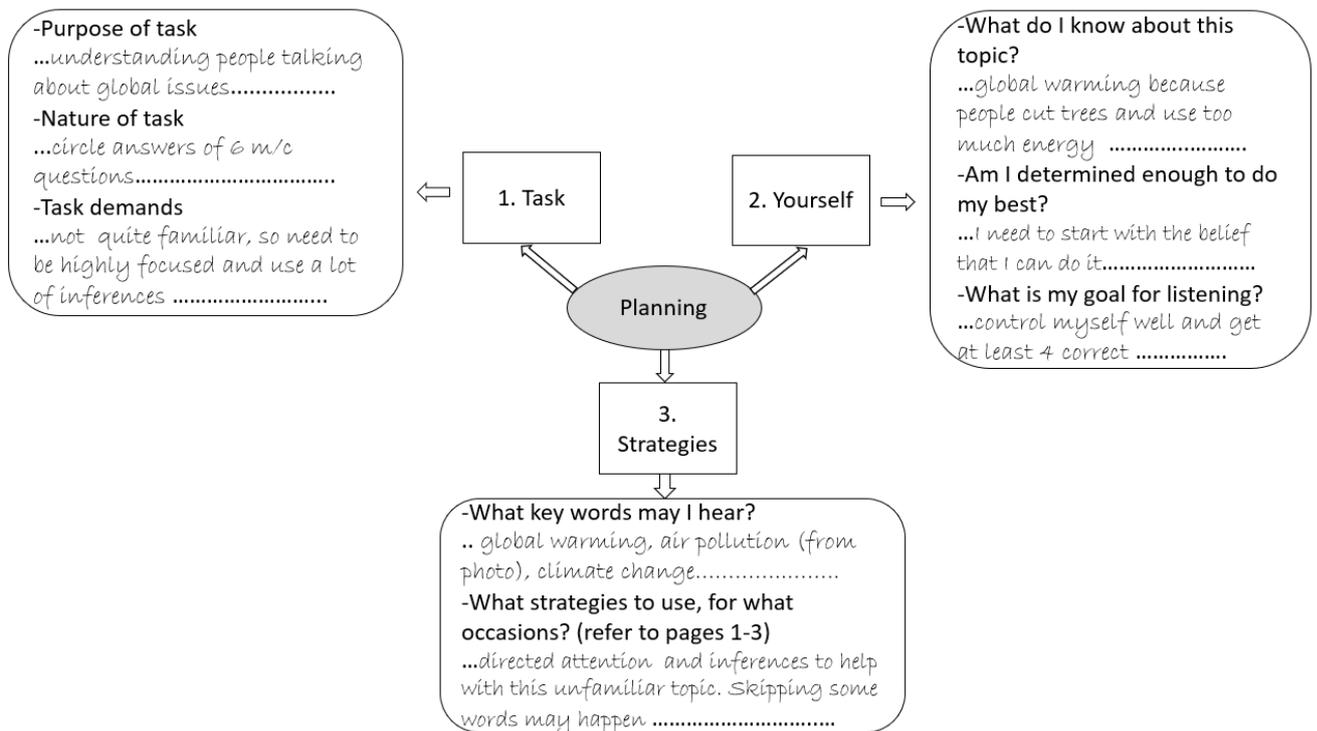…directed attention and inferences to help with this unfamiliar topic. Skipping some words may happen ……………………

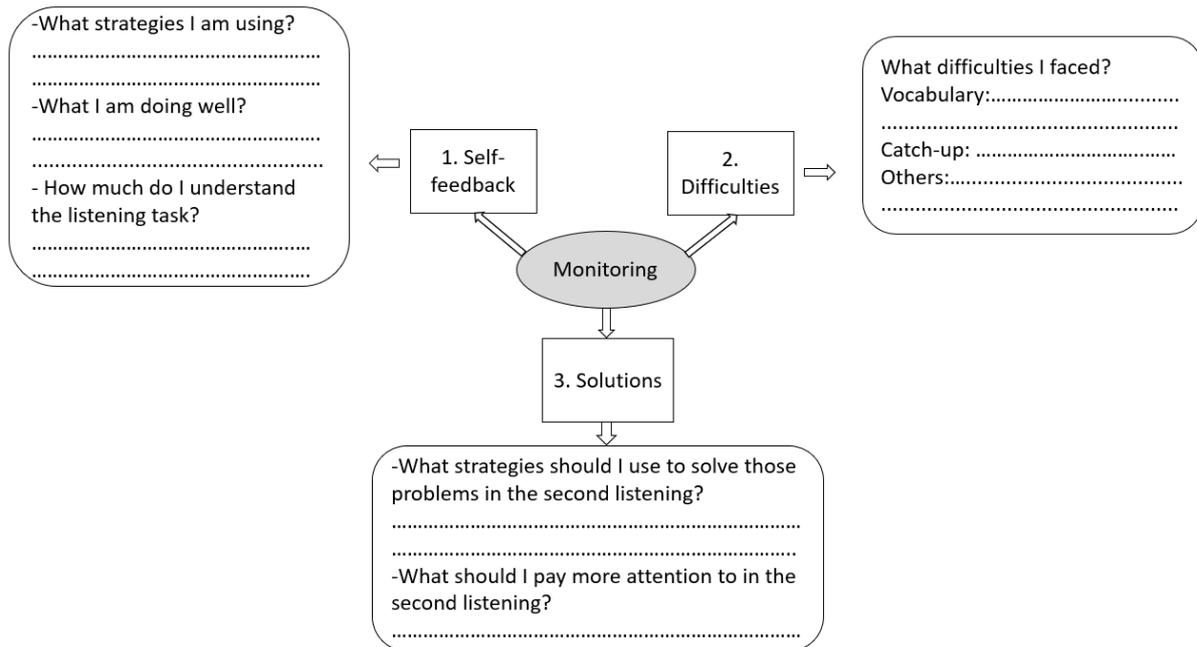**Figure A 3.4:** Samples of a metacognitive task for monitoring and a worked example

-What strategies I am using?
………………………………………………
………………………………………………
-What I am doing well?
………………………………………………
………………………………………………
- How much do I understand the listening task?
………………………………………………
………………………………………………

1. Self-feedback

2. Difficulties

What difficulties I faced?
Vocabulary:………………….............
……………………………………………………
Catch-up: …………………………………
Others:…………………………………………
…………………………………………………

Monitoring

3. Solutions

-What strategies should I use to solve those problems in the second listening?
…………………………………………………………………………
…………………………………………………………………………..
-What should I pay more attention to in the second listening?
…………………………………………………………………………

Example

-What strategies I am using?
.. *inference, directed attention, prediction, identify keyword…*
-What I am doing well?
.. *right prediction from photo, maintaining attention well…*
-How much do I understand the listening?
….*more than half now………….*

1. Self-feedback

2. Difficulties

What difficulties I faced?
Vocabulary:…*crime*…………………..
……………………………………………………
Catch-up: …*the first and third items*……………………………………
Others:…………………………………………
…………………………………………………

Monitoring

3. Solutions

-What strategies should I use to solve those problems in the second listening?
…*selective attention, listening for phrases, listening for intonation* ……………………………
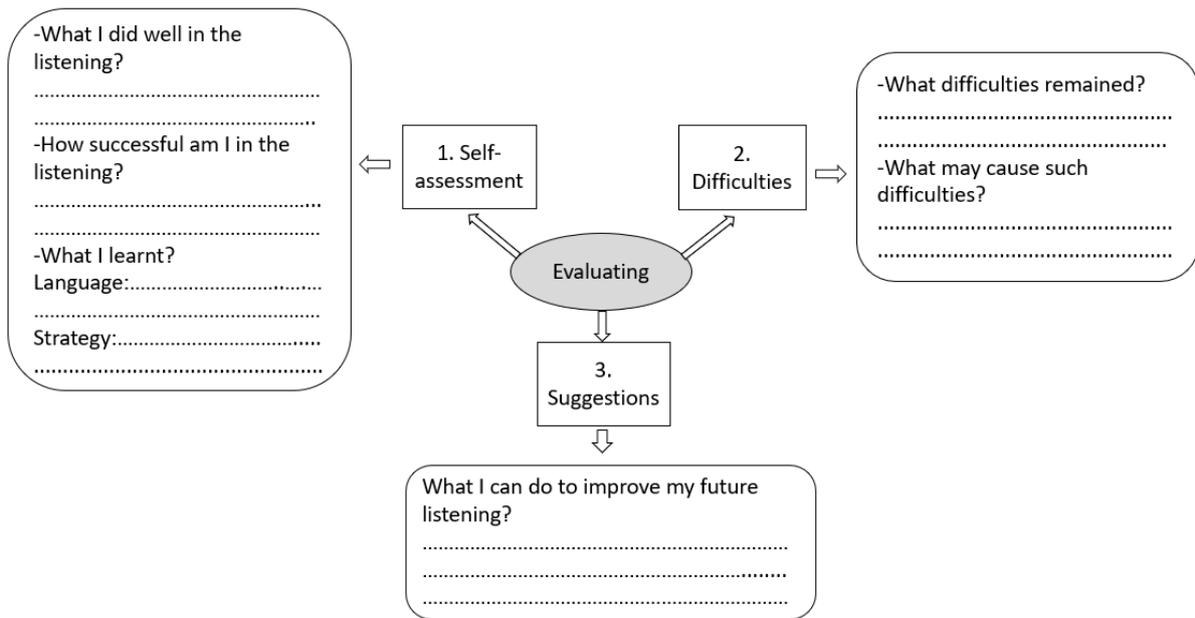-What should I pay more attention to in the second listening?
…*the first and the third* …………………….

Figure A 3.5: Samples of a metacognitive task for evaluation and a worked example

-What I did well in the listening?
…………………………………………
………………………………………….
-How successful am I in the listening?
…………………………………………..
…………………………………………
-What I learnt?
Language:……………………………
……………………………………………
Strategy:…………………………….
…………………………………………

1. Self-assessment

2. Difficulties

-What difficulties remained?
……………………………………………
……………………………………………
-What may cause such difficulties?
……………………………………………
……………………………………………

Evaluating

3. Suggestions

What I can do to improve my future listening?
……………………………………………………
…………………………………………………….
……………………………………………………

Example

-What I did well in the listening?
…Being focused and selecting strategies aft starting listening
-How successful am I in the listening?
…I am quite happy as I think I can reach my goal. I also did better in the second listening…
-What I learnt?
Language:…housing, transportation …………
Strategy:…listening for phrases, list for intonation .

1. Self-assessment

2. Difficulties

-What difficulties remained?
…catching up with some words due to fast speaking ………….
-What may cause such difficulties?
…maybe the speed and accent but some words can be skipped

Evaluating

3. Suggestions

What I can do to improve my future listening?
…monitoring is very important. I need to learn speed and nature of the audio and select strategies ex listen for intonation to help …………………………

Photo A 3.1: Online independent practices for students

Listening Strategies (EDU) G.8
EDU G.8

สตรีม งานของชั้นเรียน ผู้คน คะแนน

Session 3 : Monitoring your listening ⋮

📋	Session 3: First Listening	โพสต์เมื่อ 18 ส.ค. 2020	⋮
📋	Session 3: Monitoring Listening and Second ...	โพสต์เมื่อ 18 ส.ค. 2020	⋮
📋	Session 3: Practice Test	โพสต์เมื่อ 18 ส.ค. 2020	⋮
📋	Independent Practice 5 : Monitoring Listeni...	โพสต์เมื่อ 18 ส.ค. 2020	⋮
📋	Independent Practice 6: Monitoring Listening	โพสต์เมื่อ 18 ส.ค. 2020	⋮
📋	ไสล์ดประกอบการสอน Session 3: Monitoring Li...	โพสต์เมื่อ 21 ส.ค. 2020	⋮

Listening Strategies (EDU) G.8
EDU G.8

สตรีม งานของชั้นเรียน ผู้คน คะแนน

Session 4: Evaluating your listening ⋮

📋	Session 4: Listening Task	โพสต์เมื่อ 25 ส.ค. 2020	⋮
📋	Session 4: Evaluating Listening	โพสต์เมื่อ 25 ส.ค. 2020	⋮
📋	Session 4: Practice Test	โพสต์เมื่อ 25 ส.ค. 2020	⋮
📋	Independent Practice 7 : Evaluating Listening	โพสต์เมื่อ 25 ส.ค. 2020	⋮
📋	Independent Practice 8: Evaluating Listening	โพสต์เมื่อ 25 ส.ค. 2020	⋮
📖	ไสล์ดประกอบการสอน session 4: Evaluating yo...	โพสต์เมื่อ 25 ส.ค. 2020	⋮

Session 5: Mixed skills ⋮

📋	Session 5: Plan, Listen, Monitor	โพสต์เมื่อ 22 ก.ย. 2020	⋮
📋	Session 5: Evaluating Listening	โพสต์เมื่อ 22 ก.ย. 2020	⋮
📋	Session 5: Practice Test	โพสต์เมื่อ 22 ก.ย. 2020	⋮
📋	Independent Practice 9: Mixed Skills	โพสต์เมื่อ 22 ก.ย. 2020	⋮
📋	Independent Practice 10: Mixed Skills	โพสต์เมื่อ 22 ก.ย. 2020	⋮

Appendix 4: Samples of materials in the teachers' kit

Table A 4.1: Introduction to helpful learning strategies: Listening strategies

Strategy	Definition / Examples	Benefits
-Identifying keywords	-Finding keywords in the task's instruction and questions or in the listening text to direct your listening	-Pinpointing targets to listen for
-Listening for phrases and sentences	- Listening in larger chunks such as phrases or sentences, not word for word	-Helping to get the gist of the listening text and to cope with fast speaking tasks
-Listening for intonation, pause and linking words	- Using tone of voice, pause and linking words to guess the meaning of unknown words.	-Getting hints about the meaning of words
- Skipping	- Skipping some parts that are unimportant, difficult or you can't catch up.	-Skipping some unimportant or difficult parts can give time to better focus on the remaining parts

Table A 4.2: Introduction to helpful learning strategies: Cognitive strategies

Strategy	Definition	Benefits
Prediction: Anticipating keywords and message which may appear in the text.		
-Task-informed prediction	-Using information given in the task such as instruction, questions and photos to predict the contents	-Getting the brain activated to think about the task
-Schema-informed prediction	-Using background knowledge related to the topic to make prediction	
Focusing attention: Paying careful attention and avoid distractions.		
-Directed attention	-Being focused and maintaining attention while listening.	-Being focused helps to better take in information.
-Selective attention	-Attending to specific contents or aspects of language to aid understanding	-Fulfilling the understanding gaps especially in second listening or when listening with targets
Elaboration: Relating prior knowledge to the knowledge gained from the text in order to draw understanding of the text.		
- Personal elaboration		

	- Referring to prior personal experience	-Relevant experience or previous learning can help to draw meaning of unknown words and deal with the task more fluently
- Academic elaboration	- Using knowledge gained from academic situations such as previous lessons.	
Inferencing: Using information within the text to guess the meanings of unfamiliar words or to fill in missing information.		
- Linguistic inferencing	- Using known words to guess the meaning of unknown words.	
- Extralinguistic inferencing	- Using other information such as relationships between speakers and conversational situations to guess the meaning	- Contexts or surrounding words and sentences helps in inferring meaning of new words
- Between-parts inferencing	- Using information from different parts of the text to guess at meaning.	

Table A 4.3: Introduction to helpful learning strategies: Affective strategies

Strategy	Definition and Examples	Benefits
Self-efficacy	- Belief in oneself that they can make use of various strategies to deal with the task. Ex. "The task is challenging but I will do my best with the strategies I know".	- Awareness of strategies and self-efficacy can make one feel more competent to perform a listening task.
Self-encouragement	- Providing personal motivation through positive self-talk or satisfaction with what one has performed well. Ex. "My prediction was right".	-Being positive and feeling satisfied with own performance can keep you motivated.
Emotional control	- Being aware of one's emotions while listening and trying to avoid negative feelings and anxiety by taking a deep breath and keeping focused, etc. Ex. "I can skip this and try to focus on what I can."	-Reducing negative emotion and anxiety helps you better focus on the task.

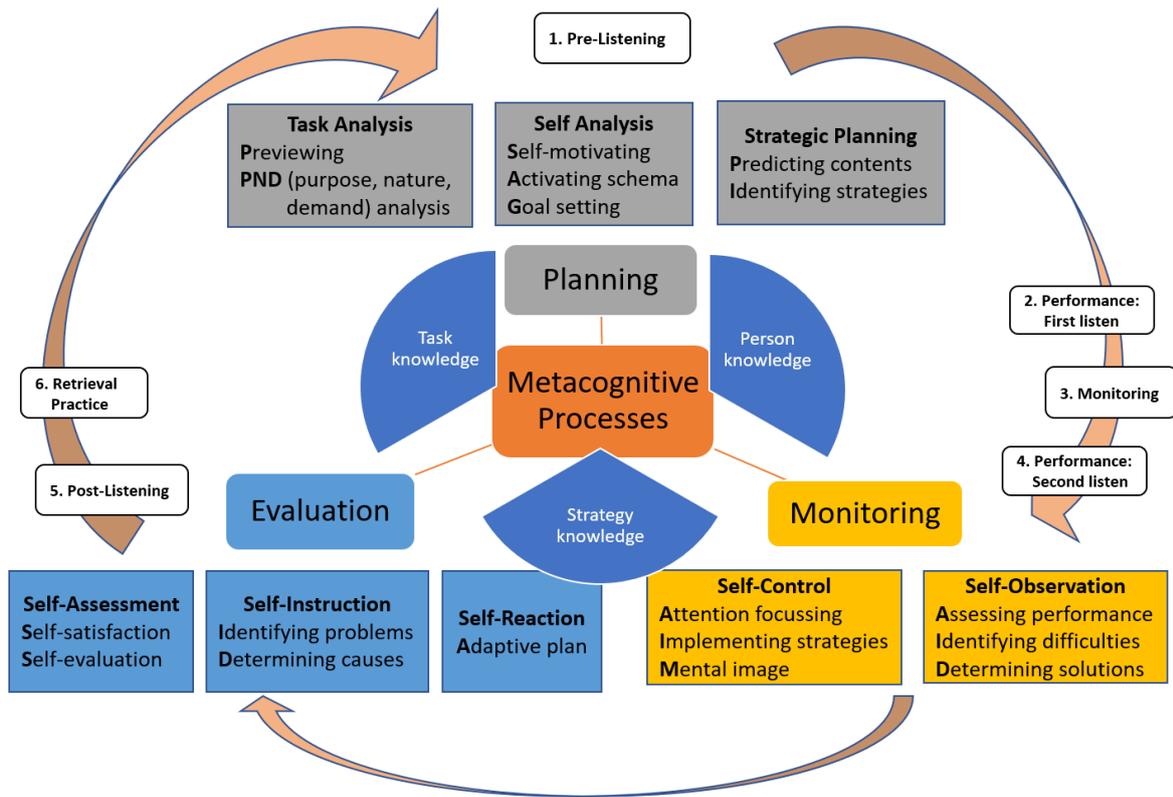**Figure A 4.1:** Overall cycle of metacognitive instruction in listening

Figure A 4.2: Sample instructional materials: Planning

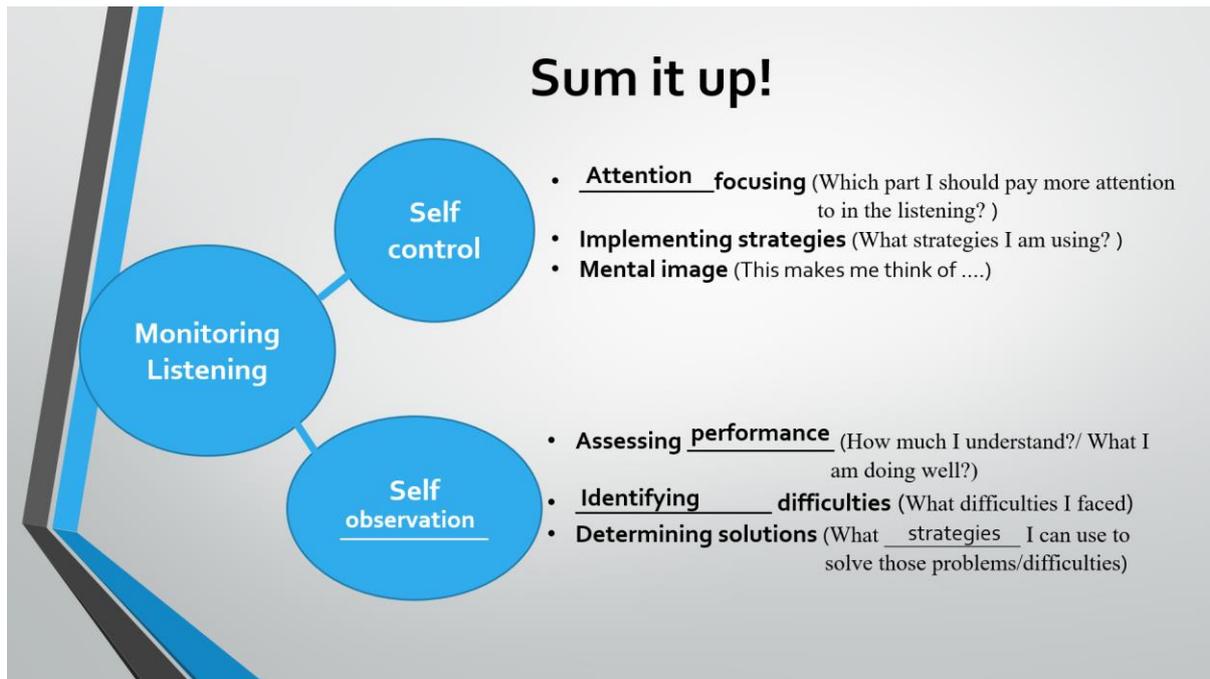**Figure A 4.3:** Sample instructional materials: Monitoring

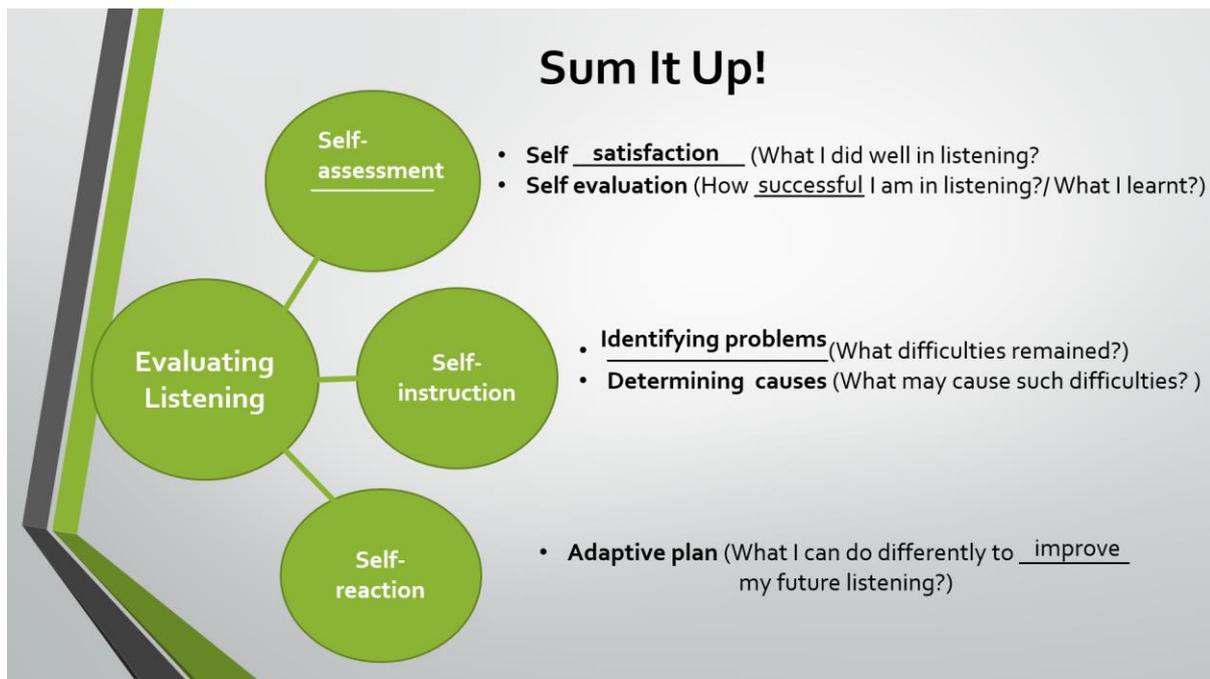**Figure A 4.3:** Sample instructional materials: Monitoring

Photo A 4.1: Sample retrieval practice task

Revision Quiz

Questions on content

1. When can you use your mobile phone?

    a. after the plane has landed

    b. after the announcement

    c. after the take-off

2. You can put your carry-on luggage _____?

    a. in the overhead space   b. beside your seat   c. under your own seat

Questions on vocabulary

1. You can keep your bag in the _____ above.

    a. apartment   **b. compartment**   c. agreement

2. The word _____ means 'to buy' something.

Answer: _____

Questions on strategies

1. In task analysis, you should think about the purpose, the nature and the _____ of tasks.

Answer: _____

2. Write a question that guides you to think about background knowledge?

Answer: _____

Appendix 5: Samples of lesson plans

Sample Lesson Plan 1: Planning

Plane Travelling

1. **Warm Up** (5 minutes)

-Introduce the topic for listening to the students.

-Tell the students that good planning before listening can be helpful for them to approach a listening task.

-Get students in pairs and discuss what they will think about to get themselves more ready to start listening to the topic.

2. **Strategy Instruction: Planning** (12 minutes)

-Explain the *planning strategy* which is shown on *page 5* of their practice book. The video link provided may also be used to introduce the strategy.

3. **Strategy Practice: Planning listening** (7 minutes)

-Inform students that they will be listening to A woman is making a pre-departure announcement on an airplane. They will be asked to listen and make appropriate responses to the 8 prompts about the announcement.

-Before listening, let students discuss *in pairs* how they are going to plan their listening for the topic based on the planning strategy presented. Ask students to write their ideas on the planning chart in their practice book (page 6).

-Encourage students to analyse the task, think about their relevant background knowledge and plan strategies which will be useful.

4. **First listening** (5 minutes)

A woman is making a pre-departure announcement on an airplane. Listen and make

appropriate responses to the 8 prompts about the announcement.

1. flight number	☐ 767 ☐ 676
2. service	☐ from London ☐ to London
3. use of laptops	☐ allowed ☐ not allowed
4. carry-on luggage	☐ allowed ☐ not allowed
5. smoking	☐ allowed ☐ not allowed
6. lunch	☐ provided ☐ need to buy
7. headphones	☐ provided ☐ need to buy
8. safety demonstration	☐ available ☐ not available

-After listening, spare a few minutes for students to think in pairs about how they worked with the first listening task and get ready for the second listening.

5. Second listening (5 minutes)

-Let students listen to the task for a second time.

6. After listening (8 minutes)

-Let the *whole class* discuss their answers and how they got them.

-Tell the correct answers and the clues for such answers.

-Have a post-listening activity as the teacher may find useful ex. looking at the transcript. (optional)

7. Retrieval Practice (12 minutes)

7. 1 Practice Test (6 minutes)

-Use a test to let students recall what they have learned in terms of vocabulary, key message from the task and strategies used during the task.

Questions on content

1. When can you use your mobile phone?

    a. after the plane has landed

    b. after the announcement

    c. after the take-off

2. You can put your carry-on luggage _____?

    a. in the overhead space   b. beside your seat   c. under your own seat

Questions on vocabulary

1. You can keep your bag in the _____ above.

    a. apartment   **b. compartment**   c. agreement

2. The word _____ means 'to buy' something.

Answer: _____

Questions on strategies

1. In task analysis, you should think about the purpose, the nature and the _____ of tasks.

Answer: _____

2. Write a question that guides you to think about background knowledge?

Answer: _____

7.2 Feedback (6 minutes)

-After the students finish the test, discuss answers to provide feedback to students.

Reference

TOEIC Test Prep. (2010). Learning Express. New York: Learning Express, LLC.

Appendix

Hello, passengers, and welcome to Cloud 9 Airlines, flight 767, with nonstop service to London. As we prepare for our flight, we ask that you avoid using your cell phones or your laptops, as these devices prevent airplane systems from working correctly. Please wait until we have landed in London to use your mobile devices. Please store your carry-on luggage in the overhead compartments or under the seat in front of you. If you need assistance, our flight attendants would be happy to help. This is a non-smoking flight and we will be offering a film as well as lunch. If you like, you can purchase headphones from one of the flight attendants to view the movie. The flight safety demonstration will begin shortly.

Appendix 5

<div align="center">

Sample Lesson Plan 2: Monitoring

Road Accident

</div>

1. **Warm Up** (5 minutes)

-Introduce the topic. Inform students that they will be listening to a woman talking about a road accident. Students will choose the best answer to complete 6 sentences about the story.

-Before listening, let students plan their listening by filling in a planning chart the appendix.

-Encourage students to analyse the task, think about their relevant background knowledge and plan strategies to use.

2. **Strategy Instruction: Monitoring** (10 minutes)

-Tell the students that while listening, effective listeners will interact with the text, monitor their own performance and make appropriate reaction to deal with the task.

-Explain the ***monitoring strategy*** which is shown on ***page 8*** of their practice book. The strategy will guide them what they may ask themselves to monitor their performance while listening and after the first listening. The video link provided may also be used to introduce the strategy.

3. **First listening** (5 minutes)

A woman is talking about a road accident. Listen to the story and choose the best answers to fill in each blank.

<div align="center">

Road Accident

</div>

1. This story is a/an _____.

      a. news   b. advertisement   c. history

2. _____ vehicles were in this accident.

      a. Two   b. Three   c. Four

3. The accident happened _____.

      a. in the morning   b. in the afternoon   c. at night

4. The word 'car collision' could simply mean _____.

      a. car park   b. car wash   c. car crash

5. The drivers were taken to the _____.

      a. police station   b. hospital   c. main street

6. There were no major _____ to traffic, and streets were reopened by ten o'clock.

      a. decisions   b. corruptions   c. disruptions

---

4. Strategy Practice: Monitoring progress (7 minutes)

-Let students to work *in pairs* to think about how they worked with the listening task and how well they understood it. Ask th10m to complete the ***monitoring chart*** on ***page 9*** of their practice book.

-Teacher encourages students to think about what strategies they find useful, what difficulties remain and what to do to solve such difficulties.

5. Second listening (5 minutes)

-Students listen again with directed and selective attention on the parts that remain unclear.

6. After listening (8 minutes)

-Let the whole class discuss their answers, how they got them and the teacher.

-Tell the correct answers and what could a clue for such answers.

- Have a post-listening activity as the teacher may find useful ex. looking at the transcript (optional)

7. Retrieval Practice (12 minutes)

7.1 Practice Test (6 minutes)

-Use a test to let students recall what they have learned in terms of vocabulary, key message from the task and strategies used during the task.

Questions on content

1. In this three-vehicle accident, there were two cars and a _____ involved.

    a. motorcycle   b. pickup truck   c. van

2. The cause of the accident is _____.

    a. heavy rain   b. heavy snow   c. not known

Questions on vocabulary

1. Which of the following could be the best meaning of the word 'injury'?

    a. harm to people   b. accident   c. death of people

2. The car accident resulted in the _____ of a few roads.

    a. collision   b. closure   c. crash

Questions on strategies

1. Apart from assessing progress, name one more monitoring strategy.

Answer: _____

2. If you have difficulties with vocabulary, what strategy can be used to help with listening?

Answer: _____

-After the students finish the test, discuss answers to provide feedback to students.

Reference

Collins (2012). *Skills for the TOEIC Test: Listening and Reading*. London: Collins
   Publishers.

Appendix

Listening Script: Road Accident

Good evening, and welcome to the six o'clock news. A three-vehicle collision during this morning's rush hour resulted in the closure of several downtown streets. A car traveling east on Main Street at approximately eight a.m. collided with a car traveling south on Oak Street. A pickup truck was also involved. The exact cause of the crash is still unknown. Firefighters rushed to the scene, and the drivers were taken to Memorial Hospital. An emergency room doctor who was interviewed by our reporter said that there were no serious injuries and all three drivers have been released from the hospital. Despite the street closure, there were no major disruptions to traffic, and streets were reopened by ten o'clock.

Appendix 5

Sample Lesson Plan 3: Evaluating

Character Trait

1. **Warm Up (5 minutes)**

-Introduce the topic. Inform students that they will be listening to three people talking about a person' character in now and in the past. Students will make appropriate response to 8 prompts.

-Before listening, let students plan their listening by filling in a planning chart the appendix.

-Encourage students to analyse the task, think about their relevant background knowledge and plan strategies to use.

2. **First listening (5 minutes)**

Listen to people talking about their friend's character now and in the past. Make a tick to identify if a character was in the past or is now.

	In the past	now
1. -Dress smart wearing a suit and a tie. -Wearing casual cloths like jeans.	☐ ☐	☐ ☐
2. - chubby and out of shape -look terrific and in good shape	☐ ☐	☐ ☐
3. - depressed and worried all the time - enthusiastic	☐ ☐	☐ ☐
4. - athletic with big muscles -overweight	☐ ☐	☐ ☐

3. **Checking progress (5 minutes)**

-Ask students to think about how they worked with the listening task and how well they understood it. Let them write their ideas on the monitoring chart in the appendix.

-The teacher encourages students to think about what strategies they find useful, what difficulties remain and what to do to solve such difficulties.

4. **Second listening (5 minutes)**

Students listen again with directed and selective attention on the parts that remain unclear.

5. Strategy Instruction: Evaluating (10 minutes)

-Tell the students that after listening, effective listeners will reflect on and evaluate their performance.

-Explain the *evaluation strategy* which is shown on *page 10* of their practice book. The video link provided may also be used to introduce the strategy.

6. Strategy Practice: Evaluating listening (8 minutes)

-Let students work *in pairs* to complete the *evaluation chart* on *page 11* of their practice book to assess how satisfied they are with the overall understanding and performance.

- The following questions can be used to elicit discussion.

    - What they did well in the listening

    -What they learnt

    - What difficulties remained and what the causes can be

    -What they can do to improve future listening

- Discuss the correct answers of the listening task with the *whole class* and what the clues for such answers.

7. Retrieval Practice (12 minutes)

7.1 Practice Test (6 minutes)

-Use a test to let students recall what they have learned in terms of vocabulary and key message from the task and strategies used during the task.

Questions on content

1. Rosie is depressed now. Maybe she has a problem with _____.

    a. her brother   b. her boyfriend   c. her work

2. Ted is over 100 kg now after he got _____.

    a. rich   b. sick   c. married

Questions on vocabulary

1. Which word can mean 'sad and worried'?

    a. enthusiastic   b. careful   c. depressed

2. After work, he doesn't want to go anywhere because he feels so _____.

    a. chubby   b. exhausted   c. terrific

Questions on strategies

1. What are the two strategies to do for self-assessment?

Answer: _____

2. After you identify difficulties, what next should you ask yourself?

Answer: _____

7.2 Feedback (6 minutes)

-After the students finish the test, discuss answers to provide feedback to students.

Reference

Richards, J. C. (2011). *Expanding Tactic for Listening: Unit 5 Character Trait*. Oxford: Oxford University Press.

Appendix

Listening Script

1.
Ever since John got that new job, you wouldn't recognize him. I guess you have to be more conservative when you have a job like that. He works really long hours now and he wears a suit and a tie. And when he comes home from work he never wants to go out. All he does is sit in his chair and watch TV. He looks totally exhausted. That's not how he was in college. Back then, he used to go around in old T-shirts and jeans. And he had that crazy green hair!

2.
I ran into Donna Simpson the other day. I could hardly recognize her. Do you remember her from high school? She was a little chubby and out of shape back then. I don't think she ever exercised or played any sports. Well, she really looks different now. She's lost a lot of weight. In fact, she looks terrific. She told me that she decided to get in better shape after she got married last year. Now she goes to the gym three times a week.

3.
Have you had a chance to talk to Rosie lately? She's gotten so depressed-you know, sad and worried all the time. It's a real change. She used to be so enthusiastic in high school. She was a cheerleader, and she was always laughing and making jokes. Now she just sits in coffee shops all by herself and just looks really sad. You can tell that something's really bothering her. It must be her boyfriend. I heard they recently broke up.

4.
Do you remember what Ted Rodgers used to look like? He used to be so athletic, with really big muscles. He went to the gym all the time and was always careful about what he ate and drank. Well, you wouldn't believe how much he's changed. I'm sure he weighs over 100 kilos now. It doesn't look like he does any exercise at all, and he eats just about anything. Someone told me he changed after he got really rich from the stock market. They said that after he got rich, he decided just to enjoy himself.

Appendix 6: Class observation protocol

Instructor: _____ Course Title/Code _____

Students' grade/Year _____ Students' major _____

Date: _____ Time _____

Topic: _____ Target strategy: _____

Number of students: _____ Number of boys: _____ Number of girls:_____

Organization	None (1)	Some (2)	Strong (3)
1. Presented appropriate introduction to/ overview of the topic/session.			
2. Presented the contents in a logical sequence			
3. Manage time appropriately.			
4. Summarized main points of the lesson			
5. Used retrieval practice to consolidate understanding			
Comments:			

Presentation and Interaction	None (1)	Some (2)	Strong (3)
1. Explained major/minor points clearly			
2. Used good examples or instructional media to clarify key points, key terms			
3. Provided effective activities and materials to facilitate learning			
4. Used pair and group discussions to promote active and collaborative learning			
5. Ask questions to encourage students thinking before, during and after a task.			
Comments:			

Target Strategy: Planning (This applies only in the planning session and mixed skills)	None (1)	Some (2)	Strong (3)
1. Explained the planning strategy clearly.			
2. Established clear learning goals (understanding, strategy).			
3. Support students (ex. asking questions, giving examples) to analyse tasks			
4. Promoted students' activation of background knowledge			
5. Used prompts (ex. questions, examples) to guide students' strategic planning			
Comments:			

Target Strategy: Monitoring (This applies only in the monitoring session and mixed skills)	None (1)	Some (2)	Strong (3)
1. Explained the monitoring strategy clearly.			
2. Support students (ex. asking questions, giving examples) to provide feedback to themselves			
3. Promoted students to identify difficulties			
4. Used prompts (ex. questions, examples) to guide students about the possible solutions			
Comments:			

Target Strategy: Evaluating (This applies only in the Evaluating session and mixed skills)	None (1)	Some (2)	Strong (3)
1. Explained the evaluating strategy clearly.			
2. Support students (ex. asking questions, giving examples) to assess themselves.			
3. Promoted students to identify the remaining difficulties.			
4. Used prompts (ex. questions, examples) to guide students' adaptive plan for further development			
Comments:			

Student Reaction	No (1)	Not sure (2)	Yes (3)
1. Most students participate in the session activities appropriately.			
2. Most students seem to respond positively to the metacognitive strategy instruction.			
3. Most students are able to accomplish the metacognitive tasks.			
Comments:			

Overall Comments (major strengths and suggestions for development):

Appendix 7: Participant information sheet

Durham
University
School of Education

Shaped by the past, creating the future

Participant Information Sheet

You are invited to take part in a research study on thinking instruction for improving English learning outcomes. Please read this form carefully and ask any questions you may have before agreeing to be in the study. The study is conducted by Mr. Meechai Wongdaeng as part of his postgraduate study at Durham University. This research project is supervised by Dr. Nadia Siddiqui and Professor Steve Higgins from the School of Education at Durham University.

The purpose of this study is to examine the effectiveness of a model of metacognitive instruction on the learning outcomes of English learners.

If you agree to be in this study, you will be exposed to a thinking-based learning environment in seven sessions of your course and will be asked to respond to a questionnaire about your learning at the beginning and at the end of the semester. You will also be asked to do a pre-test, a post-test and a delayed post-test. The test results will only be used for the purpose of the research and **_WILL NOT_** affect your grade of the course.

You are free to decide whether or not to participate. If you decide to participate, you are free to withdraw at any time without any negative consequences for you.

All responses you give or other data collected will be kept confidential. The records of this study will be electronically kept secure and private. All files containing any information you give are password protected. In any research report that may be published, no information will be included that will make it possible to identify you individually. There will be no way to connect your name to your responses at any time during or after the study.

If you have any questions, requests or concerns regarding this research, please contact me via email at meechai.wongdaeng@durham.ac.uk or by telephone at XXX.

This study has been reviewed and approved by the School of Education Ethics Sub-Committee at Durham University (date of approval: 18 December 2019)

Mr. Meechai Wongdaeng
Doctor of Education Candidate
School of Education
Durham University

Appendix 8: Consent form

Durham University
School of Education

Shaped by the past, creating the future

Declaration of Informed Consent

- I agree to participate in this study, the purpose of which is to examine the effectiveness of a model of metacognitive instruction on the English learning outcomes in Thailand.
- I have read the participant information sheet and understand the information provided.
- I have been informed that I may decline to participate and answer any questions or withdraw from the study without penalty of any kind.
- I have been informed that all of my responses will be kept confidential and secure, and that I will not be identified in any report or other publication resulting from this research.
- I have been informed that the investigator will answer any questions regarding the study and its procedures. Mr. Meechai Wongdaeng, School of Education, Durham University can be contacted via email: meechai.wongdaeng@durham.ac.uk or telephone: XXX.
- I will be provided with a copy of this form for my own records.

Any concerns about this study should be addressed to the School of Education Ethics Sub-Committee, Durham University via email to ed.ethics@durham.ac.uk.

_____

Date   Participant Name (please print)   Participant Signature

I certify that I have presented the above information to the participant and secured his or her consent.

_____

Date   Signature of Investigator

Leazes Road
Durham City, DH1 1TA

Telephone +44 (0)191 334 2000 Fax +44 (0)191 334 8311
www.durham.ac.uk
Durham University is the trading name of the University of Durham

Appendix 9: Random allocation protocol for independent allocator

Durham University
School of Education

Randomisation Protocol

This random allocation is for a research project to evaluate the effectiveness of a pedagogical intervention on the learning outcomes of a group of English learners. This requires a random allocation of the participants into either an intervention or a control group.

Group 1

The participants are from eight clusters.

-Three clusters will be randomly assigned to an intervention group (IG).

-Five clusters will be randomly assigned to a control group (CG).

Among the three intervention clusters:

-Two clusters will be randomly assigned to teacher MM.

-One cluster will be randomly assigned to teacher MW.

Group 2

There are 2 participant teachers.

-One teacher will be randomly assigned to an intervention group (IG).

-One teacher will be randomly assigned to a control group (CG).

Thank you very much for your kind cooperation.

Meechai Wongdaeng

Ed.D. Researcher, Durham University

Appendix 10: Sample of online independent practice

4. The weather today be _____. *
- ( ) a. sunny all day
- ( ) b. snow all day
- ( ) c. rainy all day

5. The low pressure should move out of the area by _____. *
- ( ) a. Monday
- ( ) b. Friday
- ( ) C. Sunday

6. The weekend is expected to be _____. *
- ( ) a. sunny
- ( ) b. rainy
- ( ) c. snowy

7. Write a word from the listening task that you don't know the meaning. *

ข้อความคำตอบสั้นๆ

8. Is the word from number 7 important for completing the task? *
- ( ) a. Yes, it is an important word I need to know to answer a question. So, I will have to guess using the infere...
- ( ) b. Not really because the questions do not ask about it. I may skip it then.

9. If you want to guess the meaning of the word in number 7, what will you do to help you make a * good guess?

ข้อความคำตอบสั้นๆ

10. When you don't give up and keep trying, what will be the results of your learning? *
- ( ) a. you can't improve if you are not good at English.
- ( ) b. if you keep practicing and try different strategies, your English skills can be improved.

Appendix 11: Coefficients of multiple regression models in section 7.2 of Chapter Seven

Table A 11.1: Coefficients for models predicting overall gain scores: intervention and control groups (for Table 7.12)

	Unstandardised B Coefficients	Standardised beta Coefficients
(Constant)	18.84	-
Admission Score	2.27	0.12
Pretest	-0.74	-0.52
Group	-3.64	-0.19

Table A 11.2: Coefficients for models predicting listening gain scores: intervention and control groups (for Table 7.14)

	Unstandardised B Coefficients	Standardised beta Coefficients
(Constant)	9.66	-
Admission Score	1.89	0.15
Pre-Listening	-0.77	-0.58
Group	-2.04	-0.16

Table A 11.3: Coefficients for models predicting overall gain scores: intervention and comparison groups (for Table 7.16)

	Unstandardized B Coefficients	Standardised beta Coefficients
(Constant)	20.73	-
Gender	1.25	0.08
Pretest	-0.57	-0.55
Group	-7.63	-0.61

Table A 11.4: Coefficients for models predicting listening gain scores: intervention and comparison groups (for Table 7.18)

	Unstandardized B Coefficients	Standardised beta Coefficients
(Constant)	13.80	-
Pre-Listening	-0.64	-0.62
Group	-5.02	-0.58

Table A 11.5: Coefficients for models predicting gain scores of intervention group
(for Table 7.20)

	Unstandardized B Coefficients	Standardised beta Coefficients
(Constant)	12.90	-
Admission Score	1.78	0.13
Pretest	-0.54	-0.46

Table A 11.6: Coefficients for models predicting listening gain scores of intervention group
(for Table 7.22)

	Unstandardized B Coefficients	Standardised beta Coefficients
(Constant)	8.69	-
First Gen Hi-Ed	-1.90	-0.20
Admission Score	1.85	0.20
Pre-Listening	-0.59	-0.52