THE EFFECTIVENESS OF ERROR CORRECTION DURING ORAL INTERACTION: EXPERIMENTAL STUDIES WITH ENGLISH L2 LEARNERS IN THE UNITED KINGDOM AND SAUDI ARABIA

Haifaa I. Faqeih, MA.

Doctor of Philosophy

University of York

Department of Education

January 2012

ABSTRACT

The current classroom experiments examined the effects of two types of oral corrective feedback (CF), recast and metalinguistic information, during oral production tasks on the learning of English modals (***will***, ***can*** and ***must***). These techniques were compared to an intervention with identical oral production tasks but in which CF was not provided. The study also investigated the extent to which instructional setting (EFL in Saudi Arabia and ESL in the United Kingdom) and learners' attitudes towards CF mediated the effect of CF on learning.

Pre-intermediate adult learners of English as a second language (ESL) in the United Kingdom (UK; n=36) and of English as a foreign language (EFL) in Saudi Arabia (SA; n=64) were randomly assigned to one of three conditions: metalinguistic information; recast, or task only (no CF).

Participants had four sessions within these conditions, over four consecutive weeks, delivered by a trained and experienced EFL teacher (the author).

In the UK context, learners' knowledge was measured, in pre, post and delayed posttests, using a free oral picture description, a timed grammaticality judgment and a written gap fill. In the SA context knowledge was measured by similar written gap fill and the same picture description, but also an elicited imitation, and un-timed grammaticality judgment with explicit knowledge probes.

In both contexts, an exit questionnaire (similar to Sheen's, 2006) was administered after each test to check awareness of the target feature being tested. In addition, an attitudinal questionnaire was used to measure the possible role of participants' attitude towards error correction and grammatical accuracy.

Tests of normality were used to decide whether parametric or non-parametric statistical tests were required. Equivalence between groups at pre-test was checked to determine whether actual scores or gain scores (or ANCOVAs, if parametric tests) should be analysed.

The results suggested that both metalinguistic information and recasts can be beneficial for the development of English modals, though effectiveness was influenced by the outcome measures used, the length of time between intervention and test, and the

context (UK and SA). Recast and metalinguistic information were generally found to be beneficial in most measures regardless of contexts. In most measures, task only group in the UK had no significant gains but in SA had significant gains. The study indicated that learners had an equal preference for recast and metalinguistic information CF in the EFL context but preference for recast was pronounced in the ESL context.

ACKNOWLEDGEMENTS

I am very grateful and thankful to Allah the Almighty for His help and endless support in enlightening the path of knowledge for me all the way from my childhood to this point of my life.

I would like to take this opportunity to show my sincere gratitude to those who helped me along the way. This dissertation would not have been possible without their support, encouragement, advice, and contribution. First and foremost, I would like to express my heartfelt gratitude to my thesis supervisor, Dr Emma Marsden, who guided me through the different stages of my PhD study with insightful comments, constructive feedback, and continuous support and encouragement.

I would also like to thank my two examiners, Dr Andrea Révész and Professor Leah Roberts, who were generous with their time and feedback on my thesis. I appreciate their insightful comments and suggestions on a number of issues raised by my analysis; all of which undoubtedly helped improve my thesis.

I would also like to thank Dr Danijela Trenkic for her remarkable comments on the documents presented in the TAG meetings and the upgrade viva. I wish to give special thanks to Dr Graham Low, who gave me some guidelines in designing the attitudinal questionnaire. Special thanks and appreciation devote to Dr Rosemary Erlam, who showed interest in my study, when part of it was presented at the TBLT conference 2009 and who sent me her comments concerning the application of the elicited imitation test.

This dissertation would not have been completed without the generous support and help from the teachers, students, and administrative staff at a language centre in the UK and a Saudi university, where I conducted the present study and collected my data. I am truly grateful for their immense interest in the study and their cooperation during the data collection period.

This dissertation is acknowledged to my family and friends. It is dedicated to the memory of my father, who had always been dreaming, since I was a child, of seeing me a doctor. His absolute love and support during my childhood were very remarkable. I owe a deep debt to my great mother whose unconditional love and never-ending support

encouraged me to finish my study. A great appreciation and tremendous sincerity go to my dearest husband, Professor Abdulmannan Saati, whose love and support have made me inspired, motivated and respected during this long journey. His long experience of being an academic professional has helped me morally and technically. Special heartfelt gratitude, love and thanks are dedicated to my intelligent sons, Dr Ammar, Dr Ammer, Eng. Aamer, my little chum Yousof, and Kholoud (my lovely daughter in law), whose love, smile, and encouragement were the intrinsic inspiration throughout this process. I deeply thank my sisters, sisters in law and brothers for their love and support.

I would also like to take this opportunity to recognize my friends and colleagues in Saudi Arabia and England for their outstanding support, kindness and prayers.

Last but not least, special thanks go to the Saudi government for their advice and financial support throughout the years of my journey. This dissertation was fully funded by the Ministry of Higher Education in Saudi Arabia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CF	Corrective Feedback
Cor	Grammatically Correct Items in GJT
EC	Error Correction
EFL	English as a Foreign Language
EI	Elicited Imitation
EI-cor	Elicited Imitation Correct Items
EI-inc	Elicited Imitation Incorrect Items
ES	Effect Size
ESL	English as a Second Language
GF	Gap Fill
GJT/UGJT	Grammaticality Judgment Test (Timed / Untimed)
Inc	Grammatically Incorrect Items in GJT
K	Number of items
KW	Kruskal-Wallis
L1	First Language
L2	Second Language
N, n	Number of participants
NNS	Non-native speakers
NS	Native speakers
Nt / ns	Not true/ not sure
M	Mean
MI	Metalinguistic Information group
MQ	Metalinguistic Questions
MW	Mann-Whitney
R	Recast group
PD	Free Oral Picture Description
UK	United Kingdom
Si	S: Student, where i=1, 2, 3, ... no of student

SA	Saudi Arabia
SD	Standard deviation
SLA	Second Language Acquisition
SPSS	Statistical Package for the Social Sciences
T	Teacher
TLs	Target Languages
TO	Task only group

CHAPTER 1: INTRODUCTION

Preface

On an anecdotal, personal level, as a teacher of English language to university students, I often wondered if the feedback I gave my students on English modals was retained in their memory. I began asking why some of these corrections and some aspects of grammar were remembered, while others were lost. I wanted to discover which corrective feedback (CF) techniques are more effective, specifically for the learning of English modals both immediately and in the longer term. To my personal knowledge, the two types of CF were common in my classroom and my colleagues': recasts (to lesser degree) and metalinguistic information, and I wanted to know which of these was more effective, and whether just using oral production tasks alone (relatively less common in my context - Saudi Arabian university English classes) would help learning.

1.1 Brief introduction to key terms, background, and rationales for the study

CF (defined as negative feedback) aims to make learners aware that their utterances contained errors or were somehow problematic (Mackey, 2006). However, it should be noted here that negative feedback is an interlocutor's interactional move that indicates explicitly or implicitly any non-target like feature in the learner's speech and need to be corrected.

There are many types of CF, and the current study sought to evaluate the effectiveness of two, which are briefly defined here. Recasts, as defined by Mackey and Goo (2007), are 'reformulations of learners' ungrammatical or inappropriate utterances which maintain their intended meanings' (p. 413). The other type of CF under focus in this study is a form of explicit feedback that provides metalinguistic information by giving grammatical rules or information related to the well-formedness of the learner's utterance (as defined by Ellis, Loewen & Erlam, 2006), henceforth 'metalinguistic information'. The two types of CF will be discussed and defined in greater depth, along with examples, in Chapter 3. In addition, studies that have looked at the types and

efficacy of CF, learners' noticing of feedback and the contribution of feedback to language acquisition in recent years will be reviewed in Chapter 3. The current study also sought to evaluate the effectiveness of recasts and metalinguistic knowledge relative to simply asking learners to engage in the oral production task without providing them with CF. This rationale will be developed further in Chapter 3, but is summarised here.

This section provides a brief summary of the background to perspective on errors and CF. Early in the twenties century, language errors were considered to be undesirable forms and it was the teacher's goal to reduce these errors by any means (George, 1972). However, in the early sixties, language errors began to be viewed by language experts in a more positive way, as being indicative of progression. Corder (1974) illustrated the significance of learner errors in several ways. He pointed out that learner errors are important for teachers as they indicate the amount of information that the learner has acquired, and teachers can then modify their instruction according to their students' needs. Errors were then seen as important, indeed inevitable, in the learning process. Similarly, Hendrickson (1978) stated that language errors are a natural part of learning and the systematic analysis of errors can help researchers and teachers to understand better the process of language acquisition. He emphasized five critical questions: (a) Should errors be corrected? (b) When should errors be corrected? (c) Which errors should be corrected? (d) How should errors be corrected? (e) Who should correct the errors? (p. 389). Along similar lines, Van Lier (1988) noted that in the late sixties and early seventies, teachers began to realize that errors might be more an indication of learners' efforts to form a new linguistic system rather than linguistic failure.

In contrast, educators and researchers have investigated the questions stated by Hendrickson for many years but Lyster and Ranta (1997) pointed out in their review that researchers were far from finding answers to these questions. There has been some divergence of thought regarding the effectiveness of feedback. For example, Krashen (1982) suggested that students do not need any feedback to progress. Contrarily, Lightbown and Spada (1990); Lydia White (1991); Carroll et al (1992) Long (1996); Lyster and Ranta (1997);Sheen (2004); Ellis (2009) suggested that feedback plays a crucial role in language learning, as it pushes learners to notice and attempt to say the targeted form, and therefore students may be more likely to repair their erroneous

utterances. In addition, the meaning-focused instruction has been questioned with regard to its effectiveness and research suggests that form-focused instruction can benefit language learners. The term Form-focused instruction (FFI) is defined by Ellis (2001, p.1) as "any planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic forms". Instructions that focus on form favour attention to linguistic structures within the context of meaning-focused, communicative activities (Ellis, 2001; Long, 1991, 1996).

The types of feedback that are most effective in terms of student learning have been investigated, for example, Carroll and Swain (1993); Ellis et al (2006) and Sheen (2006) suggested that learners would benefit from direct, explicit CF, whereas other researchers, such as Lyster and Ranta (1997), Oliver (2000), and Oliver and Mackey (2003) suggested that for particular forms in particular contexts feedback is effective when it is implicit, in the form of recasts. This debate can lead to confusion for example, as Lyster and Ranta (1997) pointed out "because of so many different approaches to feedback, second language teachers have trouble finding research that addresses practical issues of CF" (p. 38). A related practical and theoretical issue that the current study addresses is the relationship between learners' attitudes and the effects of CF.

To examine the potential influences of different learning contexts relative to learners' attitude, participants in this study were recruited from two contexts (ESL and EFL). ESL participants were studying English in a language institute in York for an average period of six months, and were exposed to oral production practice and interaction during their lessons. The EFL participants were studying in a Saudi university in Saudi Arabia and were used to a high proportion of grammar translation approaches with high levels of correction but with few opportunities for oral production practice and natural interaction in English. Another purpose the current study sought to determine whether attitudes are different in different contexts and the effectiveness of different types of error correction and oral production tasks on learners' achievement in formal language learning.

1.2 Overview

- Chapters 2, 3, 4 and 5 present a review of the relevant literature.
    - Chapter 2 examines oral CF from SLA theory perspectives.

    - Chapter 3 examines research relating to CF and SLA from pedagogical, classroom-based perspectives.
    - Chapter 4 considers the CF and individual differences.
    - Chapter 5 deals with the target form of the study 'English modals'.
- Chapter 6 presents the research questions.
- Chapter 7 presents the methods followed in the UK context and includes the descriptive statistics and data analysis used for the purpose of the current study.
- Chapter 8 discusses the experiment undertaken in SA and reports the descriptive statistics and data analysis for the purpose of the current study.
- Chapter 9 compares the results obtained in the two achievement tests that were used in both ESL and EFL contexts: the free oral picture description and gap fill tests.
- Chapter 10 provides an overall discussion for all results.
- Chapter 11 concludes and summarizes the main findings of the study, considers their theoretical and pedagogical implications, identifies a number of limitations of the study, and offers some directions for future research.

CHAPTER 2: CORRECTIVE FEEDBACK DURING ORAL PRODUCTION IN SECOND LANGUAGE ACQUISITION: THEORETICAL PERSPECTIVES

Introduction

It is widely accepted that that CF does not play an important role in first language acquisition (Krashen, 1982). Children successfully acquire their L1 through social interaction. Language acquisition, therefore, is often argued to take place implicitly while receiving language input, and CF does not play a significant or necessary role. However, for second language acquisition this is a point of considerable debate, certainly in instructed contexts. Some key aspects of this debate are reviewed in this chapter, by summarizing, and then discussing the relevance of, some theories of L2 learning. The theories have been selected on the basis that they all suggest that the three experimental treatments used in the current study (recast, metalinguistic information, and oral production alone) may be beneficial for language learning, and so each is useful in providing a theoretical rationale for the pedagogical interventions chosen. The theories / approaches reviewed include: Long's (1996) "Interaction Hypothesis", as it claimed that learning occurs when input is contained conversational and linguistic modifications; Swain's (1985) Output Hypothesis, as it suggested that language development is driven by producing output that may have a substantial role in transforming explicit knowledge into implicit knowledge; Schmidt's Noticing Hypothesis, (1990), as it suggests that CF may be effective by raising learners' awareness of features of the input and a mismatch with their output. Other sections in this chapter discuss the possible role of priming (speaker's repeated production of a previously spoken or heard structure across successive utterances) in learning and the different types of knowledge that are thought to be involved in second language learning. The aim of reviewing the theories included in this chapter is neither to falsify one specific learning theory nor to provide a complete description or critique of each theory in its entirety. Rather, specific aspects of these theories have been selected which are relevant to the current study. Nevertheless, some discussion of how the study can provide evidence which is compatible with one or more

of the theories will be included. Also, the current study may suggest that one or more of the theories may require further research and possible revision, and their relevance to classroom error correction research may be considered. The theories reviewed may also help the interpretation of the findings achieved in the current study.

It must be emphasized, however, that this is a ***classroom*** study, investigating the effectiveness of three *pedagogical* techniques in developing the acquisition of English modals for Saudi learners of English in relation to different contexts, ESL and EFL.

Section 2.1 discusses the Input Hypothesis (Krashen, 1980, 1985); section 2.2 presents the Interaction Hypothesis (Long, 1980, 1996); section 2.3 presents the Noticing Hypothesis (Schmidt 1990, 2001). The Output Hypothesis (Swain, 2005) will be presented in section 2.4, and skill acquisition theories (DeKeyser, 2007) in section 2.5. Section 2.6 discusses the role of the types of knowledge-implicit and explicit. The role of priming during interaction (discussed by McDonough 2006) will be briefly discussed in section 2.7. Finally, section 2.8 links the reviewed theories to the current study.

2.1 Krashen's Input Hypotheses and Related Arguments

It is commonly known that input, provided either by a teacher or by another learner in a formal or informal setting, is necessary for the acquisition of a second language. According to Corder, (1967) input is distinguished from intake. Input refers to what is available to the learner, whereas intake refers to what is actually internalized. The most influential theory of the role of input in SLA was proposed by Krashen in 1980. He argued that in order for L2 acquisition to proceed, learners must be exposed to comprehensible meaningful input which contains linguistic data slightly ahead of their current state of grammatical knowledge *(i + 1*) where (*i)* is the L2 learner's current linguistic competence, and (*i + 1)* is the next level of that competence achieved with comprehensible input (Krashen, 1985). He suggested that production of the target does not directly aid acquisition.

Krashen suggested that learners can make use of three kinds of contextual information: extra-linguistic information (learners' knowledge of the world and previously acquired linguistic competence); the input that can be available via

interaction; and interaction in which meaning has to be negotiated e.g. when there is a communication problem.

Given the fact that recast was supposed to inform the learner of the positive evidence (the correct forms of the TL) and does not interrupt the flow of communication, it may have the capacity to promote acquisition arguing that the positive (rather than the negative) evidence in recast that might help learners. Unlike metalinguistic feedback, the negative evidence is expected to promote learning but not acquisition.

Acquisition versus learning

Krashen distinguished between 'acquisition' and 'learning'. He believed that acquisition uses unconscious processes and that grammatical rules are not helpful. In other words, the implication of his theory is that a second language is acquired much more like the first language, in which parents do not focus on explicit instruction of the language, but instead focus on communication and meaning. It is further implied that if input is understood, and there is enough of it, the necessary grammar is automatically learnt (Krashen, 1985). Related to this is that Krashen argued that provision of error correction (=negative evidence) is not helpful, and does not improve second language performance. He believed that grammatical systems can be entirely learnt during communicative activities which encourage learners to focus on meaning rather than on form.

He believed that 'learning', on the other hand, is a conscious process that involves the memorization of many formal grammatical rules, and error correction can have a role in this. He believed that 'learning' leads to grammatical and mechanical knowledge of the language, but it does not lead to fluency (learners' ability to communicate smoothly and fluently).

Note that although Krashen distinguished between learning and acquisition, the current thesis does not attempt to do so, as it is now broadly accepted that this distinction is a matter of distinguishing between different types of knowledge such as implicit and explicit which will be addressed at the end of this chapter. The terms learning and acquisition will be used interchangeably in the current study.

Evidence supporting Krashen's Hypothesis

Krashen's hypothesis seems to be supported by some scholars in certain aspects. For example, Schwartz (1993) and Truscott (1996) also put forward similar theoretical idea suggesting that only positive evidence (a model form of the TL) is sufficient for learners to acquire L2, and that negative evidence (e.g., implicit/explicit feedback to any non-target like feature in learners' speech) does not help learning and that structures learned through error correction cannot become part of internal grammar. They propose negative effects of error correction, such as confusing the learners, causing over-use of a particular form and/or interfering with natural language acquisition processes. Others have perceived that excessive use of error correction can lead to lower motivation (e.g., Lightbown & Spada, 1993), in line with Krashen's notion of an affective filter. Following these arguments, classroom teachers should not focus intentionally or explicitly on errors of language form during class, but should instead provide comprehensible input to learners.

Challenges to *Krashen's* input hypothesis

Krashen's Input Hypothesis has been criticized for being lacking in empirical evidence and untestable owing to its vagueness (Mitchell & Myles, 2004). In addition, White (1987) criticized Krashen's hypothesis for not considering the possible benefits of providing rules, arguing that for some syntactic structures comprehensible input could not be counted on and that certain types of errors may need rules instruction. Others have pointed out that although interaction can solve communication problems through negotiation and increase comprehension, it does not mean that increased comprehension automatically leads to L2 acquisition; that is, learners may not necessarily retain the comprehended target language (Ellis, Tanaka, & Yamazaki, 1994).

2.1.1 Relevant Aspects to the Current Study

By now, the insufficiency of Krashen's Monitor Model as a whole is widely acknowledged. There is little current debate surrounding this model and the input Hypothesis in particular. However, the current study could contribute to the debate

surrounding Krashen's Input Hypothesis in several ways. It could provide evidence about:

a) Whether error correction generally is useful for learning a specific set of language forms via a comprehensible meaningful environment.

b) The type of knowledge which results from different kinds of error correction, by investigating the kind of tasks in which learners can demonstrate knowledge gained due to error correction. Specifically, the proposed study will inform our understanding of whether error correction which is more implicit and in line with L1 interaction (i.e. recasts), is more effective than error correction which is explicit and directly provides information about grammar (i.e. metalinguistic information). Krashen's notions imply that recasts should be more beneficial for acquisition and explicit knowledge provided in CF would only improve explicit knowledge and contribute only to learning. (Measuring explicit and implicit knowledge is discussed in more depth in section 2.6 below).

c) How learners feel about different types of error correction, e.g. whether they encourage or discourage them (see Chapter 4 for more information on learners 'attitude towards CF). The proposed study may also provide evidence about other negative effects of error correction.

2.2 Long's Interaction Hypothesis

The interaction approach was built on Krashen's claims. Long (1996, 2007) based his Interaction Hypothesis on the argument that input is important for acquisition to take place, particularly when learning an L2, by means of the conversational adjustments that occur when there is a comprehension problem. Long argued that input is not sufficient on its own for language learning to occur , that some types of negative evidence may be beneficial for acquisition and more attention should be given to the interaction that learners are engaged in. He updated his hypothesis in 1996, with a greater emphasis on how certain conversational strategies such as repetitions, confirmation checks, comprehension checks, clarification requests or recasts may help to solve communication problems. Given that interaction involves a number of components including negotiation, recasts, and feedback, Long argued that interactional feedback is a very important aspect

for language development because it raises learners' awareness of errors in their speech during meaning-based interaction, it helps them to notice mismatches between their inter-language and the TL, it encourages them to hypothesize the correct forms and test those hypotheses and to modify their IL (Long, 2007). In support of this approach, Long demonstrated that tasks involving a two-way exchange of information(teacher-student) lead to more conversational adjustments than do tasks involving a one-way exchange of information (teacher-centred).

For CF to be effective Long (2007) claimed that focus on form (interactional communicative instructions) and meaning should be provided simultaneously in a classroom context, and that recasts are the best strategy for the negotiation for meaning, because they are implicit, and thus do not interrupt the flow of interaction unlike explicit forms of feedback which do not assist learning as the flow of communication is interrupted.

Evidence supporting Long's Interaction Hypothesis

Long has received considerable theoretical and empirical support in the literature (e.g. Pica, 1994). On the theoretical ground, Doughty (2001) supported the use of recasts, which provide learners with opportunities to engage in 'form–meaning mapping'. Carroll (2001), also supported Long's updated version of the Interaction Hypothesis, which emphasized that feedback obtained during negotiation may facilitate L2 development as this is in line with the notion of 'failure-driven' learning. On the empirical evidence, Mackey (1999), in her investigation of question formation showed that interactors who engaged in directed negotiation were able to develop their L2 knowledge of question formation faster than non-interactors.

Challenges to the interaction hypothesis

Theoretical challenge was provided by Sheen (2006) on Long's view of the role of negative feedback as "somewhat narrow" (p.13). Long, (1996, 2006) has claimed that learners need to be equally attending to form and meaning for effective CF. Sheen viewed such a position as a problematic as it is often not clear whether the feedback is a result of communication breakdown, or it is teacher's choice. She also stated that in

many classrooms teaching contexts, teacher's choice for the corrective feedback is much more common than conversational feedback.

Empirical evidence, against the claim that implicit feedback is effective, is provided in the studies of Ellis (2006) and Sheen (2006) suggesting that the use of implicit CF does not promote acquisition. Long further argued that explicit type of feedback (corrective feedback forms that treat language as an object) are unlikely to assist learning because they interrupt the flow of communication and thereby lack the focus on form and meaning. In contrary, Ellis (2006) and Sheen (2006) suggested that the provision of a brief metalinguistic feedback by a teacher did not appear to interfere with the communicative flow of the activity in the classroom.

2.2.1 Relevant Aspects to the Current Study

The instructional materials and error correction techniques used in the current study put several aspects of Long's hypothesis and arguments into operation:

a) The study investigates whether interactional feedback may benefit learning (as it may, according to Long, enhance comprehensibility and, therefore improve learning).

b) The instructional materials developed implement Long's claim that for CF to be effective, it should be provided in activities and tasks that offer opportunities for negotiated interaction with a focus on communicating meaning.

c) Long (1996) argued that most of the studies on the type of task and negotiation of meaning between pairs inside and outside the classroom, have been carried out in Western educational institutions and little is known about the kinds of negotiation of form and meaning present in L2 interactions in other contexts. Mitchell and Myles (2004), accordingly, suggested this as a possible future avenue for error correction research. This is a relevant issue to the current study that investigates the influence of an EFL context on two CF strategies (recast and metalinguistic correction).

2.3 Schmidt's Noticing Hypothesis

Given that regular interaction (as in Long's (1996) interaction hypothesis) depends on works through learner-internal factors such as noticing, empirical research investigated the relationship between noticing and learning in these contexts. This

relationship was first investigated by Schmidt (1990, 1995) who argued that learners need consciously to essentially notice the gap between what they hear or see in the input and what they produce. The process of converting the input into intake was the base of Schmidt's Noticing Hypothesis (1990). Following Schmidt noticing hypothesis is the acquisitional value of interactional feedback which can help direct the learner's attention towards the mismatch between the target input and their own interlanguage form (based on 'noticing the gap,' Schmidt and Frota 1986). In his own experience, Schmidt (as an American learner of Portuguese in Brazil) found that instruction, interaction, and correction influenced his learning of Portuguese and that the target features in the input that he had consciously attended to during the interaction with native speakers were almost always acquired.

The relative association between noticing and L2 development in the presence of interactional feedback was further examined by SLA researchers. The data in Mackey (2006) pointed to an association between the provision of interactional feedback and learners' reports about noticing the target linguistic forms in L2 classroom context, suggesting that when interactional feedback (e.g., recasts and negotiation) is provided on problematic L2 forms, learners report noticing those forms more than when feedback is not provided. Although Mackey's data suggested a relationship between feedback and noticing and a possible link between noticing and learning, there was no clear indication that learning follows noticing for some learners' reported noticing but did not develop, and a few learners in the control group developed but did not report noticing the target items. This is a supportive point to Schmidt's (1995) warning regarding the effectiveness of noticing on learning the target items. However, Schmidt claimed that some forms may not be noticed until learners are developmentally ready, and that noticing in the input could be affected by different factors: instruction, frequency, perceptual salience, skill level, task demands and comparison.

Support for Schmidt's Noticing Hypothesis

Empirical research in support of noticing in L2 acquisition is that of Schmidt and Frota (1986), in which Schmidt analysed his own acquisition of Portuguese and found that the ability to produce L2 grammatical forms was not directly related to the grammar that he had received as input, but the grammar that he had noticed. Schmidt's arguments have been widely used in instructed L2 research. Noticing, from a theoretical point of view, according to Gass, (1988) is considered the first stage of language acquisition, but described by Batstone (1994) as the "gateway to subsequent learning" (p. 100). Lynch (2001) argued that noticing is an important component of successful language learning and VanPatten's (2007) Input Processing relies on the basic notion of attention to forms to be learnt.

Concerns and challenges about Schmidt's Noticing Hypothesis

The most detailed criticism of Schmidt's Noticing Hypothesis comes from Truscott (1998) who concluded that "the foundations of the hypothesis in cognitive psychology are weak and is not based on any rational theory of languages" (p. 104), and the Hypothesis is too vague to determine what learners must notice. Truscott argued that noticing is necessary for metalinguistic knowledge but not language competence. Truscott (1998) therefore suggested that to enhance our understanding of noticing in SLA, further investigation is needed using, for example, more exact testing of the type of knowledge gained.

2.3.1 Relevance to the Current Study

The treatments in the current study operationalized the central claim of the Noticing Hypothesis as CF may help learners to develop the TL by making them notice the mismatch between their IL forms and the TL forms. A number of researchers have taken learners' responses that incorporate the CF as evidence of noticing the feedback (e.g., Chaudron, 1977; Lyster, 1998; Lyster & Ranta, 1997; Mackey et al., 2000; Sheen, 2004, 2006). Of course, we cannot take the absence of responses as evidence that learners have failed to notice feedback, nor can we ignore the possibility that learners might copy

feedback without true understanding (e.g., Gass, 2003). Nonetheless, researchers have argued that uptake implies that learners have noticed the corrective function of recasts (Lyster & Mori, 2006) and that "a reformulated utterance from the learner gives some reason to believe that the mismatch between learner utterance and target utterance has been noticed" (Lightbown, 1998, p. 193). In the current study the recast treatment contains positive evidence that may help learners to notice a gap between the target language (TL) and their inter-language (IL) knowledge. It is, however, possible that recasts are too indirect for such noticing to occur, as they do not indicate the 'gap' sufficiently clearly or reliably to the learner. Metalinguistic information, on the other hand, provides a clear indication to learners that there is a gap between their IL and the TL, which in turn raises learners' awareness of what they are unable to say in the target language and could facilitate L2 acquisition.

2.4 Swain's Output Hypothesis

Based on both formal and informal observations in the context of immersion programs in Canada Swain, (1985, 1995, 2000, 2005) proposed the importance of the Output Hypothesis, arguing that pushing learners to produce output assists language development as it provides the learner with the opportunity to practice, to notice the gap between their IL and their TL, and to offer them the chance to test their IL hypotheses. She also claimed that language production forces learners to move from the semantic strategic processing to syntactic use of language as a result of the three functions of output (specified in Swain, 1995). Swain attributes considerable importance to the provision of CF, such as clarification requests, as these can promote pushed output and thereby help learners to develop their inter-language.

2.4.1 Relevant Aspects of the Output Hypothesis to the Current Study

The Output Hypothesis provides some justification for using oral production practice, in that opportunities for the noticing, hypothesis testing and reflective functions of output will be provided. Also, the hypothesis provides some justification for the provision of feedback from interlocutors and teachers as this may inform learners of their initial problematic utterance. This may create the condition for cognitive comparison

which enables learners to notice the gap between their inter-language output and the target language input (Izumi, 2003). In particular, metalinguistic feedback and perhaps to a lesser extent recasts (depending on how they are perceived and processed) may enhance the metalinguistic function of the learner's output, and the noticing /triggering function allowing learners to attend to comprehensible useful information in the input and modify their output by reformulating the utterance in a more target-like way.

2.5 Information Processing and Skill Acquisition Theory

Skill acquisition theory describes the process of progressing different representations from the cognitive phase to autonomous by practicing. This process helps learners to move from initial to advanced levels in learning various skills. Three stages for the skill acquisition have been posited by different researchers: cognitive, associative, and autonomous (e.g., Fitts and Posner, 1967); or declarative, procedural and autonomic (e.g., Anderson, 1982; Anderson et al. 2004). Byrne (1986) had three distinctive ways for the kills to be acquired represented in presentation, practice and production (PPP).

DeKeyser (2007) along with others highlighted three different stages in the skill acquisition process for achieving the goals of language learning, namely the declarative, procedural and the automatic stages. At the declarative knowledge stage, learners are initially presented with some information about a skill, e.g., rules about certain aspects of the language they are learning. In the next stage, learners learn how to apply these rules to specific sentences, whether it is in comprehension or production. "This procedural knowledge consists of very specific rules and can be used fast and with a lower error rate, but the disadvantage is its lack of generalizability" (DeKeyser 2007, p.3). The last stage is 'automatization', which is considered rather difficult, for it needs a large amount of practice to decrease the time required to complete the task, the percentage of errors and the amount of attention required. As a consequence of practice, the procedural knowledge becomes automatic DeKeyser (2007). He further suggested that the changes occurring in the process of learning start with a rather rapid step (e.g., at the declarative knowledge stage) followed by a much slower process (e.g., at the 'automatization' process).

The process of L2 learning has been dealt with from a cognitive psychologists' point of view by Shiffrin and Schneider (1977). They suggested that information is

processed either by controlled or automatic processing. The controlled processing requires a lot of attentional control and is constrained by limitations of short-term memory. On the other hand 'automatized' sequences are stored as units in the long term memory and can be easily accessible when a situation requires it. This is in line with DeKeyser (2007)'s arguments that repeated activation or practice is needed for knowledge to move from being controlled to automatic processing of learning.

2.5.1 Aspects of Skill Acquisition Theory Relevant to the Current Study

Skill Acquisition Theory would predict that the provision of metalinguistic information or/and recasts (when they are likely to enable learners to focus on form and to notice errors in their interlanguage production as suggested by Doughty, 2001) should represent the first stage of learning i.e., it should help learners to formulate declarative knowledge. This proposition has been given support by Révész (2012) whose study appears to be in line with the view that recasts can facilitate the encoding of new declarative knowledge.

If learners, in the current study, are then given sufficient time to practice this knowledge, this may become 'proceduralised' and then 'automatized'. If learning is documented in the experimental groups more than the task only group, then this could provide evidence which is compatible with DeKeyser's skill acquisition theory. However, the necessary amount of practice to reach the automatized stage could be beyond the scope of the current classroom study, in which learners are given relatively few opportunities to produce the target form. If so, automatised learning may not be observed (see section below on measuring explicit and implicit knowledge). Nevertheless, such findings could inform us about the issue of how much practice is required before learning can be observed according to the measures used in the current study. The current study could also provide some indication of whether the knowledge gained, tends to be more explicit or implicit and this may enhance our understanding of the outcomes of skill acquisition, as "these forms of knowledge may have been acquired implicitly or explicitly, and may or may not have been transformed from explicit to implicit 'proceduralised' and 'automatized' through large amounts of practice, or from implicit to explicit (analysed) through reflection" (DeKeyser & Juffs, 2005, p 438).

2.6 Explicit and Implicit Knowledge

As the current study set out to use a battery of measures that may tap into different knowledge types, and as the treatment types are often referred to as explicit or implicit (metalinguistic information and recast CF, respectively) , a brief review of the terms explicit and implicit knowledge is provided here.

Explicit knowledge is defined in Ellis (2004) as "the conscious awareness of what a language or language in general consists of and/or of the roles that it plays in human life" (p.229). Implicit knowledge, in the other hand, is the unconscious awareness, automatic, and implicit kind of knowledge that used in normal communication of fluent speakers (Krashen, 1982, Bialystok, 1978).

In terms of explicit and implicit learning, it is important to keep in mind that most learning activities do not promote exclusive acquisition of either implicit or explicit knowledge. For instance, explicit learning activities usually provide input, which may result in acquisition of implicit knowledge. On the other hand, activities aimed at the acquisition of implicit knowledge may result in explicit knowledge if the learners become aware of the learning target. This synchronicity of different types of knowledge of the same phenomenon and their mutual influence make it very difficult for researchers to assess the nature of the knowledge that has been learned in a specific task.

The relationship (or interface) between these two knowledge types is the subject of considerable debate (e.g. Hu (2002) and Ellis (2004, 2005)). The non-interface position is that implicit and explicit knowledge are unrelated, involve different acquisitional mechanisms, are stored in different parts of the brain (Paradis, 1994) and are accessed during performance by different processes, either automatic or controlled. In this position, conscious knowledge cannot be converted into acquired competence and thus explicit grammar instruction does not become the basis of acquisition (Kubota, 2000).

The strong interface position maintains that practice can turn learned (explicit, declarative) knowledge to acquired (implicit, automatised) knowledge. This position was first formally advanced by Sharwood Smith (1988) and then promoted by DeKeyser (1998). Bialystok (1982, 1985) argued for a direct interface between explicit and implicit knowledge, claiming that L2 learners can begin with explicit knowledge that can through

intensive comprehension practice be proceduralised, automated, and converted into implicit knowledge.

The weak-interface position proposed by Rod Ellis, (2005) argues that explicit knowledge can help to develop implicit knowledge. For example, explicit knowledge of a linguistic feature can be converted to implicit knowledge via practice only if a learner is developmentally ready to acquire the linguistic form. Ellis (2005) argued that explicit knowledge may make the relevant features salient and enable learners to notice the features in the input and the difference between their existing representation of the features and what was actually noticed in the input. The transformation process of explicit knowledge into implicit does not necessarily happen (Kubota, 2000, p.27).

The current study does not intend to test which of these 'interface' positions is valid. The summary above is provided as the results and their potential implications are discussed in the light of these issues.

Defining and operationalising explicit and implicit knowledge is also a subject of some debate. One of the most recent treatments of this issue is by Ellis (2005), and this is summarised here. Ellis suggested seven characteristics for distinguishing explicit and implicit knowledge.

I) Awareness: this can involve two different kinds: intuitive awareness of implicit grammatical rules and a conscious awareness of why a sentence is ungrammatical.

II) Nature of knowledge: this can be either declarative knowledge (e.g., abstract rules and examples) or procedural knowledge (e.g., automated).

III) Extent of systematicity and certainty implicit knowledge is considered to be more systematic and is employed with greater certainty.

IV) Accessibility of knowledge: this can be accessible in online communication where little time is available (implicit knowledge) or during controlled situations with sufficient time (explicit knowledge). (As DeKeyser's (2003) and Hulstijn (2002) emphasize, practice can speed up the execution of rules to some extent, and so explicit knowledge can, once converted to automatised knowledge, be available during online communication).

V) The use of knowledge is assumed that the type of knowledge depends on the tasks learners are asked to perform. If learners are asked to detect and correct a written

task, they will tap into their analysed knowledge, whereas in an oral task they will tap into automatic knowledge (Ellis 2005).

VI) Self-reporting: whereas learners tend to put their explicit knowledge into words, they cannot verbalize their implicit knowledge.

VII) 'Learnability' and age: explicit knowledge can be learned at any age but implicit knowledge can be learned only within the critical period. This issue is controversial. For example, Bialystok (1994) did not indicate a certain age for learning explicit knowledge but claimed age limitations on L2 learners' ability to learn implicit knowledge.

2.6.1 Measuring Explicit and Implicit Knowledge

In order to investigate learners' production of the targeted linguistic items, a number of different features have to be taken into consideration in designing the type of instruments used to measure accuracy of implicit and explicit knowledge, which will be described in this section.

Type of responses

The type of responses to the different outcome measures used to gauge the relative effect of CF were classified in Norris and Ortega (2000): 1) metalinguistic judgments (or GJTs) responses require learners to judge the grammaticality of a target structure; 2) constrained constructed responses require learners to produce the tested forms in tasks where the use of the target structure is essential; 3) selected responses that require learners to choose the correct answer among several alternatives; and 4) free constructed responses where learners are required to produce the target language freely without any constraints.

Based on the above classification, Norris and Ortega (2000) found that effect sizes associated with metalinguistic judgments and free constructed-response measures were substantially lower than those with selected-response measures and constrained constructed-response measures, whereas Li, (2010) found different results in the Meta analyzed studies that measured learners' achievements through free constructed responses produced larger effects than those that employed constrained construction responses or metalinguistic judgment tests. This might be due to the fact that during free

constructed responses, learners might "avoid producing structures about which they are uncertain, which leads to a lower error rate. On the contrary, controlled construction responses and metalinguistic judgment measures provide obligatory contexts for the tested structures, in which conditions learners do not have the freedom of choosing particular items to answer" (p.345). Similar finding was revealed in Lyster and Saito's (2010) meta-analysis suggesting larger effects for free constructed-response measures than for constructed-response measures and metalinguistic judgments. The tentative patterns, in the reviewed three meta-analyses justify the demand for "further research into the variable effects of CF as measured by production tasks that are more or less controlled as well as by other tasks designed to assess both implicit and explicit knowledge of the target language"(p.292).

Design of implicit and explicit outcome measures

Different criteria for distinguishing implicit and explicit knowledge have been set by Ellis (2004) according to how each test mapped out on these criteria. When measuring their implicit knowledge, learners are using their feelings, they do not refer to their metalinguistic information and they are under pressure. In contrast, tests of explicit knowledge aimed to measure learners' use of rules under no pressure and encourage the use of their conscious focus on form and metalinguistic knowledge. To develop a battery of tests that would provide relatively separate measures of implicit and explicit knowledge, Ellis (2005) conducted a study included a total of five different tests in the form of oral imitation (defined as Elicited Imitation by Erlam 2006), oral narrative, timed GJT, untimed GJT, and metalinguistic knowledge tests. Ellis designed those tests in accordance with four of the criteria for distinctive implicit and explicit knowledge (discussed in Ellis, 2005, p.148). Following some of the criteria established by Ellis (2005), a total of five tests were designed for the purpose of the current study.

Test battery of implicit knowledge

As the current study set out to use a battery of measures that may tap into different knowledge types, and following the criteria mentioned above, a total of three tests were designed to provide measures of learners' implicit knowledge (e.g., elicited

imitation, picture description and timed grammaticality judgment tests) and two tests were designed to provide measures of learners' explicit knowledge (e.g., gap fill and untimed grammaticality judgment tests). The following sections review each measure separately.

Elicited imitation test

Elicited imitation is a type of test that requires participants to repeat back a spoken stimulus sentence as accurately as possible (Naiman, 1974). The test for the current study was designed to provide a measure of learners' implicit knowledge as the participants would rely mainly on feel, they would be under pressure to perform in real time, would focus primarily on meaning, and would have no reason to access their metalanguage (Erlam, 2006). Based on Erlam (2006), there are certain features that distinguish reconstructive test from a test that might allow learners to rely on simple rote repetition of target stimuli. A reconstructive elicited imitation test requires test takers to focus attention first on the meaning of the utterance before repeating it. It is necessarily to have some grammatical and others ungrammatical sentences. Spontaneous correction of incorrect sentences is a powerful indication of learners' internalization of targeted language structures not rote repetition of stimuli. This type of test was designed and administered in SLA research (e.g., Ellis et al, 2006; Erlam & Loewen, 2010) to provide a measure of learners' implicit language knowledge.

Choice of test items in EI test

In order to finalize the choice of test items, several factors should be considered. Erlam (2009) piloted the test statements on native speakers. She indicated that only the items for which native speakers produced the targeted structure were retained as test items. Statements should be designed to whether participants would agree with or not. They should be loosely organized around themes and grouped together to reduce learners' attention to form. Each statement should be repeated after making the decision about the truth value of the items participants heard. This way reduces the likelihood that they will explicitly focus on linguistic form and thus access explicit language knowledge.

Length of syllables in EI test

In most elicited imitation studies sentence length is a variable that needs to be considered in relation to a specific population and test design rather than as an absolute measure (Erlam, 2006; Munnich et al, 1994). Previous research (e.g., Bley-Vroman & Chaudron, 1994) stressed on sentence length and the role it plays in preventing the possibility that participants repeat exactly what they have heard. Hameyer (1980) found that sentence length determines to what extent the test is reconstructive, as it correlates with the number of syntactic and semantic changes participants made.

Hamayan et al (1977) used sentences that are an average of nine syllables in length with Arabic learners of English across several age levels. However each sentence was accompanied by a short explicative sentence of about five syllables. Munnich et al (1994) used sentences of equal syllables (fifteen) with advanced Japanese learners. Erlam (2006) used varied statements ranged from eight to eighteen syllables averaged 13.53. She found a small and non-significant negative relationship between syllable length of individual test items and participants' ability to repeat grammatical structures correctly and correct ungrammatical, $r=-0.28$, $p=0.11$. More research on the relative effects of the length of test items and learners' repetition suggested that some adult ESL speakers could only repeat sentences of seven to eight syllables in length (for example Perkins et al., 1986). "So the general range of sentence length found in the literature for adult testing is between six syllables and nineteen syllables" (Graham, McGhee & Millard 2010, p.59).

Picture description test

The majority of the studies included in Norris and Ortega's (2000) meta-analysis used measures that 'required the application of L2 rules in highly focused and discrete ways' (p. 483). The over-reliance on this type of measurements could be a major weakness that FFI research has to overcome (Doughty, 2003). Since then, there has been a tendency to use oral tasks, mainly oral picture description tasks, as dependent variables. These tasks may differ in terms of the required response types (i.e. constrained vs. free) and therefore vary in their validity as measures of implicit knowledge, they are considered to be more appropriate than metalinguistic tasks (e.g. grammaticality judgments, or text completions). However, although oral narrative tasks are often thought

to elicit implicit knowledge, this kind of task does not guarantee that learners will not access some of their explicit knowledge (Ellis, 2004; Erlam, 2006; Sheen, 2007).

Picture description tasks were used in SLA as a measure of implicit knowledge (e.g., Lyster, (2004); Ammar & Spada, (2006); Ammar, (2008)). To measure the effects of implicit and explicit CF techniques on the participants' knowledge of English possessive determiners Ammar and Spada, (2006); Ammar, (2008) presented six pictures to each learner individually, one picture at a time. All of the interviews were tape-recorded, transcribed, and analyzed in terms of grammatical and ungrammatical use of possessive determiners. The total use of the target feature on the oral picture description varied from one learner to another. The accuracy ratio for each of the students was calculated by dividing the number of correct third person uses by the total production of third person (correct or incorrect).

Test battery of explicit knowledge

Metalinguistic questions test

Metalinguistic questions may contribute to the development of explicit knowledge as learners' awareness could be raised to the problematic part when they are asked to state directly what kind of knowledge (i.e., feel or rule) they have used to make the judgment. This process is only available on an untimed GJT and it "could help to increase the construct validity of GJTs as a measure of explicit knowledge" (Ellis 2004, p.265). A supportive finding reflects the difference between implicit and explicit knowledge found in Green and Hecht (1992) suggesting that there was a gap between learners' ability to correct errors and to verbalize the rules involved in their investigation of English grammar via GJT test. Hu (2002) suggested that when learners are asked to explain grammatical features in a task, they may possess explicit knowledge of a specific rule but they fail to state it well because they lack the necessary skill to talk about language. The provision of rules could be a source of difficulty for deriving a quantitative measure "given that learners vary in the degree of precision and accuracy with which they are able to state a rule"(Ellis, 2004, p.250).

Gap fill test

Although gap fill test measures students' explicit knowledge assuming that leaving the blanks and giving space of time might force learners to refer to their explicit information. Ellis, (2005) suggested that "fill-in the blank might invite the use of explicit knowledge, but it does not guarantee it, as learners are obviously able to complete the exercise by drawing on their implicit knowledge"(p.147). DeKeyser (1995) examined the effects of two kinds of form-focused instruction (explicit-deductive and implicit-inductive) on two kinds of rules in an artificial grammar by asking learners to complete fill-in-the-blank tests in order to demonstrate their understanding of the grammatical rules. The learners in the explicit-deductive condition provided clear evidence of their ability to produce the simple categorical rules in new contexts and did better than the learners in the implicit inductive condition.

The Gap fill test was also used by Macrory and Stone, (2000) to measure students' explicit knowledge of the French perfect tense and learners' understanding of grammatical rules. The study found weak relationship among students' performance in the gap fill exercise, and their use of the tense in free oral and written production. Learners were able to supply an auxiliary in the gap fill exercise, but omitted it in their free production.

Test battery of both explicit and implicit knowledge

Ggrammaticality judgment tests (GJTs) have been used in SLA research to provide information about L2 learners' linguistic ability. The function of grammaticality judgment test varies on whether it measures explicit knowledge, or implicit knowledge. The major definition of GJT is to "decide whether a sentence is well-formed or deviant" (Ellis, 1991, p.162). A part from this definition, Loewen (2009) has set a number of features for this type of tests, that may influence the type(s) of knowledge learners draw on when judging a sentence, such as (1) making preference judgment; (2) limiting the amount of response time; (3) providing learners with dichotomous or multiple-choice response options; (4) locating, correcting and describing the errors; and (5) indicating the degree of confidence in the judgment (p.94).

However, the type of knowledge a GJT measures differed insofar as the timed GJT was predicted to measure primarily implicit knowledge, whereas the untimed GJT was predicted to measure primarily explicit knowledge (Ellis, 2005).

Timed grammaticality judgment test

Timed GJT is predicted to measure implicit knowledge for it encourages the use of feel, time-pressured, and little need or opportunity to access metalinguistic knowledge. However, Ellis, (2004) hypothesized that when learners are asked to judge the grammaticality of a sentence speedily, they are more likely to rely on implicit knowledge (though DeKeyser (2003) notes that time pressures do not guarantee a measure of implicit knowledge) it is possible that some learners may access explicit knowledge even if they are under time pressure, Ellis (2004). In order to use timed grammaticality judgment test, Ellis (2004) pointed out two principal processes that should be involved: 1) semantic processing (understanding the meaning of the sentence) and 2) noticing (searching for the incorrect part in the sentence).

To calculate the time border for the timed grammaticality judgement test, 20% was to be added to the average native response time taken so as to allow more time for the slower processing speed of L2 learners (Ellis, 2005). Table 2.1 demonstrates the range of time allocated in different studies.

Table 2.1 Studies used timed tests

Studies	Tests	Time
Bialystok, 1979	Timed Grammaticality Judgment	3 seconds per item
Han, 2000		3.5 seconds per item
Loewen and Nabei, 2007		Time range from 1.8 - 5 seconds
Mandell, 1999		10 seconds for each item (to read, judge, and indicate the problem)
	Dehydrated sentences	15 seconds for the Dehydrated sentences
Sheen, 2007	Speeded dictation test	8 minutes for 14 items = 34.28 seconds per item
	Error correction test	15 minutes for 17 items = 52.9 seconds per item

Untimed grammaticality judgment

To measure learners' explicit knowledge, the untimed GJT is predicted to measure explicit knowledge as it encourages a high degree of awareness, unpressured, has a controlled access to explicit knowledge, and predicted that responses would likely involve metalinguistic knowledge Ellis et al (2005). Although unlimited time given to perform untimed GJT test, learners may rely on their implicit knowledge to judge a sentence. However, the construct validity of untimed GJTs as measures of explicit knowledge needs to be considered in relation to the specific tasks learners are asked to perform. Most of the studies that focused on investigating explicit knowledge via GJT tests required learners to perform several operations: a) identification of the ungrammatical sentences, b) correction of the errors, c) provision of rules and d) to indicate the degree of certainty of their judgment (Ellis, 2004, p.294). Other studies required provision of rules (for example, Green & Hecht's (1992); Masny's (1987)). To measure learners' explicit knowledge, the untimed grammaticality test was used in SLA research (e.g., Ellis et al, 2006; Erlam & Loewen, 2010) although researchers pointed to

"the importance of including measures of both types of knowledge (implicit and explicit) in experimental studies" (Ellis et al, 2006, p.339).

Grammatical and ungrammatical sentences in GJT tests

Measuring grammatical and ungrammatical sentences in the GJTs was examined by Ellis (2005) and found that they measure different types of knowledge. Grammatical sentences provide a measure of implicit knowledge as their score correlated more strongly with the imitation test, oral narrative test, and timed GJT than with the metalinguistic knowledge test. In contrast, the ungrammatical sentences provide a measure of explicit knowledge as their score correlated less strongly with the other tests, especially the imitation and oral narrative tests**.**

Ellis et al, (2006) distinguished between these two constructs in the oral imitation and untimed GJT tests, and found significant difference between the grammatical items and the ungrammatical items in the oral imitation and untimed grammaticality judgment tests. Similar result to that of Ellis et al (2006) was found in Loewen and Nabei, (2007) indicating that the ungrammatical items were significantly higher than that for grammatical items on the untimed GJT, but higher significant scores for the grammatical items than those for the ungrammatical items on the timed GJT.

2.6.2 Summary of Relevance of 'Explicit and Implicit Knowledge' to the Current Study

The current study could provide evidence in line with any of the interface positions. For example, this study might contribute some evidence that the explicit information (via the metalinguistic information CF) may facilitate the acquisition of the target feature. If such explicit information leads to gains on measures that are thought to tap into implicit knowledge (see sections 7.8.1, 7.8.2, 8.10.1 & 8.10.4), then this could be evidence compatible with a strong or weak interface position. As it was mentioned previously, measures designed to tap implicit knowledge are unlikely to be pure measures of implicit knowledge, so one needs to be cautious with making such a conclusion.

If the metalinguistic information group showed gains on the explicit measures but not on the implicit measures, this would be evidence in support of a non-interface

position, at least within the context of the amount of exposure provided in the current study. It would not provide much information about the validity of the non-interface position. The development of implicit knowledge tends to require a lot of time. Aanother interpretation for the disassociation between explicit and implicit could be based on cognitive psychology's perspective in terms of the notion of transfer-appropriate processing (see Morris, Bransford & Franks (1977) for more information on this notion). In this regards, EFL/ESL learners may appear to have the necessary knowledge to make correct responses but they are unable to transfer this knowledge to certain types of grammatical problems.

Another relevant point of this debate is that the type of knowledge that learners use (and therefore demonstrate learning of) depends on the tasks they are asked to perform (Ellis 2005). This study therefore operationalized a battery of measures which are thought to tap into different knowledge types (see sections 7.6 and 8.6 for a full discussion about each outcome measure). For example, one test required learners to detect and correct ungrammatical features without a time limit and indicate the certainty of their decision. Another test required learners to produce oral language during free oral interaction, thus exerting both time and communicative pressure. This battery of measures was designed to give some indication of whether learners' knowledge of English modals before and after the treatment was explicit (declarative) or implicit (automated).

2.7 Priming

'In the context of language use, priming refers to the phenomenon in which prior exposure to language somehow influences subsequent language processing, which may occur in the form of recognition or production' (McDonough & Trofimovich, 2009, p.1). Semantic or syntactic priming in both language comprehension and language production is when the target is recognized or produced, more quickly and accurately if it is semantically or syntactically similar to a previously encountered word or syntactic structure 'the prime'(Bock, 1986; Branigan et al.,2000). McDonough and Mackey (2006) referred syntactic priming to a "speaker's tendency to produce a previously heard or produced structure across subsequent utterances" (p.711). McDonough (2006) argued

that syntactic priming can support L2 acquisition by enabling learners to choose between a simple and an advanced form, or between a non-target-like form and a more appropriate form. In other words, when syntactic priming occurs, the structure is temporarily available, and the learner is likely to produce the more advanced or appropriate form in subsequent utterances.

Given that the priming effect implied in the processing of one item influences the processing of another item, learners' proper reaction and the representation of the model forms of the target in the recast condition and the correct subsequent utterances in accordance to the provision of metalinguistic information may prime the target forms.

Empirically, syntactic priming (or learners' response to recasts) has been shown to occur in second language (L2) speech production and recent interaction studies have reported that English L2 learners often repeat lexical items that were initially produced by an interlocutor during syntactic priming activities (McDonough and Mackey 2006). The researchers investigated the immediate full or partial responses to recasts and learners' production for a new utterance using the syntactic structure that was provided in the recast, either immediately or a few turns later. It was found that there was significant relationships between recasts and primed production and thus this positive relation were predictive of ESL question development.

The effect of the metalinguistic CF in Sheen (2006) suggested that metalinguistic comments might have been primed although more time was needed to use the explicit information regarding the target structure.

A limitation on the role of priming, suggested in Kim and McDonough (2008), in that the knowledge gained may not be generalisable to other lexical items, as EFL learners produced more passives when their prompts had the same verb that previously occurred in the researcher's passive sentence. As a follow up to the previous thought, metalinguistic information tends to be generalisable possibly because the information provided may have a priming effect, and increasing the likelihood of noticing features in the input establishment of expectations.

2.7.1 Aspects of Priming Relevant to the Current Study

Priming is thought to be one mechanism which is involved in learning, this brief review was simply provided because priming may be one explanation of the findings from this study; if it is found that recasts or metalinguistic information do facilitate short and/or prolonged learning. If the provision of CF promotes this mechanism via the different outcome measures, then this explanation will be re-visited in the discussion.

If the uptake of learners in the experimental groups were included in the study, data could have been analysed to investigate whether responses to recasts and reaction to the grammatical comments immediately or few turns later were primed and thus signified learners' development of English modals.

2.8 Summary of How the Reviewed Theories are Relevant to the Current Study

The range of theories and models of learning and knowledge discussed above have been selected as they do one or more of the following: 1) justify the choice of interventions; 2) offer potential explanations for possible findings from the current study; 3) justify the outcome measures used. (It is re-iterated that the aim is not to test any of these theories or models - the design of the study does not allow this). Key aspect of the chapter was:

1) The findings may be compatible with Krashen's claim that recasts, which are more implicit, would be of greater benefit in the longer term for acquisition, and that metalinguistic correction, which is explicit, and directly provides information about grammar, would only improve explicit knowledge about the language, and this knowledge would not be accessible during certain tasks.

2) The feedback techniques provided in my study may draw learner's attention to mismatches between input and their output. This study, as it has a control group that did the task without CF, might provide evidence in line with Long's claim that negative feedback obtained during oral communication may facilitate L2 development at least for certain language aspects. It might also inform his claims about the efficacy of recast in facilitating acquisition via a conversational exchange.

3) The findings may be in line with the Noticing Hypothesis that suggests that CF contains positive evidence, such as recasts, that may be useful in giving learners the opportunity to notice a gap between the target language and their inter-language knowledge. Metalinguistic information also provides a (possibly clearer) indication to the learner that there is a gap between their erroneous production and the TL; it may also assist learners in developing awareness at both the level of noticing and understanding. This is discussed in more depth below, with particular reference to the role of the task only group.

4) Gains in the CF groups, in particular the metalinguistic group, may support Skill Acquisition theory, particularly if the gains of the metalinguistic group are greater than the recast and task only groups. The recast, and even task-only, group may also develop some explicit knowledge, observable on particular tests. If gains are seen in measures that are thought to tap explicit knowledge and these are accompanied by gains, perhaps in the longer term, in measures of more implicit knowledge, this could provide evidence for an interface between explicit and implicit knowledge. On the other hand, gains in explicit knowledge only would suggest that within the scope of this study, there was no interface with implicit knowledge, and gains in only implicit knowledge, without gains in explicit knowledge, would be evidence in favour of a non-interface position.

5) The findings of the current study may support Swain's (1985, 1993, 1995) output hypothesis, which claims that by producing the target language, L2 learners may notice what they cannot express in the target language. This type of noticing may make L2 learners more attentive to further input, leading to more opportunities for L2 acquisition. It may show that through the production during the interaction activities, both the provision of CF (recasts and metalinguistic information) and the performance of the different tasks may help learners to notice, detect and repair their linguistic problems.

CHAPTER 3: CORRECTIVE FEEDBACK IN CLASSROOM AND LABORATORY BASED RESEARCH

3.1 The Role of Corrective Feedback in Learning a Second Language

Introduction

A large body of classroom and laboratory based evidence is available suggesting that some focus on grammatical form, via CF, can be useful. A series of studies in Canadian immersion classrooms (e.g., Lyster, 1994; Spada & Lightbown, 1993) was motivated by concern about immersion learners' relatively low accuracy in production after years of exposure to meaning-focused input and subject-content study in the target language. This introduction will look at a number of relevant publications thematically in order to present a coherent picture of what is currently known about CF effects on learning. Another principal goal of this chapter, however, is to review different types of CF (with an emphasis on distinguishing implicit versus explicit feedback) and their roles in SLA as observed in a range of classroom studies. The chapter concentrates on recast and metalinguistic information CF, as these are the foci of the current study. Several form-focused instruction and CF studies have been tabulated in Appendix L.

Section 3.2 looks at the different types of CF and the measurement of their effectiveness for SLA. Section 3.3 examines previous research which investigates the effectiveness of explicit and implicit CF in both classroom and laboratory settings. Section 3.4 focuses on the nature of the control group in classroom studies to date: some using a test only group or some using the intervention tasks (-CF + tests). The chapter, therefore, identifies important gaps that remain in this large body of research and that are addressed in the current study.

3.2 Types of Corrective Feedback

The term feedback has been used to refer to both negative and positive feedback, error treatment, error correction and implicit and explicit feedback, occurring in both natural and instructional settings. It has been defined by Mackey (2007) as the reactive

information that learners receive in regard to the linguistic and/or communicative success or failure of their utterances.

In a comprehensive study of CF in Canadian immersion classrooms (Lyster and Ranta, 1997, pp. 46-49) distinguished six different types of CF:

1. ***Explicit correction*** refers to the explicit condition of the correct form. As the teacher provides the correct form, and clearly indicates that what the student said was incorrect (e.g., "Oh, you mean," "You should say").
2. ***Recasts*** involve the teacher's reformulation of all or part of a student utterance, minus the error (e.g., S: you must to ask him, T: you must ask him).
3. ***Clarification requests*** indicate to students either that the teacher has misunderstood their utterance or that the utterance is ill-formed in some way and that a reformulation is necessary (e.g., excuse me, pardon?).
4. ***Metalinguistic Feedback*** contains either comments, information, or questions related to the 'well-formedness' of the student's utterance, without explicitly providing the correct form (e.g., you need to have a simple form of the verb after modals).
5. ***Elicitation*** has three different techniques:
   i) Eliciting completion of their own utterance by strategically pausing to allow students to 'fill in the blank' (e.g., "No, not that. It's a . . .").
   ii) Using questions to elicit the correct forms (e.g., "How do we say *X* in French?").
   iii) Teachers occasionally ask students to reformulate their utterance.
6. ***Repetition*** refers to the teacher's repetition, in isolation, of the student's erroneous utterance. In most cases, teachers adjust their intonation so as to highlight the error (e.g., must to?).

These types have been categorised according to their degree of explicitness (overtness) to the learner. Any indication to the learners that their use of the target language is non-standard, draw learners' attention to language in two ways: implicitly and/or explicitly (Loewen, 2005). So, if a language learner says, '*He can paints the room,' the CF can be given explicitly by giving explicit or metalinguistic information, for example, 'you should use the infinitive form of a verb after modals.' Or, feedback can be given

implicitly, for example recasting is used such as 'ok, he can paint the room' please refer to Appendix H for more examples. Recasts, clarification requests, elicitation, and repetition are described as implicit feedback. Explicit correction and metalinguistic feedback fall within the category of explicit feedback. However, Erlam and Loewen (2010) have distinguished between implicit and explicit recasts depending on the number of moves and the intonation of the CF. The operationalization of recast may verify the degree of its implicitness or explicitness, for example in Erlam and Loewen (2010) the implicit recasts (interrogative) "consisted of correction of the error made, given with rising intonation whereas explicit recast (declarative) consists of two uninterrupted feedback moves. In the first, the student's error was repeated with rising intonation. In the second, a correction was provided in declarative form (p.886)".

CF can sit within a range of different approaches to grammar teaching or form-focussed instruction ("any planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic form" (Ellis, 2001, p. 1)). For example, it could be employed as part of 'focus on forms' (a focus on lexis, structures, or functions in the absence of a communicative context) or 'focus-on-form instruction' (attention to linguistic structures within meaning-focused, communicative activities (Ellis, 2001; Long, 1991, 1996)). The current study makes no claims about the general approach to grammar teaching adopted as the experimental interventions share characteristics of different approaches (e.g. the linguistic focus was the intentional organising principle for the sequence of tasks, but CF occurred within meaning-based tasks). In other words, the study seems to have been conducted in task-supported contexts.

Different types of feedback may impact on different aspect of language knowledge and use, as it will be discussed next.

Measuring the Effectiveness of CF

To gauge the effectiveness of implicit or explicit CF, learners' reactions to feedback have often been used as indicators, such as: intake (what learners notice in input becomes intake for learning), uptake (learners' immediate reaction to teacher's different types of CF) and repair (learners' modification for a problematic form in a target language

(Hall, 2007)). It is sometimes assumed that these reactions indicate that the learner has noticed the non-target form. However, a learner may not have noticed the error, but simply repeat the teacher's or a learner may not respond even though s/he does understand the error (Lochtman, 2002). Using learners' reactions is not, therefore, always a very good indication of whether the CF has resulted in learning.

Another way to measure the effectiveness of the CF is by comparing pre- and posttest scores. One problem with such tests is that different tests can tap into different types of knowledge. For instance Ellis (2007) notes, several studies investigating the effectiveness of implicit versus explicit feedback used the kinds of tests that favour the use of explicit knowledge (see section 2.6 for a discussion of implicit and explicit knowledge). Since implicit knowledge is said to underlie the ability to communicate fluently and confidently in an L2, CF effectiveness should be measured for impact on implicit knowledge (Ellis, 2005). Moreover, delayed posttest are needed to determine the extent to which CF is long lasting (Mackey & Goo 2007).

To this end, the current study used a battery of tests that are thought to tap different knowledge types (the terms 'explicit' and 'implicit' knowledge are adopted), and were used six or seven weeks after the intervention. The term implicit refers to "knowledge that learners are only intuitively aware of and that is easily accessible through automatic processing, whereas explicit knowledge consists of knowledge that learners are consciously aware of and that is typically only available through controlled processing" (Ellis et al 2006, p.340).

The following sections review previous studies that have operated different CF techniques in classroom based or laboratory environments. The review includes CF studies that have investigated recasts and metalinguistic knowledge (as these are the focus of the current study). (Note that the studies reviewed have not necessarily investigated both these CF techniques together).

Given that the effectiveness of CF can depend on the instructional setting, it may also be affected by the instructional or social context, such as ESL or EFL as suggested by Ellis and Sheen (2006), these variables will be addressed in Chapter 4.

Generally, learners believe that making errors is an integral and natural part of learning a foreign language and they can learn a lot from their mistakes (Bargiel-

Matusiewicz & Bargiel-Firlit, 2009). It is, then essential for SLA researchers to find out how learners perceive error correction and the types of CF provided in response to their erroneous utterances. These perceptions and beliefs will be dealt with in Chapter 4 as this is part of the study's underlying principles.

3.3 Studies on Corrective Feedback in Classroom and Laboratory Settings

A growing number of studies (Ellis & Sheen, 2006; Nicholas, Lightbown, & Spada, 2001; Spada, 1997; Spada & Lightbown, 2009) pointed out that the effectiveness of CF and subsequent learning outcomes are notably different in laboratory and classroom settings. The meta analysis of the effectiveness of CF in SLA by Li (2010) confirmed that empirical data from classroom and laboratory settings generate essentially different results. It has been suggested that "lab-based studies yielded a substantially larger effect than classroom-or group-based studies" (p.345). This might be because classroom is more distracting, CF in classroom is not always directed towards individual learners and thus might not be easily identified especially implicit type of CF (Nicholas et al., 2001), whereas laboratory contexts in which CF is delivered one–on-one making it possibly easier to perceive. Another reason for the advantage of laboratory over classroom context is that "variables can be more easily or better controlled and the quality of treatment might be better than in the classroom" (Li, 2010, p.345).

However, Gass, Mackey and Ross-Feldman (2005), comparing groups of learners engaged in learner-learner interaction, found no difference in the effectiveness of feedback given in classroom or laboratory settings. In contrast, Mackey and Goo (2007) found that lab-based studies, collected for the meta-analysis, showed a greater effect than classroom-based studies.

From a pedagogical perspective, Ellis et al 2006 argued that it is important to examine corrective feedback within the classroom context and that ecological validity can only be achieved through classroom-based research (p.365).

It is, however, difficult from the previous reviewed issues to argue that classroom observation studies give more insight into the effectiveness of CF than laboratory observation studies. To this end the current pedagogical study examined the corrective feedback within a classroom-based context that might ensure the ecological validity.

Previous studies on recasts

Recasts (a reformulation of learners' erroneous utterance into correct utterance) have been the most widely CF type investigated in L2 research due, in part, to the fact that they are the most frequently used form of feedback in both caretaker-child (Bohannon & Stanowicz, 1988; Farrar, 1992; Marcus, 1993) and teacher-student interactions (Ellis, Basturkmen, & Loewen, 2001; Lyster & Ranta, 1997). They also have been researched in both laboratory and classroom contexts in terms of (a) L2 development as gauged by pre-tests and/or posttests (e.g., Ammar, 2008; Ammar & Spada, 2006; Carroll & Swain, 1993; Doughty & Varela, 1998; Ellis, 2007; Ellis, Loewen & Erlam, 2006; Ishida, 2004; Leeman, 2003; Loewen & Nabei, 2007; Long, Inagaki, & Ortega, 1998; Lyster & Izquierdo, 2009; Mackey,1999; Mackey & Philp, 1998; McDonough & Mackey, 2006; Morris & Tarone, 2003;Nassaji, 2009; Sheen, 2007) and (b) immediate learner responses to the recasts, such as uptake and modified output (e.g., Ellis, Basturkmen & Loewen, 2001; Lyster, 1998; Lyster & Ranta, 1997; Nassaji, 2007; Oliver, 1995, 2000; Panova & Lyster, 2002; Pica, Holliday, Lewis & Morgenthaler, 1989; Sheen, 2004).

Given that recasts have recently been the topic of interactional work in the SLA indicating the usefulness of this type in developing L2 knowledge, studies that compared recasts to other types of CF yielded mixed findings with great demands for more research regarding the effectiveness of recasts in L2 learning. Some of these studies are discussed in more depth here as they are relevant to the current study. More information about other studies can be found in Appendix L.

Recasts vs. other types of CF

Although learners find recasts as "criticism and even mockery" (Morris & Tarone, 2003, p. 325) rather than a feedback, it apparently, tends to be the most frequently used corrective feedback technique in response to students' errors even though they have tended to result in little uptake. This superiority could be attributed to the fact that recast often does not interrupt the flow of communication and thus facilitate form meaning mapping (Erlam & Loewen, 2010). The subsequent sections shed light on some studies

that compared the relative effects of recasts to other interactional type of implicit and explicit CF, followed by review literature on studies that examined the relative association of interaction, learning opportunities and task complexity to SLA.

Recasts and negotiated interactional input

To find out whether learners who received intensive recasts in comprehensible interactional input would benefit more than those who received interactionally modified input, Mackey and Philp (1998) presented a short-term study exploring the effects of negotiated interaction on the production and development of question forms in English as a second language (ESL). The researchers administered a pretest, three posttests (immediate, one week after, and four weeks after), and using information gap tasks to elicit questions. The potential rationales behind the study were to investigate: (a) the effect of recasts on learners' short-term interlanguage (IL) development, and (b) the nature and content of learners' immediate responses to recasts. Thirty-five adult ESL learners, ranged in age from15 to 30, from two private English language schools in Sydney, Australia participated in the study. They were of mixed backgrounds from beginner and lower intermediate intensive English language classes. Participants involved in 15 to 25 minute sessions for one afternoon each day for 1 week, then three more sessions over 3 weeks. Participants were paired with a NS and performed three tasks. The results suggested that learners benefited more from interaction with recasts than interaction alone in production of targeted higher-level morphosyntactic forms. The researchers also found that recast was beneficial for short-term interlanguage development even when they are not immediately incorporated by learners. However, Mackey and Philp (1998) pointed out that it is "difficult to identify whether learners who repeated the recast were actually perceived the recast as feedback or simply another way of saying the same thing" (p.351). *Note,* it is important to point out that this study will be revisited in section 3.4 relative to the inclusion of a group who participated in the interaction activities alone but had no CF.

In a research that found a small example of success using recasts, Long, Inagaki and Ortega (1998) did two empirical studies; one involved 24 second-semester learners of Japanese distributed in five groups. The distribution of the groups was made taking into account the type of feedback provided (model or recast) and the target forms. The first

group received a first treatment with recasts in adjective ordering production and a second treatment with models and locatives as the target form. The second group had models and locatives in the first treatment and recasts and adjective ordering in the second treatment. In the third group, recasts in locatives were provided in the first treatment and models in adjective ordering in the second one. Finally, the last experimental group had models and adjective ordering in the first treatment and recasts and locatives in the second one. The analysis on the oral picture-description task posttests revealed advantages for the treatment groups over the control on the adjective ordering target form but not on the locative form.

The second study included 30 undergraduate third-semester Spanish speakers studying English involved in task-based interaction assigned to four treatment groups and a control group. The treatment groups alternated recasts and models with the two target forms (object topicalization and adverb placement) in the same way the Japanese experiment did. The results of pretest, posttest, and control group design of this study revealed that not only did the recast and model group outperform the control one, but also that the effects of recasts were more than preemptive positive input (models) in achieving short-term improvements for the adverb placement. The findings obtained in these two studies suggest that the type of form used as target may play an important role in the effects that recast have on L2 development. However, the results in these studies need to be taken with caution due to the low number of participants in each treatment group and to some methodological issues related to the design of the experiment itself such as the number of items in the tests and the election of the target forms.

The nature of the interaction and the role of the learner are important factors, together with the type of structures that may be affected through interaction. Mackey (1999) carried out a direct exploration of the relationship between conversational interaction and L2 development. Learners of English were engaged in communicative tasks, with word order in WH questions being the targeted structure and with opportunities for interaction between participants. Adult ESL learners (n=34) of varying L1 backgrounds were divided into four experimental groups and one control group according to interactionally modified input, feedback from negotiated meaning, and learners proficiency level. The treatment consisted of one session per day for 1 week, one

session 1 week later, and a final session 3 weeks later. Each treatment session lasted approximately 15–25 minutes and consisted of different types of information-gap tasks in which NS-NNS dyads were given three tasks to perform. The design of the study involved pretest-treatment, posttest1 (immediate), posttest2 (2 weeks later) and posttest 3(5 weeks later). In the test sessions, participants carried out "spot the difference" tasks, in which each participant had a similar picture with 10 differences. The pictures were hidden from the view of the partner. The NNS was required to find the differences between the two pictures by asking questions. Mackey found that interactionally modified input produced a positive effect on the development of question formation but this is only evident for learners who are developmentally ready**.** The results confirmed that conversational negotiated interaction did facilitate second language development and provided direct empirical support for the claims of the interaction hypothesis (Long, 1996).

Recasts vs. explicit CF

Previous research points to the different effects of recast on children and adults, and that recasts may be more beneficial for children than adults (e.g., Mackey & Philp, (1998); Mackey and Oliver (2002)), Carroll and Swain (1993) found that adult language learners require more explicit explanation because of their previous learning preferences.

In their study, Carroll and Swain (1993) carried out an investigation comparing recasts, or as they put it "reformulated correct responses to mistakes" with other types of explicit negative feedback (given below) on adult Spanish speakers (n=100) learning English as a second language at the low intermediate ESL classes in the Toronto area. The time between initial and final testing was only 1 week. The subjects were divided into five groups according to the type of feedback they received when they made an error in the dative verbs; group (A) received direct metalinguistic feedback, Group (B) were told that their response was wrong, group (C) were given recasts whenever they made a mistake, group (D) were asked if they were sure that their response was correct when they made a mistake (indirect metalinguistic feedback), and group (E), a comparison group, received no feedback. The experimental sessions consisted of four parts: a feedback session and a guessing session followed by a second feedback and a second

guessing session. The results found that in initial feedback sessions, groups A (explicit metalinguistic explanations) and D (explicit correction) performed significantly better than the comparison groups. It was found that, on short-term recall, Group A performed significantly better than all groups except Group C (recasts). This is possibly due to the fact that adult learners of language require more explicit explanation. In short, Carroll and Swain (1993) suggested that explicit instruction combined with explicit metalinguistic feedback may be beneficial for students to understand complicated rules and that adult learners use feedback to learn specific and abstract linguistic generalizations and to correctly narrow the application of those rules.

Recasts vs. prompts

Given that prompts might help in eliciting the target structure more than recast, the effectiveness of these two techniques was investigated in different publications (e.g., Lyster, 2004; Ammar & Spada, 2006; Ammar, 2008; Lyster & Izquierdo, 2009; Dilans, 2010).

The effectiveness of recasts over prompts was examined by Lyster (2004) on 10–11 years old fifth-grade students (n=148) participating in a quasi-experimental classroom study. He investigated the effects of four types of form-focused instruction (FFI) and corrective feedback- 1) FFI + recasts; 2) FFI + prompts; 3) FFI only; and 4) no feedback on the acquisition of French grammatical gender in immersion classrooms in Quebec, Canada. Pretests, immediate posttests, and delayed post tests were administered to the experimental group as well as to a comparison group to compare classes on three measures: a cloze test, a composition task and an oral interview. Lyster assessed the effectiveness of recasts over prompts (including elicitations, metalinguistic cues, clarification requests, and repetitions where the teacher adjusts intonation to highlight the error) following a 5-week period of classroom-based instruction. Results showed a significant increase in the ability of students exposed to FFI to assign grammatical gender. In terms of feedback type, recasts were shown to be less effective than prompts in leading to improvements, especially on the written production tasks. In other words, the FFI group with prompts was superior to the control in all measures in both posttests, whereas the FFI with recasts group outperformed the control group only on the posttest

written measure and the delayed-posttest oral measures. Lyster related this finding to the ambiguity of recasts and potential difficulties involved in noticing morphosyntactic errors, and he concluded that recasts might not be the most effective type of feedback to use in communicatively oriented classrooms in comparison to other feedback types such as prompts. He concluded that prompts are superior to recasts and further argued that prompts work better for acquisition by pushing learners to self-repair.

A similar result to Lyster (2004) was observed regarding the superiority of prompts over recast on L2 development. In a quasi-experimental study Ammar (2008) examined the differential impact of recasts in comparison to prompts and no corrective feedback on francophone learners' acquisition of English third person possessive determiners. Sixty-four students from three intact intensive English as a second language classes carried out communicative activities during which they received corrective feedback according to the condition they were assigned to. The instructional intervention was spread over a four-week period and consisted of two main parts: a three-phase instruction session that lasted 45 minutes and 11 communicative activities, each of which lasted 30 to 45 minutes. A pre-test-treatment–immediate posttest –delayed posttest design was used to identify the effects of prompts and recasts. An oral picture-description task and a computerized fill-in-the-blank test were administered prior to the treatment and immediately after it ended. Four weeks later the oral picture description task was re administered. Analyses of individual participants' oral data revealed that prompts and recasts are more effective than no feedback and that prompts may be more effective than recasts in leading to L2 morphosyntactic development especially for low-proficiency learners. Data from the computerized task showed that prompts allowed learners to repossess possessive determiner knowledge faster than recasts.

Different from the previous two studies, a more recent study by Lyster and Izquierdo (2009) indicating no significant different effects between the recast and the prompt groups. The researchers investigated the differential effect of prompts and recasts on the acquisition of grammatical gender by adult L2 learners (n=25) of French. Learners were exposed to a three hour form-focused instructional treatment on the target, and each learner participated in three different oral tasks in dyadic interaction with native or near native speaker of French who provided either prompts or recast in response to their

errors. Pretests and immediate and delayed posttests included two oral production tasks and a computerized reaction-time binary-choice test. The results showed that both recast and prompts groups significantly improved accuracy and reaction-time scores over time, regardless of the type of feedback as these two types of CF provided different opportunities for second language learning. This result could be attributed to the fact that learners receiving recasts benefited from the repeated exposure to positive models as well as from opportunities to infer negative evidence, whereas learners receiving prompts benefited from the repeated exposure to negative evidence as well as from opportunities to produce modified output.

Given that the target structure might play a role in the efficacy of the CF type, in a more recent work, Dilans, (2010) investigated the effects of oral corrective feedback (CF) in the form of prompts and recasts on second language (L2) vocabulary development. A population of intermediate adult learners of ESL in a community college located in the US Southwest was used. The participants (n=23) were designated to three groups: prompts, recasts, and control. The treatment consisted of a four-step vocabulary activity during which prompts, recasts or no feedback was provided, respectively. The quasi-experimental study employed pretest-posttest-delayed posttest design plus a background questionnaire. The treatment outcomes were tested in terms of measures based on a three-dimensional second language vocabulary development model. The findings indicated that prompts and recasts were equally beneficial in short-term and that prompts were slightly more advantageous in a longer term than recasts in facilitating L2 vocabulary development. However, from both CF groups, only the prompts group demonstrated significant increases on all three dimensions of L2 vocabulary development as they were operationalized for the purpose of the study. An explanation for the stronger effect of prompts over recast could be possibly because of their (pushed) output-generating orientation. In addition, opportunities for learners to produce language related to the meanings of the targets appear to be conducive to learning, more than attending to teacher input. The result of Dilans' is different to Li's (2010) meta-analysis suggesting superiority for the explicit CF over the implicit type of CF in short term. The long term effects for the prompts group over recast is also different to Li's (2010) whose meta-

analysis implied that the effect of implicit type of CF did not fade or increase over a long term.

Recasts and task complexity

Robinson's cognition hypothesis (2001) makes the claim that task complexity is a strong variable affecting the occurrence of interaction and learning opportunities. Some complex tasks may prompt learners to look for more and more help during meaning-based communicative tasks. Given that task features have essentially strong effects on learners' L2 production and the processes of language acquisition in both research and classroom contexts, the construct of task complexity has been an important consideration in designing tasks. Task complexity, however, could be treated as a moderator variable between interactional feedback (e.g., recast) and learner uptake. This rationale was examined by Révész (2009) who investigated the relative effectiveness of recasts and task complexity which, in turn, can affect L2 learning outcomes. In this regard, Révész (2009) emphasised on the relationship between task variable +/- contextual support combined with recasts and how this relation affects L2 morphosyntactic development (the past progressive) on (n=90) adult learners of EFL. There were four comparison groups and one control group (participated in the testing sessions only). The implementation of tasks for the comparison groups differed as to (1) whether learners received recasts while describing photos, and (2) whether they could view the photos while describing them. The pre-, post-and delayed posttests included grammaticality judgment, fill –in- the gap, written production, and two oral production tasks. The results yielded two main findings. First, learners who received recasts in the absence of photo out-performed learners received recasts in the presence of photo, on the three testing tasks in both posttests. Second, the group that viewed photos but did not receive recasts yielded some advantage over the group who neither viewed photos nor received recasts. When the recast group was compared to the non-recast group in the presence and absence of photo support, the +/- photo recast group exhibited a considerable gain from the pretest to the posttest while the +/- photo non recast groups showed a small increase. On the delayed posttest, the +photo recast group and the −photo non recast group maintained their respective gains. In contrast, the −photo recast group and the +photo non recast

group showed respectively a slight decrease and increase in their use of the past progressive. Finally, the results indicated that the −photo recast group achieved the highest gain, followed by the +photo recast group, the +photo non recast group, and the −photo non recast group. A revisit for this study will take place in section 3.4 in accordance to the inclusion of a group who involved in the interaction activities as well as the experimental groups. This study will also be revisited in Chapter 4 regarding learners' perceptions on the present or absent of contextual support.

Kim (2009) examined the effects of task complexity and task condition on the occurrence of interaction-driven learning opportunities and L2 development in EFL task-based language classrooms on the acquisition of two morphosyntactic structures in English: questions and past tense. The study employed a pretest-posttest design on Korean university students from four intact English classes who were randomly assigned to three experimental groups (i.e., simple, +complex, ++complex) and one comparison group (n=191 for questions and n=l86 for past tense). Kim suggested that in order to meet communication demands while completing complex tasks, learners may notice their interlanguage gaps, and solve linguistic problems by negotiation and interactional feedback. The results of the study indicated that leaner-learner interaction in task-based language teaching was found beneficial in L2 development. It was also suggested that more complex tasks facilitate more learning opportunities and develop language learning. The findings of the study suggested that task complexity is an important variable which can impact the amount of meaningful interaction as well as learners' attention to linguistic forms in task-based language teaching contexts. The findings indicated that carefully designed pedagogic tasks can influence the amount and quantity of learning opportunities for developmentally advanced forms that occur in L2 classrooms, both in terms of interaction and production practice. Teachers should not take a dominant role in task-based language teaching classrooms, but rather assist learners' interlanguage development by providing appropriate feedback.

Previous research on recasts and metalinguistic information

Some research suggests that more explicit instructional techniques are more beneficial than more implicit techniques. For example, Norris and Ortega (2000)'s meta-

analysis found a larger effect size for explicit instruction *(d=0.96*) than implicit instruction (*d = 0.81).* Specifically, Norris and Ortega (2000) meta-analysed the effect sizes of some subgroups of studies examining the efficacy of recasts and metalinguistic feedback. The results suggested a larger effect size for the metalinguistic feedback than for recasts. Similar result was found in the most recent meta-analysis by Li (2010), which "revealed that explicit feedback worked better than implicit feedback over a short-term and that the effects of implicit feedback did not fade or even increased over a long term"(p.348).

In contrary, Mackey and Goo (2007) meta-analyzed the effect of different feedback types on immediate posttests and have found that the mean effect size was 0.96 for recasts (this is partly due to the fact that interactionist theories, such as Long's Interaction Hypothesis, claim that learners have to be primarily focused on meaning for CF to facilitate development of linguistic competence), 0.47 for metalinguistic feedback, and 0.52 for negotiation. Although Li (2010) pointed out that recasts have tended to receive considerable attention in SLA research, whilst explicit feedback such as explicit correction and metalinguistic information feedback have had much less interest, Mackey and Goo (2007) stated that it was premature to make conclusive arguments for the superior effects of recasts based on their meta-analyzed data. One issue that has been quite widely investigated in relation to CF techniques is whether more explicit means of correction (such as metalinguistic information) compares to more implicit means of correction (such as recasts). Given that there is a dearth of research on explicit type of CF, metalinguistic information in particular, and that findings appear to be mixed in terms of whether it is more or equally effective as recasts, the current study aims to improve our understanding of the effectiveness of this kind of CF.

Evidence that metalinguistic information is more effective than recasts

Given that explicit CF is to direct learners' attention and to provide explicit metalinguistic explanations and rule-based types of instruction such as consciousness-raising tasks, and that implicit CF aims to attract learners' attention to form and/or to the induction of the rules underlying a target form without telling the learner overtly what the target form is or how it is formulated. Several studies (for example, Ellis et al, 2006;

Ellis, 2007, Sheen, 2006, 2007) have compared the provision of implicit type of CF with more explicit types of feedback and have shown that learners learn better from the explicit feedback as they pushed learners to modify their non-target-like output without providing L2 models (Egi, 2010) and identify the nature and locus of the errors (Ammar & Spada, 2006).

A study by Ellis et al. (2006) investigated the relative effects of implicit and explicit CF on the acquisition of past tense by low intermediate ESL learners in New Zealand. The study involved three classes of students (*n* =34) with the mean age of 25 years. Implicit feedback was provided in the form of recasts and explicit feedback in the form of metalinguistic information feedback. To measure learners' performance two types of testing instrument were administered prior to the instruction, 1 day after the instruction, and again 2 weeks later and were used in the form of: (1) an oral imitation test was designed to measure learners' implicit knowledge; and (2) an un-timed grammaticality judgment test and a metalinguistic knowledge test were designed to measure learners' explicit knowledge. The treatment involved two different half-hour communicative 'focused tasks'. The tasks included a gap that required learners to focus primarily on meaning and to make use of their own linguistic resources. The results indicated that learners in the metalinguistic information group outperformed those in the recasts group, "most likely due to recognition of the overtly corrective nature of metalinguistic feedback" (Gass & Selinker, 2008, p.337). The findings indicated that metalinguistic explanation benefited implicit as well as explicit knowledge and pointed to the importance of including measures of both types of knowledge in experimental studies.

Similarly, Sheen (2006, 2007) examined the effects of CF, language aptitude, and learners' attitudes on the acquisition of English articles. The study used adult learners of various first language (L1) backgrounds in intermediate ESL classes. A design of pre-test, immediate- posttest and delayed- posttest was used to measure learners' outcomes. Tests were in the form of a speed dictation test, a writing test, and an error correction test. In addition, an aptitude test was administered prior to the pre-test session. There were two treatment sessions. Each session involved a narrative stimulus, which consisted of seven indefinite and definite articles, to elicit the target structure of the study and lasted 30-40

minutes. Learners were then asked to retell each narrative to the class. Whenever a student made an error in article usage; the teacher corrected the error using either a recast (in the recast group) or metalinguistic feedback (in the metalinguistic information group). Recasts, in Sheen, were provided in the context of a communicative activity, as the teacher's reformulation of a student's erroneous utterance but preserving the meaning of the original, whereas metalinguistic corrections were provided by giving the correct form, plus metalinguistic information. A control group who completed the tests only was used as a comparison group. The findings showed that whereas implicit oral recasts involving article errors did not facilitate learning, the metalinguistic feedback with provision of the correct form can be effective in helping learners to improve the grammatical accuracy of English articles. This study will be revisited in section 4.1 vis-à-vis learners' attitudes towards the type of CF.

Evidence that there is no difference between recasts and metalinguistic information

In light of the mixed findings concerning the efficacy of the different types of CF, and based on the above positive indication of metalinguistic information CF over recasts. Closer examination of studies that generate no different effects between the types of CF will be discussed in this section. In a quasi-experimental study, Loewen and Erlam (2006) investigated the effectiveness of two types of corrective feedback, i.e., recasts (implicit feedback) vs. meta-linguistic information (explicit feedback) during small group text-chat interaction with 31 elementary L2 learners of English on regular past tense. A third group served as a control. After taking pre-test, the participants went through 56 minutes of corrective feedback session, where they received either type of corrective feedback while completing the two tasks; story retelling after seeing a picture with written narratives and verbal description of the pictures. Their learning on English regular past tense was measured by timed and untimed GJT tests (see Loewen and Erlam, 2006 for more details on these measures) in three different times. The pre-test occurred two days before the treatment, posttest 1 occurred immediately after the treatment and posttest 2 took place two weeks later. Analysis displayed no statistically significant gains in response to either type of feedback nor was there significant difference in gains over time. Analysis suggested that these findings may have been influenced by the learners' low proficiency with the target form. The researchers further argued that a possible

reason behind this result "may be due to the struggle that the instructor had at times in keeping the students on task, only one group appeared to be on task throughout the session" (p.10).

Similar role of explicit and implicit CF in L2 acquisition was found in a study by Loewen and Nabei (2007). Two intact English classes at a Japanese university provided 66 participants for the study. English question formation was chosen as the target for the feedback. A total of ten groups were chosen to practice in meaning focused activities and each group was randomly assigned one treatment option. Three groups received recasts, two received clarification requests, two received metalinguistic feedback, and three received no feedback. The treatment duration was 30 minutes and the two treatment tasks were spot -the-difference and a guess-the-storyline task. Administration of pre-and posttests included three different instruments: a timed grammaticality judgment test, an untimed grammaticality judgment test, and an oral production task. It was found that although all three feedback groups (clarification requests, recasts and metalinguistic feedback) improved more than their non-feedback comparison groups, no feedback group significantly outperformed the others. In their discussion of this lack of comparative advantage, Loewen and Nabei suggested that the brevity of the treatment (30 minutes) may have limited the ability to elicit sufficient differential effects. Furthermore, institutional constraints, which prevented the administration of a delayed posttest, also meant that there was no opportunity to observe comparative advantages that may have emerged over time.

In a computer-based study, Sauro (2009) investigated the impact of recasts and metalinguistic information on the development of L2 knowledge on adult learners of English (n=23) from a first year undergraduate English grammar and translation course at Malmö University College in Malmö, Sweden. The participants were divided into three groups: a recast, a metalinguistic feedback and a control group. In addition, 9 native English-speaking interlocutors were recruited from graduate programs at the University Of Pennsylvania Graduate School Of Education to interact with and provide corrective feedback to the Swedish participants via synchronous written CMC. During task-based interaction via text-chat, the learners received focused corrective feedback on omission of the zero articles with abstract non-count nouns (e.g., employment, global warming, and

culture). Two computer-mediated collaborative writing activities completed by participants when paired with their native English-speaking interlocutors, as well as computer-delivered acceptability judgment pre-, post- and delayed posttests used to measure learning of target form knowledge. Results showed neither type was significantly more effective than the other in either the immediate term or over time in target form knowledge, although the metalinguistic group showed significant immediate gains relative to the control condition.

3.4 The Nature of the Control (or Comparison Group) in Previous Studies

The nature of the control groups has varied in CF studies to date. A few studies have included a "testing control" group in which learners took only the pre-tests and posttests without being involved in any instructional treatment and had no opportunity to practice the target structure and, thus, received no feedback (e.g., Ellis et al., 2006, p.350; Sheen, 2006). In some studies, learners in the "interaction control group" (e.g., Mackey & Oliver, 2002, p.468) or "comparison" (e.g., Erlam & Loewen, 2010, p.883) carried out the same task-based activities as the experimental groups with no CF. Other studies have used a control group which had regular class instruction without feedback (e.g., Ammar & Spada, 2006) or received feedback on structures other than the target structure (e.g., vocabulary errors in Yang & Lyster, 2010).

It has been noted that (e.g., Ellis et al, 2006; Sheen, 2006), studies in which there was no control group that did just the experimental task (but without feedback) cannot isolate the effects of the CF from any effects of the experimental task itself. The current study therefore aimed to address this problem by including a group who took the tests and engaged in the interaction tasks but had no feedback. Only few published studies have been located that have used such a control group that engaged in the same activities as the experimental groups but without feedback (i.e. Mackey & Philp (1998); Mackey & Oliver (2002); Révész (2009); Erlam & Loewen (2010); and Yang & Lyster (2010)). These are reviewed in some detail here as they are highly relevant to the current study.

In a study carried out on 34 adults ESL learners in a private school in Australia by Mackey & Philp (1998) to explore the relationship between conversational interaction and L2 development, it was found that interaction +recast were more beneficial than

interaction alone in facilitating the production of the target forms (see section 3.3 for more information on this study).

Mackey and Oliver (2002) conducted a study on interactional feedback and children's L2 development. In the study, the authors examined the effects of interactional feedback on children's development using a pre-test and three posttests design. Twenty-two learners carried out communicative tasks that provided contexts for targeted forms and interactional feedback. Each test and treatment session (one session per day, for three days) lasted approximately thirty minutes. The study involved two groups who received similar amounts and types of input and opportunities for output, but differed in terms of the feedback provided: 1) the interaction and feedback group, which received interactional feedback (including negotiation and recasts) in response to their non-target-like production of question forms; and 2) an interaction control group who interacted but received no feedback. In the interaction and feedback group, children asked whatever questions were necessary to carry out the tasks and the native speaker answered their questions and asked her own when necessary. In the interaction control group, learners carried out the same task-based activities as the experimental group. The input for this group was fully modified to reduce problems and allow learners opportunities to hear and produce questions. The interaction control group received "very little feedback on English question formation" (p.471).

The result in Mackey and Oliver (2002) is in line with the result found in Mackey and Philp (1998) confirming the usefulness of recast in an interactional activities rather than the interactional activities alone to L2 development regardless of age. It appeared, however, that child learners in the interaction and feedback group showed more constant development than the interaction control group in terms of question formation. Mackey and Oliver have noted that the increase in production of questions at higher levels was observed for the interaction and feedback group in the immediate as well as delayed posttests. Contrary to this result, Mackey (1999) and McDonough (2001) found that the effects of interaction for adults did not appear immediately. Mackey and Oliver also found that "it may be possible that interactional feedback leads to development more quickly for child learners than for adults suggesting that children in their study "seemed to be able to utilize feedback for interlanguage restructuring more quickly than adults

exposed to similar feedback in the studies by Mackey (1999) and McDonough (2001)"(p.473).

Along similar lines, Révész (2009) examined the relative effectiveness of recasts by comparing recast groups with and without contextual support to non-recast groups with and without contextual support (see section 3.3 for more details). She found that learners received recast +/- contextual support showed greater development in their ability to use the target feature than learners who had received no recast +/- contextual supports.

Different results concerning the relative effectiveness between the provision of implicit and explicit type of CF and successful learning were found in Erlam and Loewen (2010). The researchers provided some evidence that a comparison group which carried out the same task-based activities as the experimental groups with no CF may be as effective as tasks with CF. In a quasi-experimental laboratory–based study, learning of French noun–adjective agreement was investigated. Fifty participants completed a pre-test, two treatment sessions, an immediate posttest, and a delayed posttest. The testing battery involved three instruments: an oral imitation test, an untimed grammaticality judgment test (GJT), and a spontaneous production test. Students completed a written background questionnaire at the pre-test and participated in mini oral interviews after the posttest and the delayed posttest about the type of feedback (results on questionnaire and interviews will be presented in Chapter 4). Two types of feedback were provided in the form of implicit feedback, as a single recast with interrogative intonation, and explicit feedback in which the error was repeated with rising intonation, and then a recast was provided in declarative form. The two experimental groups, involved in communicative tasks designed to elicit the target structures with CF and a comparison group performed the same tasks but received no CF. A series of four tasks were used in the interaction sessions to create obligatory occasions for noun–adjective agreement. All interaction sessions were audio-recorded. Results of the study indicated that "the type of feedback students received did not have a differential impact on learning. In addition, the presence or absence of feedback did not appear to influence test performance. The fact that the comparison group made gains along with the treatment groups suggests that participating in the testing sessions and/or in the interactive activities was beneficial"(p.895).

Another study, which gives support to the provision of CF over the absence of CF in interactional tasks, was conducted by Yang and Lyster (2010) in EFL classrooms at a university level in northern China. Participants included 72 undergraduate EFL students were divided into three groups, two CF groups, where teachers consistently provided one type of feedback (i.e., either recasts or prompts), and a control group (performed the same communicative classroom activities but without receiving CF on past tense errors). The treatment sessions consisted of four form–focused production activities for approximately two hours over a period of two weeks. Oral and written outcome measures were administered at pre-, post- and delayed posttests (two weeks interval). In the oral production test, learners were required to retell a story based on a series of word cues (adapted from an online grammar exercise), whereas in the written production test, learners were required to compose a story in fifteen to twenty minutes. The results indicated that "the effects of prompts were larger than those of recasts for increasing accuracy in the use of regular past tense forms, whereas prompts and recasts had similar effects on improving accuracy in the use of irregular past tense forms"(p.236). The advantages for prompts over recasts and no feedback suggests that "the differential effects of CF treatments were further mediated by (a) the extent of immediate self-repair, (b) varying degrees of CF saliency, and (c) the nature of the structural targets"(p.255).

The evidence about the relative effectiveness of oral interaction alone versus oral interaction with CF is, therefore, mixed, and the current study sought to increase our understanding in this area. In sum, in line with three recent meta-analysis (Russell & Spada, 2006; Mackey & Goo, 2007; Li, 2010) that call for more research concerning the effectiveness of CF, the current study explores the effectiveness of recast and metalinguistic information in the acquisition of English modals in FFI settings, and compares these with oral production tasks alone.

CHAPTER 4: CORRECTIVE FEEDBACK, CONTEXT, AND INDIVIDUAL DIFFERENCES

Introduction

To fully understand the role of CF in SLA, we need to determine whether individual differences such as apprehension and learners' attitudes influence the effects of different kinds of CF. One of the rationales of the current study was to explore the impact of learners' attitudes towards CF and their beliefs about the effectiveness of CF on the actual effectiveness of the different types of CF as measured by language accuracy (e.g., Schulz, 2001; Havranek & Cesnik, 2001; Sheen, 2006; Amador, 2008; and Loewen et al, 2009).

Learners' attitudes towards error corrections (amongst other factors) may affect learning outcomes, and their attitudes could be influenced by learners' cultural and educational background. It has been suggested by Gass and Selinker (2008) that "in any learning situation, not all humans are equally motivated to learn languages, nor are they equally motivated to learn a specific language" (p.165). Thus, teachers should be sensitive to students' attitudes to language, particularly to error correction although it might be argued that learners' preference may not be what is actually best for acquisition (Truscott, 1996). Previous research has reported the influence of cultural background on learning outcomes and learning styles (e.g., Dunn & Griggs, 1995; Reid, 1987; Bedell & Oxford, 1996; Woodrow & Sham, 2001) and a long tradition of research (following Gardner, 1985) has shown that two social psychological variables-attitude and motivation - play a key role in second language learning.

The current study focuses on the effects of operating two types of CF in relation to L2 learning outcomes, and on any influence of learning contexts and learners' attitudes to CF on this relationship. The subsequent section reviews previous work in this area. (It is acknowledged that teachers' attitudes to CF may have an impact on learning, but as this is not a focus of the current study, this will not be discussed in any details here).

4.1 Learners' Attitudes toward and Perceptions of Corrective Feedback

A few studies have found discrepancies between teachers' and students' attitudes to CF. For example, Schulz's (1996) study revealed that students' attitudes toward grammar instruction and error correction were more favourable than their teachers' attitudes; that is, learners want more error correction as (90%) of them had a positive attitude towards error correction. In the same vein, Ancker's (2000) surveyed teachers' and students' perception in 15 countries. The survey asked whether teachers should correct every error students make when using English and results showed a 25% positive response for teachers and a 76% positive response for the students. Teachers were concerned about the negative impact of correction on students' confidence and motivation, whereas the students wanted correction as they wanted to speak English correctly.

Given that CF could be provided implicitly, explicitly or together, it is of interest to find out whether learners have different attitudes to and perceptions of different types of corrective feedback. Another, larger, body of research has pursued "how learners perceive feedback and whether their perceptions affect their subsequent L2 development"(Mackey et al.2000, p.471). The next sections review some of SLA studies investigating learners' attitudes towards either explicit or implicit CF.

Studies finding learners' preference for explicit CF

Sheen (2006) designed a questionnaire, using a Likert scale (1-6), to measure language anxiety, attitudes towards error correction and grammatical accuracy and whether learners perceive teacher's correction as helpful and important. The results showed that positive attitudes towards error correction and grammatical accuracy were stronger in the explicit CF group than in the implicit CF group. Sheen found that learners with positive attitudes towards error correction benefitted more from metalinguistic feedback than recasts. She also argued that attitudes towards error correction and grammatical accuracy cannot be expected to have any mediating effect if learners are not aware they are being corrected.

A preference for explicit CF was also revealed in a study by Amador (2008) who investigated the preference of twenty-three college students of English on twenty error correction techniques. They were presented in dialogue form and took place in a classroom. The results were in line with Sheen's (2006) study, indicating a preference for explicit corrective feedback techniques. As this study was also measuring learners' preference for the sources of the CF, it will be revisited in the subsequent section.

Studies finding learners' preference for implicit CF

To explore how learners' perceptions about recast provided during task-based dyadic interaction might depend on the focus of the feedback, Mackey et al., (2000) examined this issue by collecting data of stimulated recall protocols with two groups of L2 learners. Data were collected from ten learners of English as a second language and seven learners of Italian as a foreign language. It was found that learners were most accurate in their perceptions about lexical and phonological feedback, and much less accurate in terms of their perceptions about morphosyntactic feedback. The authors suggested that this might have been because morphosyntax often does not interfere with understanding in the same manner as incorrect pronunciation or inaccurate lexical items. Based on these results, Mackey et al. argued that both the nature and the target of the feedback might affect the accuracy of learners'perceptions.

In a report indicating the great correlation between learners' characteristics such as proficiency level, verbal intelligence, and attitude towards correction, and the success of CF, Havranek and Cesnik (2001) conducted a comprehensive developmental study with two-hundred-and-seven native German speakers studying EFL. They compared the effects of recasts, repetition + recasts, and elicitation via measuring the success of error correction on learners' performance in a subsequent test. The study reported that corrective feedback was likely to benefit learners who had a positive attitude towards error correction and high language proficiency.

Learners' perceptions of recasts were also investigated by Philp (2003), who examined the extent to which nonnative speakers (NNSs) notice native speakers' NSs' reformulations of their interlanguage grammar through recasts in dyadic interaction. The study involved thirty three 33 adult ESL learners participating in oral communication

tasks in NS-NNS dyads and received recasts of their non-target like question forms. It was found that learners did not always notice recasts, and if they noticed them, they often did not notice every detail. The authors suggested that a variety of learners' variables may account for how students perceive recasts, if they are noticed at all, such as limitations in working memory, unfamiliar input, multiple corrections, complex changes in the recast, learners' level, processing biases of the learner, and grammatical forms in the recast that were beyond the learners' interlanguage grammar.

A recent investigation was carried out by Egi (2010) to examine the relationship between learners' perceptions of recasts and their responses to the recasts. Twenty four foreign language learners of Japanese engaged in task-based interactions during which they received recasts of their errors. Each learner then watched video clips of the recast episodes and commented on them. Analysis was taken in relation to learners' responses to the recasts: uptake, repair, and modified output. In recast episodes where they produced uptake, their reports indicated that they perceived the recasts as corrective feedback significantly more frequently compared to cases where they did not produce uptake.

Learners' attitude towards the source (or supplier) of feedback

The current study investigates teacher's correction, rather than peers' correction. This decision was taken partly because previous research suggests that learners perceive teachers' CF more positively, whereas peers' correction can be influenced by negative or positive peer relationships, the corrections may not be trusted and some learners may perceive recasts as criticism, which prevents them from using it as helpful information (Amador, 2008; Morris & Tarrone, 2003).

Learners' attitude towards the tasks in which CF is embedded

A body of research into task-based teaching and learning has demonstrated that tasks can elicit interactional features, negotiation of meaning, and attention to form, (e.g., Bygate, Skehan, & Swain, 2001; Van den Branden, 2006).

Given that CF is embedded in classroom interactional tasks, I provide a brief review of some of the research that has looked at learners' attitudes to grammar instruction, CF and tasks more generally. The review is provided as it informed the

design of the current study. Loewen et al (2009) investigated the beliefs of L2 learners regarding the role of grammar instruction and error correction. Seven-hundred-and-fifty-four L2 students at an American university completed a questionnaire consisting of thirty-seven Likert-scale items and four open ended prompts. Six underlying factors were used to investigate differences in beliefs among learners studying one of fourteen TLs: efficacy of grammar, negative attitude toward error correction, priority of communication, importance of grammar, importance of grammatical accuracy, and negative attitude toward grammar instruction. Differences were found between the ESL and foreign language learners in the area of attitudes to practice and speaking. Practice or speaking in grammar instruction was not favoured by the ESL learners, whereas it was by foreign language learners, particularly with the Less Commonly Taught Languages, Arabic and Japanese. ESL learners were less convinced about the need for grammar instruction and error correction and were keener to improve communicative skills than were foreign language learners. It was also found that learners of Chinese and Arabic were more positive about grammar instruction and error correction than were learners of other languages. "This difference might be attributable to the fact that these two languages are non-Indo-European languages and are perceived to be more challenging than languages such as German or Spanish" (Loewen et al,2009 , p. 102).

The availability or lack of contextual support might make a difference in terms of task difficulty and attitudes to task. Révész (2009) administered an exit questionnaire to half of the participants after completing the posttest to obtain information concerning the participants' perspectives on the test task. In response to the question about whether the presence or absence of photos made the description task easier, the large majority of participants (twenty nine of thirty three) reported that they found it less difficult to describe the photos when they were able to view them. The rest of the participants (12.1%) felt that the availability or lack of contextual support did not make a difference in terms of task difficulty. The majority of the participants felt that describing the photos without contextual support was more difficult, because it forced them to simultaneously focus on speech production and memorization, which, in turn, made it more challenging to concentrate on task completion. (In relation to the effectiveness of recast and task complexity, this study has been discussed earlier in section 3.3).

4.2 The Role of Contexts and Culture and Corrective Feedback

Only a little previous research can inform the question of whether the use and effectiveness of CF varies between different contexts. For example, Liu (2007) surveyed eight hundred teachers of English from forty-two countries and found that EFL teachers tended to focus more on linguistic forms than ESL teachers and learners' responses to it. Many studies have been conducted in foreign language contexts but few in second language contexts. Mackey and Goo (2007) indicated that 71% (n = 21) of the studies selected for their meta-analysis were carried out in foreign language (FL) contexts, while 29% (n = 8) were implemented in second language (SL) contexts. The researchers suggested that studies conducted in FL contexts appeared to produce stronger evidence for the effects of interaction than research conducted in L2 contexts. This difference, according to the researchers, was statistically significant for the immediate posttests, and a large mean effect size for L2 contexts was observed only on the short-term delayed posttests. They also found that learning as a result of CF in foreign language contexts, may be more effective with lexical than with grammar learning.

Similar results have been found in a meta-analysis of thirty-three primary studies including twenty-two published studies and eleven Ph.D. dissertations. Li (2010) found that studies conducted in foreign language contexts produced larger effect sizes than those in second language contexts. He defined a foreign language setting as one "where the learner studies a language that is not the primary language of the linguistic community (e.g., an L1 Korean speaker learning English in Korea); while a second language setting is one in which the learner's target language is the primary language of the linguistic community (e.g., an L1 Korean speaker learning English in the United States)" (p.315). He suggested that different CF types may have different effects as the dynamics, expectations and aims within these two settings are different.

Sheen (2004) looked at the occurrence of recasts, learners 'uptake and repair across four instructional contexts: French Immersion with children in Canada (Lyster & Ranta, 1997); ESL with adults in Canada (Panova & Lyster, 2002); Intensive ESL with young adults in New Zealand (Ellis, Basturkmen & Loewen, 2001); and EFL with older

adults in Korea (new data). She found that the instructors in the Korean EFL context provided many more recasts than the instructors in the ESL and immersion programs. Her findings also suggested that the extent to which recasts lead to learner uptake may be greater in contexts where the focus of the recasts is more salient, and where students are oriented to attending to linguistic form rather than meaning.

To compare learners' preference for particular types of correction between two distinct cultural groups in different learning contexts, Lennane (2007) examined whether cultural background had an effect on Taiwanese and Quebecois language learners' preferences for particular CF, as well as on certain beliefs and attitudes surrounding the use of error correction in the language classroom. The study involved one-hundred-and-thirty-seven Taiwanese EFL students, ninety-seven ESL Quebecois students, twelve Taiwanese English instructors and twelve native English teachers in Quebec. All participants completed two questionnaires, the first eliciting overall preferences and attitudes to CF, and the second eliciting preferences for specific types of feedback aurally modelled through a digital recording designed for the purpose of the study. In addition, a sub-sample of participants was selected for follow-up interviews. The results revealed similar preference within both cultural groups supporting the use of error correction in the classroom. However, the preference for the rate of error correction was lower for the Taiwanese students than for Quebecois students. This may be because the Taiwanese students were less used to communicative interactions, which would feasibly lead to increased opportunities for error correction; while Quebecois students may have been more accustomed to interactional classroom environments thus leading to increased instances of error correction. It was also found that phonological errors were of greater importance compared to grammatical or lexical errors for both Taiwanese students and Taiwanese non-native teachers of English (NNTEs). Contrarily, Quebecois students and native teachers of English (NTEs) in Quebec felt that grammatical errors were the most important. As for the type of CF, Taiwanese students' preference for explicit correction was significantly stronger than that of their Quebecois counterparts in response to a pronunciation error. Both Quebecois and Taiwanese students ranked explicit correction significantly higher than recasts and prompts, whether the error was grammatical or

phonological, and ranked recasts significantly higher than prompts, but only in response to a pronunciation error.

In sum, this review has suggested that learners who were used to being oriented to a certain type of CF developed further than those who were not used to this orientation (Sheen, 2004) and that corrective feedback was more effective in FL contexts than in SL contexts (Li, 2010). This was also supported in the more positive attitudes that learners in FL contexts had toward error correction than learners in SL contexts (Loewen et al., 2009), making it more likely for the effects of feedback to be incorporated.

Notably, the studies mentioned above have not systematically investigated the relationship between attitudes to CF and their effectiveness. Given that we still have "little understanding of how the learning environment affects inter-language development" (Sheen, 2004, p.264) and CF efficacy, one of the objectives for this study was to examine the possible variation in attitudes towards and effectiveness of CF in two different instructional contexts (UK and SA). To date, little research has focused on possible cross-cultural influences on learners' preference for particular forms of error correction, and none has used, specifically, recast and metalinguistic information CF techniques, with same background population (both from Saud Arabia, in the current study) but learning in two different contexts (EFL and ESL) that investigates efficacy using a battery of outcome measures.

In the current study, data collected about learners' perceptions about CF techniques will be presented in later chapters. However, prior to the study, the professional experience of the author (15 years of teaching at the university level) in Saudi Arabia, the situation is broadly as follows: Teachers, for Saudi students, are the only source of knowledge, therefore "Saudi students find it difficult to accept a teacher who does not play a dominant leading role" (Alahmadi, 2007, p.4). The provision of explicit information, including metalinguistic information, is very frequently used in Saudi classrooms, which are often teacher-centred. The classroom teacher frequently repeats the rules and provides examples, with a great emphasis on form rather than meaning and therefore students are not prepared for communication (Al-Awadh, 2000). Interactional meaningful activities are absent in most Saudi classes for structural and cultural factors such as the large number of students and the lack of experienced and

trained teachers. The small contribution of students, if there were, to the classroom discussion is "pre memorized and should not necessarily stem from communication breakdown"(Alahmadi, 2007, p.4). Recasts are less frequent than metalinguistic CF.

In the UK context, learners are probably familiar with both these CF techniques, and also with oral production tasks. However, recast was the most frequent feedback type used by teachers in ESL classrooms (e.g., Sheen 2004; Suzuki, 2004). In an oral interview, some Saudi learners, studying English in a language centre in the UK, supported the use of both types of CF by some native teachers of English in response to learners' errors. Oral interaction is available in the UK context as it is believed that learners are given the chance to use and practice English and what s/he has learned in a language classroom, everywhere and with everyone from the waiter to the grocer to the bus driver.

Given the above, several scenarios justify the need to compare the effectiveness of recasts, metalinguistic and oral production tasks alone, across these two different educational contexts. In the Saudi Arabian context, as students are not very familiar with recasts as a CF technique (partly because of the lack of oral production practice in class) it may be that recasts will not be perceived as CF by the students, and so reduce their effectiveness in promoting learning. Their familiarity with metalinguistic information as a CF technique may mean that this is more. On the other hand, it is also feasible that the novelty of recasts *increases* their impact, and the familiarity of metalinguistic information *reduces* its impact.

The lower familiarity with oral production tasks themselves in the SA context may mean that this novelty produces a large observable impact on learning; on the other hand, it could mean that the lack of familiarity is not conducive to eliciting the target form, therefore leading to few opportunities for practice and correction.

It is acknowledged that these are not precise hypotheses. They are provided as a narrative justification for the comparison between the two contexts, in addition to the calls for such comparisons in the literature reviewed above.

CHAPTER 5: THE TARGET FORM - ENGLISH MODALS

Introduction

Acquiring modals, form and function, is an essential part in the process of learning English. Modal verbs have numerous subtleties depending on the context in which they occur (Kreidler, 1999). This, in fact, is one of the factors that make mastering modals a challenge for ESL/EFL learners. Celce-Murcia and Larsen-Freeman (1999) define modal auxiliaries as "one of the most difficult structures that an ESL/EFL teacher has to deal with" (p.80). They state some of the problems ESL/EFL learners face: (1) students have been told that the third person singular present tense verb in English requires an *'s'* ending, so learners will tend incorrectly to generalize this rule to modals (2) modals precede a lexical verb without an intervening *to* infinitive, but ESL learners will often use a *'to*' following modals (3) differences with learners' native language can cause problems and (4) the different meanings each modal can have make it so difficult for EFL/ESL learners to incorporate meanings and functions.

It is important to note here that the focus of the current study is on the form of modals rather than the meaning associated with them. In addition, the study focuses on the form of 'deontic' (social interaction) rather than 'epistemic' (logical probability) modalities.

This chapter provides a definition of English modals as the target form, the properties of both English and Arabic modals, some typical learners' errors with some explanation of these errors and finally reviews some of the previous research in this area.

5.1 Some Working Definitions of the Forms under Focus

English Modals

Shaffer (2004) defines modal and modality as the conceptual domains of necessity and possibility. These domains can be expressed in a given language by modals (lexemes or auxiliaries) or grammatical mood (inflectional coding on the verb). Bowen and McCreary (1977) describe the usage of modal verbs in English as unique, partly because of gaps and overlaps, partly because of distinctive signification and also because

they are unique in their structural and semantic distributions. Modals, thus, do not work in the same way in English as they do in many other languages such as German, Spanish, Panjabi, Farsi and Arabic.

Properties of English Modals

According to Stevenson (1987), there are nine true auxiliaries of mood, also called modals: can, could, may, might, shall, should, will, would, and must. They are used with a main verb to express ideas such as ability, certainty, possibility, intention, or necessity etc. English modal auxiliaries differ from other verbs (main verbs) as follows:

1 **Inversion:** Modals can undergo inversion in interrogative clauses:

Can you come tomorrow?

Must they leave?

Other main verbs cannot undergo inversion in interrogative clauses:

*Leave they?

2 **Agreement**: Modals do not inflect for the subject-verb agreement morpheme –s:

*He cans, musts, wills etc

Other main verbs must have an agreement between the subject and the verb as having an *-s* for the third person singular:

He goes to the market.

3 **Non-finite form**: Modals always take infinitive verb forms as their compliment:

He will go home.

Other main English verbs take to infinitive after:

He wants to go home.

She is planning to study abroad.

4 **Negation**: The negative marker comes directly after the modal:

He cannot do the assignment.

Other main verbs take the negative markers before:

*They walk not.

5 **Double modals** are not allowed:

*He may will come.

Arabic Modal

When English modals were compared with their equivalents in Arabic, Western writers claim that modal auxiliaries are non-existent in Arabic (see for example Smith, 1987, 2001). Abunowara (1996, 2005) suggested two reasons behind this idea: 1) modality in Arabic has not yet attracted much attention from either Arabic or western linguists, 2) Arabic modal forms do not have distinctive syntactic characteristics that the modal verbs have in English for Arabic modalities have verbal meaning and, therefore, tend to become similar to verbs. Thus, their function is performed by normal verbs, often impersonal, or prepositions followed by the subjunctive (present) tense. In other words, English modals are accomplished with different structures in Arabic and the verb that follows is a finite verb form, whereas English uses non-finite verb forms. But, whether they are verbs, adjectives, or particles, these forms express a wide range of 'modal' meanings (see Abunowara, 1996). Also Saeed (2009) pinpointed the lack of an equivalent modal system in Arabic as the main cause. Another factor was argued to be the treatment of modals in EFL textbooks, as the focus on forms approach rather than on functions and meanings in most of the textbooks is thought to present a challenge even for advanced learners. Thus it is challenging for ESL/EFL teachers to impart knowledge on modal auxiliaries and for the learners to use the knowledge in their work.

The following categories of modalities in Arabic were outlined by Anghelescu (1999):

- Epistemic modalities or modalities of assertion such as: it is known, admitted that. According to their linguistic function, they are divided into the following:
    a. Certainty, meaning for example: *sahih anna* (true that) and *min almua'kkad anna* (certain that*)*, etc.
    b. Doubt or uncertainty, meaning for example: *zaana* (to believe)
    c. Modalities expressing anticipation: *min al – muhtamal* (probable that)
- Alethic modalities: it is absolutely true, possibly true and this could be classified as follows:
    a. Modalities expressing necessity,e.g., *min ad-daruri, la buda*.
    b. Modalities expressing possibility*: min al – mumkin, yumkin*

c. The particle *qad* which is associated with the imperfect

i. La'alla , rubba and *rubbama*

- Deontic modalities indicate obligation and, permitted and can be divided into:
  a. obligation such as*: la buda min, labuda an, yajib an, yanbaghi an*
  b. permission such as: *yumkinu, yumkinuka, min al jaiz, yajouz*
- Evaluative modalities which means it is good that (*ahsun, min almustahsan*)
- Boulomaic modalities that indicate desirability *(min al marghub fihi)*

Properties of Arabic modal verbs

Given that modals in Arabic and English have totally different system, some properties of Arabic modal verbs and their English counterpart modals are illustrated in the following examples:

1- At the negative level, e.g., (Alharbi, 2002)

Arabic script:	*Ahmad*	*La*	*Yastatiig*	*an*	*yalab.*
English transliteration:	Ahmad	Not	Can	that	play.
Proper English:	Ahmad cannot play.				

2- At the declarative level, (e.g., Saeed, 2009, p.76)

Arabic script:	*ᶜalia*	*'an*	*'adhhaba*	*'alaan*
English transliteration:	On-me	that	go-I	now
Proper English:	*I must go now*			

3- At the interrogative level, (current data)

Arabic script:	*Hal*	*tastatiig*	*un*	*talab*	*?*
English transliteration:	Qword	can you	that	play	?
Proper English:	Can you play?				

5.2 Typical Errors

The errors of Arab lower intermediate learners of English to produce modals demonstrate the difficulties they have, and suggest that their complex forms and functions, and the lack of such a system in Arabic may be reasons for this.

Modals in Arabic are accomplished with different structures. The verb that follows a modal is a finite verb form, whereas English uses non-finite verb forms. In fact, in Arabic, modality is not frequently used in everyday speech as in English. For example, in Arabic, we could say:

Open the door please.

Instead of

Can / could / would you open the door please?

Errors from the current data

The modal system in Arabic therefore carries a low functional load contrary to English (Abunowara, 1996). Arabic speakers tend to use the non-native like finites, add regular verb endings, use auxiliaries, and over-use '*that'* clause with them. The following examples (taken from the data collected in this study) illustrate these problems:

1- The insertion of another auxiliary verb with English modals

- *The son of my uncle did not can live there.

2- The use of –ing ending with the main verb after or with English modals

- *If she can painting.
- *She is canning.

3- The use of an –s ending in the main verb with third person singular or with English modals

- *She will cuts mark.
- *Anna cans climb a mountain.

4- The use of past form after English modals

- *He should told the people around.

5- The use of two auxiliary modals in a sentence

- *The police will should find the killer.

6- The use of to infinitive after must and should

- *The players must to wear the uniform.

7- The use of able to with English modals

- *Bell cannot able to buy nice clothes.

8- The question form of modals takes the same form as the sentence

- *How she can get the lesson.

9- The use of that clause

- *I can that I go.

5.3 Possible Explanations of These Errors

The errors noted above might be due to one or more of L1 transfer, overgeneralization, ignorance of rules, incomplete application of rules and simplification (as suggested by Abunowara, 1996). Abunowara found that errors within the target structure for intermediate learners were due to transfer ,while for advanced learners they were more likely to be due to overgeneralization for example, learners create a non-acceptable structure on the basis of his / her experience of another structure. Learners are taught to add the suffix –s to the verb in present simple tense with third person singular. They then apply the same rule either to the verbs following a modal or to the modal verb itself, e.g., *he can reads. Or, *he cans read. Another example is the use of to- infinitive. Learners are taught as the ‘ought to’ construction: He ought to come tomorrow. So a

learner might think it is appropriate to use *to* with must and should e.g.*He must to come tomorrow.

Abunowara (1996, 2005) also argued that teaching methods may be responsible for some of the errors. However, as the focus of the current investigation is not a study of the nature and sources of difficulties that EFL/ESL learners encounter when learning English modals, further discussion of these is beyond the scope of the thesis.

5.4 Previous Teaching and Learning Research in This Area

This section reviews previous research that has looked at the acquisition of modals amongst L2 learners.

Previous research on English modal auxiliaries

Bahns (1982) described longitudinal data collected by Wode in 1875 on naturalistic L2 English by four native German children aged 4-9. The focus was on how the children dealt with English modals in negation and interrogation and the order in which modals are acquired. The results were compared with other longitudinal studies of L2 acquisition. The study showed a developmental sequence for negation and questions using modal auxiliaries, in which 'can' was the most frequently used and the first to be acquired. This result is compatible with Saeed (2009) who indicated that the uses of *can*, which express *ability*, are found to be relatively easy for learners. This finding is also in line with Vethamani et al (2008), which will be discussed later.

The acquisition of English modality among Panjabi speaking pupils in primary and secondary schools, and the relative effects of age, function, context, geographical areas and L1 and L2, was documented in a study by Gibbs (1990). Responses were elicited for four root modality functions: ability, permission, possibility and hypothetical possibility, at the declarative, negative and interrogative environments. The results indicated that the earlier the age at which students are exposed to English modals the better their overall performance. The results showed that pupils at primary school outperformed secondary school pupils. The result also demonstrated that "English L2 pupils follow a similar developmental pattern for modals to the LI pupils in this study, with LI performance at all times ahead of that for Panjabi-speaking pupils, usually by

about two years. Given that the English mother-tongue pupils have had five years' more exposure; it appears that the second language subjects pass through the developmental stages more quickly. Thus the latter follow the same order but have a faster rate of acquisition." (p.309)

Hinkel (1995) reported a study determining whether NNS and NS usage of modals vary in relation to each other in the contexts of different topics for four hundred and fifty-five essays written by speakers of Chinese, Japanese, Korean, Indonesian and Vietnamese, which were compared to two hundred and eighty essays on similar topics written by NSs of American English. She found that native and non-native speakers of English use ***must,*** *have to*, and ***should*** in different contexts. She concluded that usage of the root modals ***must***, have to, ***should***, ought to, and need to in NS and NNS writing appears to be culture and context dependent. Hinkel emphasized the importance of having ESL/EFL students learn about the use of modals in context rather than working on the forms.

Another study (Vethamani et al, 2008) investigated the use of modals by Malaysian students in written tasks at two different competent levels (low and high). The findings showed that the most commonly used modals for the two competent levels are *can* and *could* and their negative forms. It also showed that the present tense modals were most apparent at the lower level and the past tense form was more dominant at the higher level. Vethamani et al (2008) found that students had difficulty in deciding the appropriate modals for specific functions, and that students were relatively more successful at using modals to express ability and certainty and that modals of probability or possibility showed lower frequency in writing.

Saeed (2009) investigated the extent to which University Arab learners of English have mastered English modals at the levels of recognition and using the appropriate modal verb. The study used a questionnaire which was distributed to 50 English major university students who had studied English for 12-14 years and who had scored 500 or more on the TOEFL. The 40 contextualized items in each version of the questionnaire attempted to test the major functions of modals: *possibility, ability, permission/offering, requesting*, and *suggesting /advising*. The students' responses were examined and each answer in both the recognition and production versions of the questionnaire was given a

grade of 1 or 0 depending on its correctness and incorrectness. Findings of the study reveal that the students have considerable difficulty recognizing the correct uses of modals. The results on learners' performance revealed that the average percentage of correct responses in the recognition part was 64% against 49% in the production part.

The above review has broadly informed the choice of linguistic structure for the current study. However, based on an anecdotal observation and the above erroneous production made by Saudi learners, I have considered the modal auxiliaries ***can***, ***will***, and ***must*** in declarative, negative and interrogative forms, as they seem to cause a wider range of difficulties for Arab learners of English.

Possible solutions to the problem

There could be several possible avenues to try to improve learners' use of English modals Abunowara (1996) suggested that "the more familiar a learner is with the target language, the more his reliance on his native will decrease" (p. 382). This is in line with DeKeyser (2007) who emphasized the need for practice that allows integration of form and meaning in a communicative setting. Another avenue may be to design 'referential Processing Instruction' activities (Van Patten, 2004) to provide learners with "opportunities to derive meaning from form in the input, whereas equivalent input without such opportunities did not"(Marsden, 2006, p.551). Vethamani et al (2008) recommended the incorporation of forms and functions in provision of sufficient exercises to allow practice and understanding of modals.

However, the notion of familiarity via practice was not supported by the results obtained in Saeed (2009). The findings of his analysis revealed that the performance of the students in both forms of the measuring instrument was remarkably low commenting "such low results are shocking, given that these students are supposed to have attained a level of proficiency that should enable them to perform better"(p.92).

However, the literature search for the current study did not locate any research to date that has empirically tested the effectiveness of an educational meaningful intervention to improve learners' use of certain English modal auxiliaries. Further, only one study that was located has looked at, specifically, Arab EFL learners' recognition and

production of English modal verbs, carried out by Saeed (2009). The current study begins to address these gaps.

CHAPTER 6: RESEARCH QUESTIONS

Following the literature reviews in the previous chapters, the following research questions were formulated:

RQ1 Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic?

- 1a) What is the effectiveness of these three intervention types relative to each other?
- 1b) Are any gains maintained after a delay of about 7 weeks?
- 1c) Are gains observed differentially on different outcome measures?

RQ2 Are results observed differentially in different contexts: EFL in SA, and ESL in the UK?

RQ3 What are learners' opinions about the different feedback techniques?

- 3a) Do opinions differ according to the context in which the study was done?

EXPERIMENT 1: UNITED KINGDOM CONTEXT

CHAPTER 7: EXPERIMENT 1

Introduction

This chapter reports the methods and data analysis for the study undertaken in the UK (ESL) context.

The structure of the study undertaken in the UK will be described. Chronologically, the pilots were, of course, carried out first, however, for clarity; they will be described once the main features of the study have been laid out. Section 7.1 describes the ESL research site. In section 7.2, recruiting the participants and their educational background will be laid out. Section 7.3 focuses on the design of the study including timing, randomization, and group sizes. Section 7.4 will describe the interventional instructional materials that were used in the treatment sessions. The two different styles of CF treatment and procedures, supported with authentic examples, will be discussed in section 7.5. The battery of tests and the order of their application and the procedures of the different pilot studies conducted to measure the validity and strengths of tests and activities will be presented in section 7.6, the data analysis and inferential statistics used will be presented in section 7.7. Detailed results and analysis in regard to the research questions will be discussed in section 7.8.

7.1 ESL Research Site

This was an English language centre in York, UK to which learners come from different backgrounds and ethnic groups. Some have experience of previous education and some come from prior occupations. The percentage of Saudi students in this centre is very high. There are fourteen classes for six different levels at the centre, from beginners to advanced, with an average class size of eleven students. The students are placed at each level following a placement test. The centre offers English language courses for short and long term as well as training sessions for student teachers. The number of students at the pre-intermediate level is usually higher than at other levels, therefore, this level was chosen for the study. In addition, the pilot studies suggested that this level of student were at the stage of trying to use modals, though made many errors in their use.

7.2 Participants

The participants in this study share the same L1 (Arabic), though came from different high schools and places in Saudi Arabia. English is taught in Saudi Arabia as a foreign language, with learners typically spending six to twelve years learning English for between six to eight forty-five minute periods per week. (In 2006, a policy established to start English from the last year of elementary school (year six), except for private schools, which usually start English from early years of schooling, though this cohort is still too young to be involved in the current study). English is the medium of instruction in various colleges and universities in Saudi Arabia, the use of English is compulsory in the fields of science and technology and nowadays, it is commonly used in business and marketing as well.

The subjects in the UK setting were thirty-four males and two females, aged between nineteen and thirty four (average = 24). They were assigned to different groups randomly (discussed in detail in section 7.3.2).

7.3 Design of Study

7.3.1 Timing

This study used an experimental classroom design. The study involved three levels of between-subjects variables (i.e. treatment conditions) and three levels of within-subject variables (i.e. testing times). A battery of pre-intervention tests, posttests and delayed posttests were used to assess students' acquisition of English modals.

The study duration was thirteen to fourteen weeks. Consent forms were signed in the first week followed by pre-treatment tests and four consecutive weeks of classroom intervention sessions. Immediate posttests were administered two days after the intervention sessions followed by a background and attitudinal questionnaires. Delayed posttests were administered after a six to seven week interval after the post tests. Each test was followed by an exit questionnaire to investigate whether the test had raised awareness about the focus and purpose of the test (Figure 7.1).

Tasks for corrective feedback and non-corrective feedback groups	Weeks													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Consent Form														
Pre-Tests and Exit Questionnaire														
Treatment Sessions														
Posttest, Exit Questionnaire and Attitudinal Questionnaire														
Interval (normal instruction continued)														
Delayed posttests														

Figure 7.1 Design of the study

7.3.2 Selection of Participants and Randomization

The Language Institute agreed that its students could participate, and students volunteered to take part. Participants were randomly assigned to the different conditions. Each participant was coded with a number, and then each number was written on a small piece of paper and put in a box. A child picked a number out of the box without looking and the number was assigned, in order, to the recast, metalinguistic, then task-only group.

7.3.3 Group Sizes and Background Characteristics

After several visits and meetings with the centre administrators and the students, forty-four pre-intermediate students were recruited from an English language centre in York, of which eight students dropped out, leaving thirty-six students**.** The participants were divided into three groups: one who would receive CF via recasts only (n=13), another group who would receive CF via metalinguistic information only (n=13), and one group who would do the tasks only with no CF (n=10). In the current study, each group was also divided into smaller group with six to eight students maximum in each group as larger rooms were not available at the time of the treatment.

The average age of the participants was twenty-four. Demographic details of all participants involved in the study are, regardless of the group they were in, shown in Figure 7.2. All students in the UK setting came from the same language background though they originated from different regions and economical status. Based on the information given in the questionnaire, 75% of the students had not travelled abroad before coming to the UK, 75% of the students had studied English in schools (private or public) back in Saudi Arabia, but 25% were learning English in schools as well as attending English courses after school. As for the period of exposure to English, 22% of the participants had studied English for six years, whereas 28% were introduced to English for seven to nine years, and 28% were introduced to English more than twelve years, but 22% of the participants did not respond to this question.

It is likely that participants had been exposed to some amount of teaching of English modals at secondary level (e.g., Say it in English, and English for Saudi Arabia, Ministry of Education, 1998, 1999). The extent of this knowledge is shown in the presentation of the test results in subsequent chapters.

Several reasons were given for coming to study English in the UK: to train learners to speak and write academically well (e.g., in higher education), to enable them to acquire knowledge, arts and useful inventions, to be able to get a job promotion, to prepare themselves for the IELTS, and to be able to communicate with people from different parts of the world. The average period for the study group of staying in the language centre was seven months.

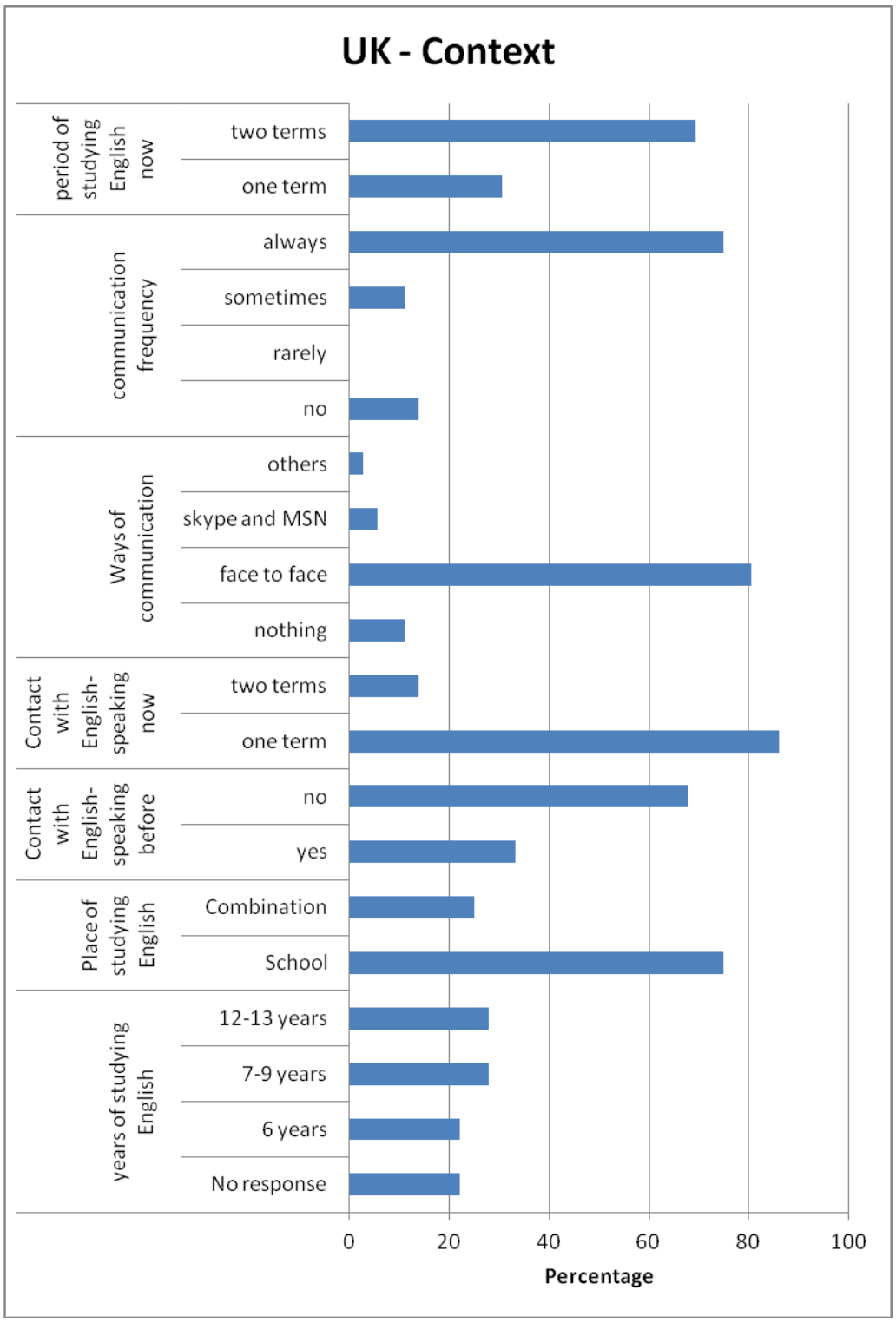


Figure 7.2 Demographic and background details of the participants in UK context

7.4 Interventional Instructional Materials

The intervention materials consisted of oral production tasks, following Nunan's (2004) definitions of task as "a piece of classroom work that involves learners in comprehending, producing or interacting in the target language while their attention is focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form"(p.4). Advice was taken from ten native speakers, and the tasks were piloted with non-native students. These pilots suggested they elicited obligatory contexts for modals. The following factors were taken into consideration in designing each task: 1) it should elicit the target structure, 2) should be appropriate for the students' background and proficiency level, 3) it should be of interest to them in the sense that they feel comfortable communicating, and 4) should have group work rather than individual.

Another design feature of the tasks was that most of the activities included in the interventional sessions started with warm-up activities in a question and answer format. This was done to lead learners to easily engage in the activities, produce more target language and be familiar with the upcoming tasks following the idea that "task familiarity could give rise to greater linguistic complexity or elaboration of the discourse" (Samuda & Bygate, 2007, p.111).

The tasks were two-way oral interactional tasks, where both (or all, in tasks where more than two students were involved) students have information to give and to receive. This included convergent tasks in which all students are working cooperatively towards an agreed goal, and divergent tasks in which students can come to different conclusions or outcomes.

The major objective of these sessions was to provide opportunities for the students to make rules, give instructions, make suggestions, give advice, and propose a future plan. All these could be elicited by using ***can***, ***must***, ***will*** and ***should*** (was elicited only in the revision session). These forms were first produced by the teacher during the interactional activities in two sources of input: 1) the warm up activities, tasks' instructions and the direction at the beginning of each session, 2) the provision of the metalinguistic information and the reformulation of the incorrect utterances or learners' repairs. These activities took place in both contexts, EFL and ESL.

The following examples point to the role of the teacher (the researcher) in the classroom. Directions and instructions were given to learners to know exactly what they needed to do in the activities:

1) In an activity that focuses on the use of ***must***,

Researcher: The Saudi Mission is trying to issue a pamphlet including some of the rules that must be followed by students who live with host families.

2) In an input that requires the use of the target form ***can***,

Researcher: In this activity each participant will be given a package of words. You need to use these words to tell your next partner what you can or cannot do.

3) An example of an input that requires the use of ***will,***

Researcher: We will help Enas in planning a trip. Each one of you has to tell her what she will need to do according to the pictures you have in your package.

Each group engaged in these interaction tasks in a classroom setting, one per week for four consecutive weeks. Each session lasted for forty-five minutes, equalling three hours of intervention in total. Norris and Ortega (2000) pointed out that the instructional treatment sessions within type-of-intervention research designs on average ranged from less than one hour to around four hours, and so the current study fits within this range. Another reason for this amount of intervention was that Ellis et al (2006); Sheen (2006); Yang (2008) suggested that longer treatment than one hour or two hours, might be required for a significant improvement evident in the implicit instruction such as recast.

The following subsections provide a sample of the materials (full interactional material sessions are included in Appendix B).

7.4.1 Session 1 (Can / Cannot)

In this session, five different activities targeted the use of **can/cannot**. Participants were required to answer questions, put words in context, create scenarios out of pictures, give suggestions and retell sentences. Figure 7.3 illustrates an activity aimed at expressing ability in different sporting activities.

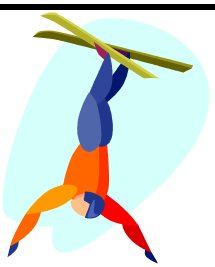
			
Flip forwards	Basketball	Tractor	Tennis

Figure 7.3 The use of can in expressing ability

7.4.2 Session 2 (must / must not)

In this session, learners had to explain rules in different contexts (e.g., a game, living with a host family and give instructions via four different activities). Figure 7.4 provides a sample of the activity that provided some opportunities to use **must / must not** in constructing rules of living with a host family.

		
No alcohol	No smoking	Lock door

Figure 7.4 The use of must in making rules

7.4.3 Session 3 (will / will not)

This session included four activities providing opportunities for using **will / will not** via: a restaurant conversation, future plan for a trip, parents' and students' concerns about going to college, and teachers' reactions towards a horrid student. Figure 7.5 is a sample of the planning a trip activity.

		
Accommodation	Food	Friends

Figure 7.5 The use of will in planning a trip

7.4.4 Session 4 (can/cannot, will / will not, must/ must not and should)

This session consisted of activities that provided opportunities for all the modals elicited in the previous sessions. Students were asked to come up with different ideas, give advice, match sentences with scenarios, and answer questions in a few sentences. Table 7.1 demonstrates one such activity. (Examples of ***should*** were elicited in this session to give the students an idea of the difference usage of ***should*** and ***must***).

Table 7.1 The use of can, will, must and should in a revision session

What will you do in the holiday? And where will you go?
What must you do in Saudi Arabia to obtain a driver's license?
Name two things that you couldn't do last year that you can do this year.

7.5 Oral Corrective Feedback Treatments and Procedures

The intention was to control the CF techniques as much as possible, to remain faithful to the conditions across different sub-groups within the same condition (and across the different contexts, UK and Saudi Arabia). However, as this was a classroom-based study, clearly this was not entirely possible, as learners do not respond to tasks in identical or predictable ways. The following sections lay out how the CF techniques were operationalized.

7.5.1 Recast Condition

Recasts can be presented in a full or partial form. In this study, recasts were full as illustrated in the example below. Any incorrect utterances apart from modals were ignored, except when a student requested some correctional information. This helped to maintain the engagement of the participants, and increased the external validity of the study.

When a student used a wrong lexical verb after a modal the correct forms for both were provided as shown below in the examples taken from CF episodes involving ***must can*** *and* ***will***. For more examples please refer to Appendix H.

Students non-native like production	**Teacher's correction**	**Students subsequent utterance**
*S1: my friend cannot song.	T: he cannot sing a song.	S1: he cannot sing a song.
*S2: you must to ask him.	T: you must ask him.	S2: you must ask him.
*S3: I will learning English.	T: so you will learn and study English.	S3: My parents' concern first comes here (this was a topic continuation rather than a reformulation).

Note the students did not always repair their utterances following a recast, though the examples here all illustrate correct uptake.

7.5.2 Metalinguistic Information Condition

This drew attention to finding a challengeable way to create activities that could be helpful for the students in maintaining their acquisition of English modals in an interactional communicative environment with the provision of rules simultaneously.

The following rules about the formal properties of English modals were included in the metalinguistic feedback: a) do not add *–s* in the third person singular present, b) have no impersonal forms (infinitive, gerund and participle), c) must be followed by bare infinitives (simple forms) of other verbs, d) can never be followed directly by another modal, and e) form the interrogative and negative forms without an auxiliary verb.

Some grammatical terms were simplified as some students seemed unfamiliar with them (e.g., the use of the term *helping verb* instead of *auxiliary*).

The following instances illustrate the procedure of providing metalinguistic information; please refer to Appendix H for more examples.

Students non-native like production	**Teacher's correction**	**Students subsequent utterance**
*S1: He can play flip.	T: Okay after can, after modals, we have to have one verb and flip here is a verb.	S1: Flip! He can... he can flip backwards.
*S5: you must when you listen to music you must do not make it loud.	T: with must we do not use do auxiliary. We use one main verb.	S5: Hum. You must make the music low and you must smoke outside the house and you must keep the bed tide every morning and the room keep it clean
*S7: he will not have a big dinner or big meal just snack or something light, and when he got to the hotel, he will made dinner.	T: You need to have a simple form of the verb after modals	S8: he will make

Note the student did not always take up the correction in their subsequent utterance.

7.5.3 Task Only Group (No Feedback)

This group was included to provide a baseline for comparison with the CF groups. The task only group was given all the same interactional activities that were given to the two CF groups but they were not provided with any CF.

The following table demonstrates three examples from the activities where errors were not followed by correction. More examples can be found in Appendix H.

Students wrong production

*S1: I think I cannot to be played.

T: Ok, what about the other one?

*S2: Ahmad is should do a timetable daily for do anything every day for do his homework and talking with his parents and help his mother or his father and do any sport and will he... He will be busy at every day.

*S3: I think I will said to Ahmad he should to be near to his son.

T: so he has to set a time for him, yeah?

It was sometimes difficult to maintain 'no correction', as some students would commit mistakes and ask if they were right or wrong. However, corrections were never provided for the target form.

7.6 Testing Instruments and Procedures

As discussed in chapter 3, Ellis et al. (2006) emphasized the importance of including a battery of tests to tap into different knowledge types, for example, implicit and explicit knowledge of learners. The tests were piloted with native and non-native participants, discussed in section 7.6.6. Same versions of the test were administered at three different times, pre-, post and delayed post-tests to assure same level of difficulty. To reduce learners' awareness of the target structure and the access to explicit knowledge, the tests were undertaken in the following order: free oral picture description, timed grammaticality judgment and gap fill as shown in Table 7.2. All 36 participants took each test. Each test will be discussed in detail in the following sections.

Table 7.2 Order of tests administered in the UK context

TESTS	NUMBER OF ITEMS	NUMBER OF PARTICIPANTS	FREQUENCY OF TESTS
Picture Description	7 Pictures	36	3 Times (Pre, Post, and Delayed post)
Timed Grammaticality Judgment	39 (3 training, 18 modals + 18 distracters)		
Gap Fill	17 (11 Modals + 6 distracters)		

7.6.1 Free Oral Picture Description Test

This test was undertaken on a one-to-one basis. It consisted of seven different pictures to elicit the use of modals. A participant was asked to look at the picture and give a full description. The approximate time for each participant was fifteen to twenty minutes but there was no time constraint and each participant took as long as they needed.

A digital mini-disc recorder was used for recording participants' production then for transcribing and analysing the data. Only sentences containing an obligatory context for a modal were transcribed.

Scoring

The accuracy of using English modals was determined for each individual participant, using suppliance in obligatory contexts scoring. The numerator and denominator therefore varied from one participant to another. If the target structure was supplied correctly a score of one was given. If the student failed to supply a modal in a context where a modal should have been used, or failed to produce a main verb, a zero score was given. Inter-language scoring was used whereby if a learner produced an obligatory context and used a modal and a main verb, but made an error such as inflecting either the modal or the verb, inserting *be* or *do*, or adding *to*. If a learner self-corrected, then only the initial incorrect production was scored as this would provide a

better measure of learners' implicit knowledge (Ellis, 2007). That is, learner's initial oral production (correct or incorrect) implies that reliance was on implicit unconscious knowledge, whereas self-correction implies reliance was on explicit conscious knowledge.

These proportions were then converted to a percentage for each participant. The scoring formula, adapted from Ellis and Barkhuizen (2005), was calculated as follows:

$$\frac{N \quad correctly \ given \ in \ context}{Total \ obligatory \ context} \times 100 = percentage \ accuracy$$

A copy of the test and scoring guidelines are included in Appendices (D) and (I).

7.6.2 Timed Grammaticality Judgment Test

This was a pen-and-paper test consisting of thirty-nine items, with eighteen containing modals, nine of which were grammatically correct and nine grammatically incorrect, eighteen distracters and three warm-up sentences (two incorrect and one correct) as shown in Table 7.3 below. The incorrect items were true productions that students have committed in the different pilots, and included errors in the following areas: (1) word order; (2) insertion of *'to'* after modals and before the main verb; (3) the use of two auxiliary modals; (4) the use of an *–s* ending in modals or with the main verb and *has* with third person singular; (5) the use of *'- ing'* form with the main verb; and (6) the insertion of another auxiliary verb '*be* or *do'*.

The order of the sentences was randomly scrambled. One item only was presented per page (e.g., Ellis et al, 2006). At the testing site, pens were distributed with each booklet and the instructions were read to the participants loudly. Participants were asked not to go back to the previous items that have already been answered or not to do the next one or turn to the next page unless the time beeper was heard (the time, ONLY, for each item was programmed on a Microsoft Power Point). Students were allowed to ask any relevant questions and to start when the clock began.

Test takers were required to (1) Indicate whether each sentence was right, wrong, or that they were not sure by ticking the appropriate box, (2) Underline the incorrect part(s), and (3) Write out the correction (s). In fact, the third condition was made believing that if learners were asked to indicate whether the sentences are right or wrong, only, they might guess, which would not provide a valid indication of their knowledge.

Allocated time varied from one item to another, ranging from six to thirty-four seconds. Although the apportioned time for the test items might seem long, there are many studies which used longer response times as shown in Table 2.1. (It should also be mentioned here that students taking the pilot test and pre-test expressed frustration at the time limit, demonstrating that they were indeed under time pressure).

Scoring

Scoring used the number of target items as the denominator (18). Distracters and training sentences were excluded. Learners' responses were scored as correct (1 mark), partially correct (0.5 marks) or incorrect (0 mark). For incorrect items, if a learner ticked correct for an incorrect item, indicated no answer, or ticked not sure, the answer was scored (0 marks). A learner was scored 0.5 marks for only ticking the correct box and underlining the relevant error but not providing a correction; for writing a correction only but not indicating the specific error; or for only ticking wrong but not underlining the error or providing the correction. The answer was scored (1 mark) if a learner underlined and provided a correction; or ticked wrong and provided a correction; or underlined, ticked and wrote out the correction. Students' scores were converted to a percentage. A copy of the test and scoring guidelines is in Appendices (F) and (I).

Table 7.3 Items in GJT test

Target Feature	Types	# of items
Warm-up	Grammatical	2
	Ungrammatical	1
Can / cannot	Grammatical	3
	Ungrammatical	3
Will / will not	Grammatical	3
	Ungrammatical	3
Must/ must not	Grammatical	3
	Ungrammatical	3
Distracters (13 tenses, 1 possessive, 1 sub/verb agreement, 1 adjective,1 pronoun and 1 passive)	Grammatical	9
	Ungrammatical	9
Total Items		39

7.6.3 Gap Fill Test

The gap fill test consisted of seventeen items, eleven targeted modals: ***will/ will not***, ***can/cannot***, and ***must/ must not***, and six distracters. The order of items was scrambled. Students were asked to look at the pictures provided next to each sentence, then fill in the blanks with the missing information; some blanks needed one word, others two or three, and some needed no word at all. The length of each blank was intended to be the same so that participants would elicit no information on the size of word/s that was/were missing.

Scoring

The denominator for the pre, post-and delayed tests was eleven (i.e. excluding the distracters). Inter-language scoring was used: A fully syntactically and semantically correct answer was given (1 mark). A partially correct response was given (0.5 mark), such as providing modals with a lexical verb that is inflected (*s, ed* or *ing)*, insertion of *be* or *do*, addition of *to* or the use of semantically incorrect verb. A learner was given zero for providing a modal but no verb. Scores for each participant were converted into a percentage score. A copy of the test and scoring guidelines are included in Appendices (E) and (I).

7.6.4 Exit Questionnaire

The exit and the attitudinal questionnaires for the current study conceptually replicated those used by Sheen (2006), though some adaptations were made. The exit questionnaire in Sheen's was designed to test whether learners were aware of the target of the tests and the treatments by giving them two questions. In her questionnaire, learners were given four options in the first question. The second question was an open ended for learners were asked to say what they learned from the tests and the treatments. Her questionnaire was administered after the error analysis test in the delayed post session.

Unlike Sheen, (2006) the exit questionnaire in experiment 1 of the current study consisted of one open ended question to find out if learners were able to identify the study's target structure, suggesting that the provision of options might indicate that the

test was testing more than one thing. The questionnaire for the current study, however, was administered after each written test (GJT and GF) at all three testing times. Please refer to Appendix G.

7.6.5 Background and Attitudinal Questionnaire

Although the attitudinal questionnaire for the current study replicated the one used by Sheen (2006), there were some differences in the content areas and the point of the scales. Sheen focused on measuring language anxiety and attitudes towards CF and grammatical accuracy. The attitudinal questionnaire in the current study focused on measuring attitudes towards three constructs: content of the activities, learners' opinions about error correction and accuracy generally and learners' opinions about the CF techniques used during the intervention. Sheen used six point Likert Scale, whereas five point Likert Scale was used in this study. Given that agreement has to be reached as to what to ask within a framework or model encompassing the research questions to be addressed and tested by the information obtained, the following table illustrates the relevant questions that were replicated from Sheen (2006).

Q	Current Study	Q	Sheen 2006
10	I feel it is better for me to know the corrections of my errors.	10	To improve my English, it is necessary that I learn from my own errors
5	I feel it is my teacher's duty to correct my errors all the time.	11	I want my teacher to correct my English errors all the time.
6	I feel frustrated when you correct me.	14	It bothers me when the teacher corrects my errors.
7	I feel better when you give me the rules.	17	I like to learn English by analysing my errors.
17	I prefer providing me with rules and information		
14	Having my errors corrected is the best way to learn English.	22	The best way to learn English is when the teacher corrects my errors.
9	I feel nervous about speaking after you have corrected my errors.	27	I am afraid of speaking right after the teacher corrects my errors.
20	I need to finish the activities fast so I can attend my other classes.	31	I always look forward to our English classes.

11	I feel that I am not used to being corrected when I do grammatical mistakes.	32	I do not like it if the teacher always focuses on grammatical errors.

To measure the validity of the questionnaire, it was first piloted on students in the same language centre for the main study, but with participants from an earlier cohort. The questionnaire was administered in English for this study was measuring learning of English language for Saudi learners and it was undertaken in an English language centre in the UK and in the Department of English in SA. In case of English lexical difficulties, an Arabic translation was made by the researcher.

Some of the questions elicited biographic data relating to age, gender, educational background, prior exposure to English, residency in the host environment or the country of origin, degree of contact with English outside the classroom, and attendance of English classes (for more demographic details see Figure 7.2).

The other items elicited three attitude 'constructs' (areas of interest): 1) attitudes toward activities, 2) learners' feelings towards error correction and grammatical accuracy generally, and 3) attitudes towards the CF provided in the study as shown in the following tables. Twenty-one five-point Likert scale (strongly disagree to strongly agree) items covered these areas, along with questions asking about participants' personal data.

1) Learners' opinions about the content of the activities:

Q	**The actual questions**
1	The activities are interesting.
2	The activities are not up to my level.
3	The activities are easy.
4	The activities are short.
20	I need to finish the activities fast so I can attend my other classes.

2) Learners' opinions about error correction generally:

Q	**The actual questions**
5	I feel it is my teacher's duty to correct my errors all the time.
6	I feel frustrated when you correct me.
8	I feel discouraged when I repeat the same errors.
9	I feel nervous about speaking after you have corrected my errors.
10	I feel it is better for me to know the corrections of my errors.
14	Having my errors corrected is the best way to learn English.

18 I think the most helpful way is correcting my errors directly.
19 I need a lot of time to think about my mistakes.

3) Learners' opinions about the different CF techniques:

Q	The actual questions
7	I feel better when you give me the rules.
11	I feel that I am not used to being corrected when I do grammatical mistakes.
12	I feel that this way of correction is new for me.
13	I am benefitting from your corrections.
15	I feel most comfortable with your direct corrections.
16	The corrections you have been providing are not important.
17	I prefer providing me with rules and information.
21	What you are doing does not improve my English.

Coding and Scoring for the Attitudinal Questionnaire

To code the attitudinal questionnaire, each group was given a number for example metalinguistic group was coded 1, recast group was coded 2 and task only group was coded 3. In addition, each participant was also given a number, e.g., S1, S2, S3 ……. etc. However, the ID for a student in the metalinguistic group was 11, 12, or 13…..etc.

In regards to the different questions, each question was also given a number from 1-21 as this was the last question of the questionnaire. Each response was also coded depending on the rank of the scale as the questionnaire was five Likert scale starting from strongly disagree(1) to strongly agree (5).

In terms of scoring the questionnaire, for each attitude construct, several questionnaire items were used to improve the validity and reliability of the measures. To obtain the scores for these three constructs, the following procedure was followed:

1) To calculate the mean scores, responses were added up for each survey response then divided the total sum by the total amount of questions.

2) Some questions may be counted as reverse points. They should be marked as five instead of being marked as one and so on. In this case, questions two, eleven, sixteen and twenty one were reversed.

Analysis

Each set of scores is analysed to investigate group differences, using one-way ANOVAs. In addition, Pearson correlation coefficients between the attitudinal scores and

the immediate post test scores on all tests for all groups were calculated as the questionnaire was carried out at posttest only.

7.6.6 Piloting the Intervention Materials and Testing Instruments

The initial design of the tests and the intervention materials started at the end of year 2007, and were first piloted in Saudi Arabia. In year 2008, another two pilots were undertaken in the UK to examine the validity of the tests and interventional materials, on native participants and non-native students. This section is placed here (perhaps counter-intuitively after the methods) as the preceding sections on the final tests and materials was first necessary to communicate to the reader the aims and nature of the materials and tests, so as to avoid repetition.

The pilots had three main objectives: to find out the most appropriate class level for the study; to check the intervention materials and tests elicited contexts for the target form and their difficulty seemed appropriate.

Three pilot studies were undertaken with English native speakers, and three pilot studies were undertaken with Saudi learners across the two different contexts as shown in Table 7.4.

Table 7.4 Summary of pilot studies

<table>
<tr><td rowspan="3">TESTS</td><td colspan="4">PILOT STUDIES</td></tr>
<tr><td colspan="2">Participants in UK</td><td colspan="2">Participants in SA</td></tr>
<tr><td>Native</td><td>Non-native</td><td>Native</td><td>Non-native</td></tr>
<tr><td>PD</td><td rowspan="3">8</td><td rowspan="3">6</td><td rowspan="3">2</td><td rowspan="3">9</td></tr>
<tr><td>GJT</td></tr>
<tr><td>GF</td></tr>
<tr><td>EI
(carried out in SA only)</td><td>6</td><td>9</td><td></td><td></td></tr>
</table>

Piloting tests and intervention materials started in Saudi Arabia on nine Saudi participants. Many significant changes were made to the testing instruments and intervention materials. The style of oral production test was completely changed to include more pictures that help in eliciting the target structure. For the grammaticality judgment test, the types of items were changed to include some genuine incorrect sentences produced by actual learners and some distracter sentences to distract learners' attention from the target structure. The number of items increased and the test become timed.

A new background and attitudinal questionnaire was designed after consulting Dr Graham Low concerning the scale and the categories of the questionnaire. Five -point Likert scales were chosen to be included in the study as mentioned in section 7.6.5.

A second pilot was conducted on ten native speakers to measure the validity of the materials and tests in eliciting English modals and the difficulty of the tests. The native speakers of English completed the oral production, gap fill and grammaticality judgment tests. They were asked to suggest possible modifications, so as to elicit particular English modals. Many suggestions were given regarding the picture description till the final version of the test was reached. For example, the pilots revealed that the pictures that were designed to elicit ***'must/ must not'*** actually elicited *'have to',* and modifications were made to address this.

In order to establish the time limit for each item in the timed grammaticality judgment test, test takers were asked to undertake each item as quickly as possible, measuring the time for each item with a digital watch.

The gap fill was altered to include a picture next to each sentence, more gaps and more sentences. Discourse completion and dehydrated tests were also piloted but were dismissed as they did not elicit knowledge of the target structure.

After each test the natives were asked, orally, whether they were aware of the target structure. The results showed that native participants did not know what the test was about except one participant. It was decided to introduce this 'exit questionnaire' after each test in the second pilot and the main study conducted in an ESL and EFL context. The intervention materials were improved to engage more interaction.

Another pilot on Saudi students in an English centre in York, UK was carried out. After a series of visits and emails, six participants from the intermediate level expressed their willingness to take part. A consent form was signed by the staff and participants. One of the reasons for selecting the pre-intermediate students was that this was the largest cohort.

The order of the tests in the pilot study was free oral picture description, timed grammaticality judgment and gap fills. This was partly done in order to reduce the likelihood that participants would become aware of the target structure, particularly relevant for the measures in which I hoped to reduce the influence of explicit knowledge. The same order was maintained for the main study in the ESL context but changed in the EFL context.

In the timed grammaticality judgment test, learners experienced difficulty with the time allocated for each test item. They expressed their unfamiliarity with being under pressure, and they thought that it was beyond their level. Since the time for each item had been measured on native speakers, who found it sufficient, and the aim was to exert considerable time pressure, the time limits were kept, but I increased the number of the warm-up items to three to familiarize the students with the time pressure.

During piloting it was found that no participants became aware of the focus of the tests (modals), very few participants expressed realisation that they were doing the same test three times, and no negative feelings were expressed about this. Therefore, a single version of the test was used for the pre, post and delayed posttests, in the main study.

The length of the interactional sessions in the second pilot study was increased to two hours over four continuous days. However, results indicated that thirty minutes for each session was not enough, partly for logistical reasons (arriving at the lesson on time) and partly because insufficient interaction and opportunities for correction was occurring. Consequently, the time was increased to forty-five minutes per session. This was also addressing one of the limitations found by Ellis (2006) and Sheen (2006) who found that the length of time for their communicative tasks should have been longer.

The intervention activities were carried out one-to-one for the pilot, digitally recorded. For the main study, the intervention was carried out in small groups, for two reasons: 1) meeting each participant individually was difficult to arrange 2) participants

were less likely to join the study as the one-to-one arrangement would have meant that some of them had to miss their regular classes. The participants in the main study were therefore divided into the three groups (recast, metalinguistic information, and task only) and the tests and intervention sessions ran during their spare time in groups, except for the free oral picture description test which was done one-to-one.

7.7 Statistical Methods

Mean scores and standard deviations were calculated for percentage scores from pre-, post- and delayed posttests for all measures.

A range of statistical tests were used to assess change over time and between conditions. Parametric or non-parametric tests were used where necessary. The following sections justify the choices made regarding the statistical tests.

Normality of distribution and note on use of non-parametric statistics

Before conducting these statistical tests, the normality of the distributions was tested, using Shapiro-Wilk test statistic (as this test is more accurate than Kolmogorov-Smirnov, Field (2009, p.546).

If data were non-normally distributed, non-parametric tests were used (Rasinger (2008) and Field (2009) suggest the use of non-parametric tests in case of non-normal distributions). These statistical tests are not often used in published research in L2 pedagogy research, and so some justification of their use is included. Norris and Ortega (2000) noted that "despite frequently low sample sizes, only limited use was made of non-parametric statistics (12% of studies)" (p.460), suggesting that non-parametric statistics should be used in cases where appropriate. Mackey and Goo (2007) reported "the most frequently used statistical method for main analysis was ANOVA (50%), followed by ANCOVA (15%), t-test (12%), chi-square (12%), Kruskal-Wallis (a non-parametric version of ANOVA) (6%), MANOVA (3%), and logistic regression (3%)" (p. 422). Discussing this low use of non-parametric tests, Larson-Hall (2010) suggests "many researchers are reluctant to use non-parametric statistics because they have heard that they have less power than parametric statistics" (p.58). (If this is true, then any claims made would be based on more *conservative* inferential statistics. But, in fact, this

is probably preferable to making claims based on statistics where differences are found when none are there). However, Larson-Hall argues "it is not accurate to say that non-parametric tests always have less power than parametric ones ...one cannot make a blanket statement about which kind of test is more powerful - it all depends on the circumstances"(p.58-9). Arguing further for the use of parametric and non-parametric statistics where appropriate, she noted "using either a parametric test *or* a non-parametric test when the data do not follow the assumptions can result in the loss of power to find statistical differences when they do in fact exist"(p.58).

A few published studies have employed non-parametric tests for specific outcome measures if the assumptions were not met for parametric tests for those particular measures. For example, Marsden (2005, 2006) used Friedman and Wilcoxon tests; McDonough (2007) used Kruskal- Wallis and Mann-Whitney tests; Sachs and Suh (2007) used the Mann-Whitney test because "Shapiro-Wilk's tests ... indicated non-normal distributions" (p.215).

Baseline equivalence

To find out whether groups have had similar baseline scores an ANOVA (if the data was normally distributed) or Kruskal-Wallis (if non-normally distributed) was used to compare pre-test results across groups (or contexts, UK *versus* SA). If baselines were statistically significantly different, then gain scores (i.e. post minus pre-test scores; delayed posttest minus pre-test scores; delayed posttest minus post test scores) were analysed using non-parametric tests (if data was non-normal) or ANCOVAs with the pre-test as a covariate (if data was normal).

Test reliability

The reliability of the scoring of oral and written measurements was carried out. Twenty five percent of the data for each measure was scored by native and native-like speakers of English. The decision of the researcher and the independent raters were submitted to Cronbach's alpha coefficients, the reliability coefficient on all measures are presented in the table below. This was considered sufficiently high for the researcher to independently code the remainder of the data.

Outcome measures in UK	Inter-rater reliability coefficient (r)
PD	**0.95**
GJT	**0.97**
GF	**0.99**

Summary of analysis processes

Following tests of the normality of the data and the equivalence of the baselines, the following decision process was used (full coverage of these required tests are laid out in Appendix J).

1- If the data was normally distributed and the baselines the same, repeated measures ANOVAs plus Planned Contrasts to test for interactions between time of test and group were conducted.

2- If the data was normally distributed and the baselines are not the same, a repeated measure ANCOVA was used with the pre-test as a covariate and posttest and delayed posttest as dependent variables interchangeably.

3- If the distribution was not normal, and the baselines were the same:

   a) To compare changes over time within each group: the data was split into different groups, and then a Friedman test was carried out (equivalent to a repeated measures ANOVA). If this showed a statistically significant difference then Wilcoxon test (paired test for within-subject comparisons) was used

   b) To compare posttest scores and delayed posttest across groups: a Kruskal-Wallis test was carried out (equivalent to an ANOVA). If this showed a statistically significant difference then a Mann-Whitney (paired test for between-subject comparisons) was used

4- If the distribution was not normal and baselines were not the same, gain scores were used. In these situations, gain scores were presented and three sets of gain scores were compared: pre to post; pre to delayed post; and post to delayed posttests, using Kruskall-Wallis and Mann-Whitney tests.

The use of one-tailed and two tailed tests

For the paired comparisons between the pre and post tests and between the pre and delayed posttests, one-tailed tests were used, as the hypothesis was that there would be gains in scores and a positive direction for the data was predicted. If it had not been possible to make any such predictions, a non-directional (two-tailed) test would have been appropriate and statistically significant difference in any direction (increase or decrease) needed to be detected. For more discussion see Butler (1985) or any other introductory statistics textbook.

In terms of the research questions, the following predictions (P) are proposed:

RQ1:	Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic?
Prediction 1	Recast, metalinguistic information and interaction tasks alone will help Saudi learners to develop their learning of English modals over time (one tailed tests will be used, where necessary).
RQ1a:	What is the effectiveness of these three intervention types relative to each other?
Prediction 1a	There will be a group difference between the interactional groups (recast and task only) and the metalinguistic information group in the UK as learners were used to communicative environment but no difference between the three groups in SA as the three intervention types were new for the EFL group (one tailed tests will be used, where necessary).
RQ1.b	Are any gains maintained after a delay of about 7 weeks?
Prediction 1b	All three groups will make no change between posttest and delayed posttest on all measures (two tailed test will be used where necessary) in both contexts.
RQ1c	Are gains observed differentially on different outcome measures?
Prediction 1c	In UK, similar gains of implicit and explicit outcome measures will be found as learners were exposed to language in a native English environment.

	In SA, explicit outcome measures will be found different from implicit outcome measures as learners were used to being instructed in a focus on forms classroom environment.
RQ2	Are results observed differentially in different contexts: EFL in SA, and ESL in the UK?
Prediction 2	There will be a difference between learners in an ESL context and EFL context on the "implicit measures" (i.e. free oral picture description) due to environment and learners' fluency but no difference on the "explicit measure" (i.e. gap fill) due to time constrained.
RQ3	What are learners' opinions about the different feedback techniques? a) Do opinions differ according to the context in which the study was done?
Prediction3	Learners in an EFL context have positive attitude towards CF more than learners in an ESL context.

7.8 Descriptive Statistics and Data Analysis

This section presents the results of test of normality and the baseline for all outcome measures used in the UK context, and the relative results and analysis to the following research questions:

Q1) Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic?

(1a) What is the effectiveness of these three intervention types relative to each other?

(1b) Are any gains maintained after a delay of about 7 weeks?

(1c) Are gains observed differentially on different outcome measures?

Q3) What are learners' opinions about the different feedback techniques?

Normality of data for all outcome measures

The results of the Shapiro-Wilk test are presented in Appendix J. The test of normality indicated that the assumption of normality was violated in the data from UK

context in free oral picture description test, timed grammaticality judgment (correct and incorrect items), and gap fill test, thus non-parametric tests were carried out for those measures. The data in the overall timed grammaticality judgment was normal, thus parametric tests were carried out.

Baseline parity

The results of ANOVA suggested significant similar baseline across different groups for the overall timed GJT test F (2, 35) =1.20, p=0.31. The results of Kruskal-Wallis suggested similar baselines for picture description test (H(2)=2.71,p=0.26), gap fill test (H(2)=1.60,p=0.45), and the grammatically incorrect items in timed GJT test(H(2)=1.46,p=0.48. For all these measures, the actual scores were analysed, rather than gain scores.

There was one exception to this, the test of normality was violated and the results of Kruskal-Wallis suggested different baseline across different groups for the grammatically correct items in timed GJT test (H (2) =7.59, p=0.02), thus the gain scores were analysed. Please refer to Appendix J**.**

Data on the different measures will be presented in tabular and graphical form, analysed, and discussed for each test in the following order: free oral picture description test, the timed grammaticality judgment test (total score, and grammatically correct and grammatically incorrect separately), and the gap fill test.

7.8.1 Free Oral Picture Description Test

Descriptive results

The mean scores and standard deviations are shown in Table 7.5 and presented graphically in Figure 7.6.

Table 7.5 Mean scores on free oral picture description test (PD)

Treatment group	N	Pre-test		Post-test		D Post-test	
		M	SD	M	SD	M	SD
Metalinguistic	13	85.34	14.96	94.77	6.59	93.52	9.35
Recast	13	88.19	8.90	91.22	6.63	95.84	4.41
Task only	10	84.93	30.25	94.15	5.74	96.14	2.67

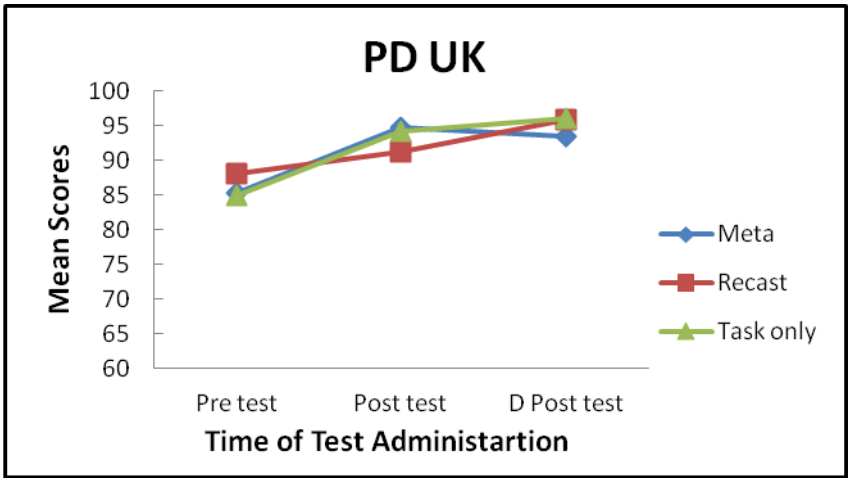


Figure 7.6 Mean scores on picture description

Analysis

The results of the Friedman test indicated a significant difference between testing times for the metalinguistic information group, χ^2 (2) = 10.92, p<0.01 and recast group, χ^2 (2) = 12.67, p<0.01. In contrast, there was no significant difference for the task only group, χ^2 (2) = 4.20, p = 0.12.

For the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre-test and posttest (Z = -2.76, p<0.01) and between pre-test and delayed posttest (Z = -2.82, p <0.01), but not between posttest and delayed posttest (Z = -0.94, p = 0.38).

For the recast group there was a borderline statistically significant difference between pre-test and posttest ($Z = -1.49$, $p = 0.08$), a significant difference between pre-test and delayed posttest ($Z = -2.67$, $p < 0.01$) and between posttest and delayed posttest ($Z = -3.06$, $p < 0.01$).

To find out whether there was a significant difference between the three groups at the time of posttest and delayed posttests, the Kruskal-Wallis test showed no differences between the three groups at posttest, $H(2) = 3.18$, $p = 0.20$ nor at delayed posttest $H(2) = 0.27$, $p = 0.87$. Detailed information is included in Appendix K (Table K.1).

Summary

Results suggested that the provision of CF via recasts and metalinguistic information helped the development of English modals, whereas the interaction tasks alone did not.

The results also indicated a significant gain for the metalinguistic information group in the pre- posttest time, but the recast group showed gains after the post test.

However, when groups were compared, the results indicated that the three intervention types did not seem to lead to different scores. This seems to contradict the finding that the metalinguistic group and the recast group made gains on tests but the task-only group did not. The lack of statistically significant gains in the task-only group may have been due to the high standard deviation in the pre-test. This wide variation may have meant the statistically significant change was more difficult to detect.

Although the results of PD test might simply reflect the fact that learners possessed ceiling levels of English modals knowledge at the beginning of the study, it is possible that the different corrective feedback and the interaction tasks alone increased learners' awareness of the grammatical target, thus encouraging them to monitor their output using their knowledge. However, this cannot be true as when learners were asked at the end of the final test if they were aware of which grammatical structures the test was measuring, no one was able to identify English modals. Second, as Table 7.5 shows, there is clear evidence that all the groups improved in the post test and delayed posttest.

7.8.2 Timed Grammaticality Judgment

Descriptive results

The accuracy of the mean scores and the standard deviations for the overall timed GJT scores (k=18) are provided in Table 7.6 and presented graphically in Figure 7.7.

Table 7.6 Mean scores on timed (GJT) test

Treatment group	N	Pre-test		Post-test		D Post-test	
		M	SD	M	SD	M	SD
Metalinguistic	13	54.27	13.25	64.10	11.70	62.39	13.06
Recast	13	47.86	11.46	49.57	16.89	60.26	18.12
Task only	10	45.56	18.29	54.44	22.50	58.33	18.93

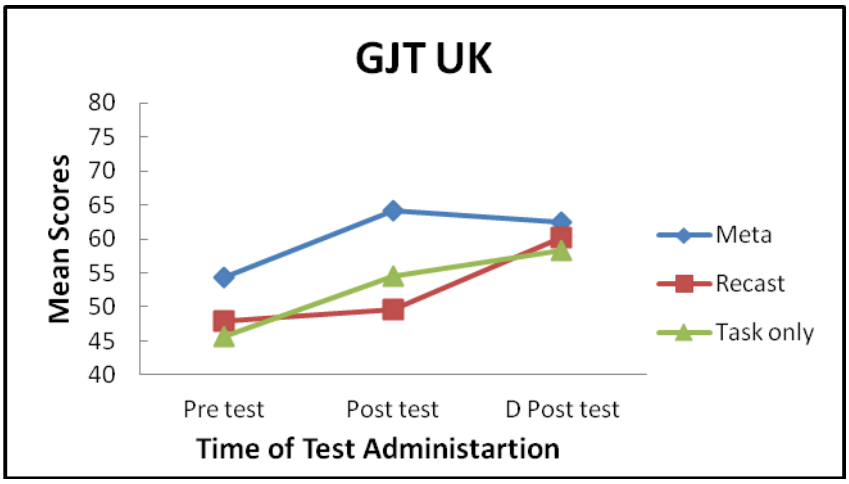


Figure 7.7 Mean scores on timed GJT test

Analysis

Overall scores

The mixed design ANOVA indicated a significant difference between testing times, F (2, 66) = 9.70, p <0.01, but not between the treatment groups, F (2, 33) = 1.32, p = 0.28. There was no significant interaction effect between tests and treatment groups, F (4, 66) = 1.21, p = 0.32.

The tests of within-subjects contrasts suggested a significant difference between pre-test and posttest F (1, 33) = 8.39, p <0.01and pre-test and delayed posttest F (1, 33) = 16.01, p <0.01but no significant difference between posttest and delayed posttest F (1, 33) = 2.98, p = 0.09. As there was no interaction between test and treatment, these gains were made regardless of the group learners were in. Please refer to Appendix K (Table K.3) for more information.

Summary

The results suggested that all the groups showed improvement as a result of the treatment in the form of interaction tasks alone, recast feedback and metalinguistic information feedback. The overall scores improved significantly on post testing for all groups, and these gains were maintained at delayed posttest, though no further gains were made between post tests and delayed posttest.

A planned contrast suggested a significant difference between metalinguistic information group and the recast group (the metalinguistic group scoring higher) on the post-testing time (p = 0.04). However, it is emphasised that as there was no overall interaction, such a planned contrast must be interpreted with caution as an indication of a possible trend, not a reliable difference.

7.8.3 Correct Items in Timed Grammaticality Judgment Test

Descriptive results

The mean scores on the correct items (k=9) of timed GJT are presented in Table 7.7 and in Figure 7.8. The gain scores are used because the baseline scores were not the same across groups.

The gain scores are presented in Table 7.8 and Figure 7.9 (Note the data from the gains scores of correct GJT items violated the assumption of normality, and so non-parametric tests were used). It is also important to note that to calculate the gain scores, pre mean scores were subtracted from post mean and delayed-post mean scores and the post mean scores were subtracted from delayed post mean scores.

Table 7.7 Mean scores on correct items in timed GJT test

Treatment group	N	Pre-test		Post-test		D Post-test	
		M	SD	M	SD	M	SD
Metalinguistic	13	82.05	17.30	84.62	12.45	75.21	12.95
Recast	13	70.09	19.45	64.96	22.61	72.65	21.57
Task only	10	56.67	19.21	64.44	21.47	77.78	16.56

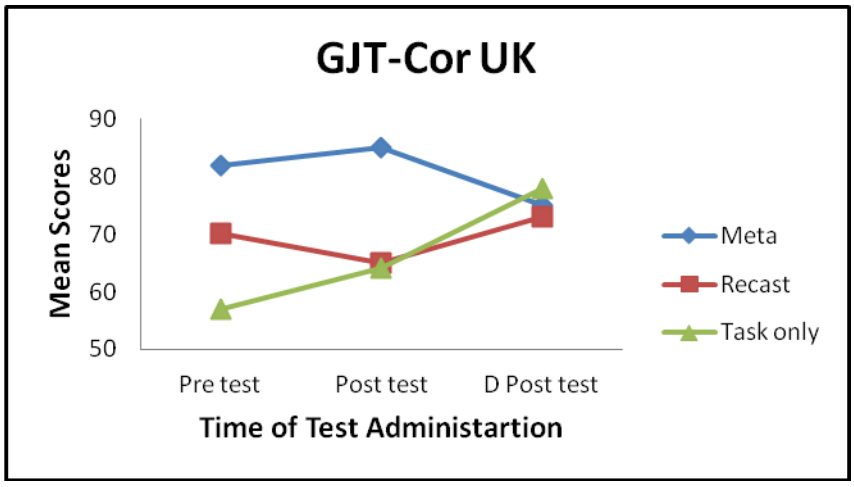


Figure 7.8 Mean scores on correct items in timed GJT test

Table 7.8 Gain scores on correct items in timed GJT test

Treatment group	N	Gain Scores on Correct Items		
		pre- post gains	pre -d post gains	post- d post gains
Metalinguistic	13	2.56	-6.84	-9.41
Recast	13	-5.13	2.56	7.69
Task only	10	7.77	21.11	13.34

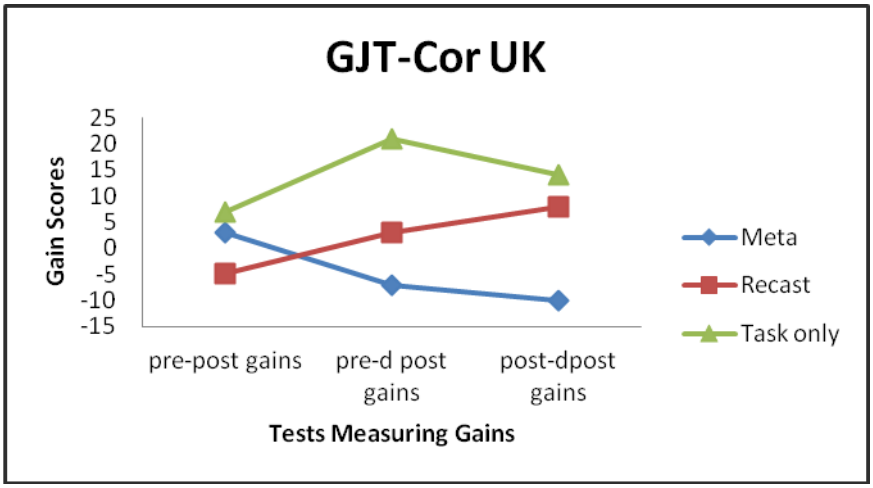


Figure.7.9 Gain scores on correct items in timed GJT test

Analysis

The results of the Kruskal-Wallis suggested no significant difference between groups on pre-post gains (H (2) = 1.87, p = 0.39). There was significant difference between groups on pre-delayed post gains (H (2) = 7.66, p = 0.02), and a trend toward significant difference between groups on post-delayed post gains (H (2) = 5 .70, p = 0.06).

Between the metalinguistic information and task only groups, the Mann-Whitney test showed a significant difference on pre-delayed post gains (Z = -2.73, p = 0 .01) and

a significant difference on post-delayed post gains (Z = -2.22, p = 0.03) on the direction of task only group.

Between the recast and the task-only groups, the Mann-Whitney showed a borderline statistically significant difference in the pre-delayed post gains (Z=-1.78, p=0.08) on the direction of task only group, whereas no significant difference between the recast and task only groups on post-delayed post gains (Z=-0.69, p=0.52).

Between the recast and the metalinguistic groups there was a significant difference in post-delayed post gains (Z=-1.76, p=0.04) on the direction of recast group, but no significant difference between pre-delayed post gains (Z=-1.02, p=0.28). For more information, please refer to Appendix K (Table K.4).

Summary

The results suggested that no group made significant gains on the immediate posttest, but gains were evident in the longer term.

As for group differences, the results indicated a beneficial role of the implicit CF (recast) and the interactional activities (tasks alone) at the delayed post-testing times for recast and task-only groups.

7.8.4 Incorrect Items in Timed Grammaticality Judgment Test

Descriptive results

The mean scores for correctly changing the incorrect items (k=9) and the standard deviation for all groups are shown in Table 7.9 and in Figure 7.10. Note the baseline data for the incorrect timed GJT items was the same, thus raw scores were used.

Table 7.9 Mean scores on incorrect items in timed GJT test

Treatment group	N	Pre-test		Post-test		D Post-test	
		M	SD	M	SD	M	SD
Metalinguistic	13	26.50	19.00	43.59	22.89	49.57	19.04
Recast	13	25.64	15.96	34.19	20.52	47.86	27.36
Task only	10	34.44	21.88	44.44	35.14	38.89	28.81

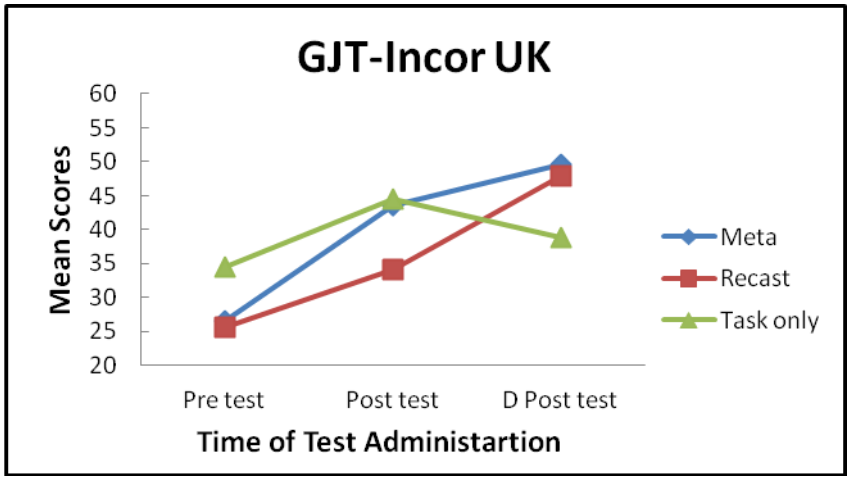


Figure 7.10 Mean scores on incorrect items in timed GJT test

Analysis

The results of the Friedman test indicated a significant difference between testing times for the metalinguistic information group, χ^2 (2) = 11.6, p<0.01and recast group, χ^2 (2) = 9.17, p = 0.01. In contrast, there was no significant difference for the task only group, χ^2 (2) = 0.79, p =0.67.

For the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre-test and posttest (Z = -2.56, p = 0.01), and between pre-test and delayed posttest (Z = -2.92, p <0.01), whereas no significant difference between posttest and delayed posttest (Z=-1.47, p=0.14).

For the recast group there was a statistically significant difference only between pre-test and delayed posttest (Z = -2.83, p = 0.01), whereas no statistical significant differences between pre-test and posttest (Z = -1.48, p = 0.14), and between posttest and delayed posttest (Z = -1.72, p = 0.09).

To find out whether there was a significant difference between the three groups at the time of posttest and delayed posttest, the Kruskal-Wallis test showed no significant differences between the three groups at post-test, H(2) = 1.05, p = 0.59 nor at delayed posttest H(2) = 1.62, p = 0.45. Detailed information is included in Appendix K (Table K.5).

Summary

Results suggested that the provision of CF via recasts and metalinguistic information helped the development of English modals, whereas the interaction tasks alone did not.

The results also indicated a significant gain for the metalinguistic information group in the short- and longer-terms, but the recast group showed gains in the longer-term only.

However, when groups were compared, the results indicated that the three intervention types did not seem to lead to different scores. This seems to contradict the finding that the metalinguistic group and the recast group made gains on tests but the task-only group did not. Please refer to Appendix K (Table K.5) for more information.

7.8.5 Gap Fill Test

Descriptive results

The results of the mean scores on gap fill items (k=11) and the standard deviation are given in Table 7.10 and Figure 7.11.

Table 7.10 Mean scores on gap fill (GF)

Treatment group	N	Pre-test		Post-test		D Post-test	
		M	SD	M	SD	M	SD
Metalinguistic	13	52.10	20.38	65.38	19.77	72.73	21.88
Recast	13	59.44	20.62	78.67	12.77	76.22	18.10
Task only	10	61.36	24.64	66.82	28.91	69.55	21.65

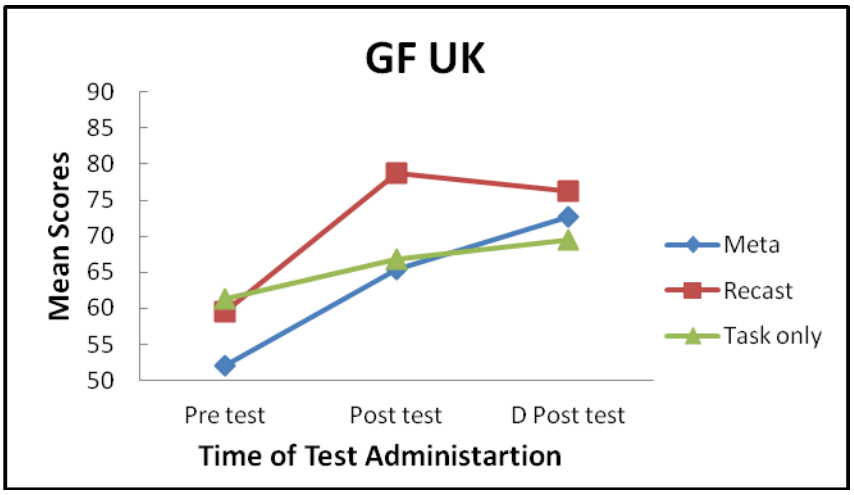


Figure 7.11 Mean scores on gap fill test

Analysis

The results of the Friedman test indicated a significant difference between testing times for the metalinguistic information group, χ^2 (2) = 9.18, p = 0.01 and recast group, χ^2 (2) = 7.68, p = 0.02. In contrast, there was no significant difference for the task only group, χ^2 (2) = 1.88, p =0.39.

For the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre-test and delayed posttest (Z = -2.69, p = 0.01), whereas no significant differences between pre-test and posttest (Z = -1.68, p = 0.09), and between posttest and delayed posttest (Z=-1.34, p=0.18).

For the recast group there was a statistically significant difference between pre-test and posttest (Z = -2.47, p = 0.01), and between pre-test and delayed posttest (Z = -

2.49, p = 0.01), whereas no significant difference between posttest and delayed posttest (Z = -0.63, p = 0.53).

To find out whether there was a significant difference between the three groups at the time of post-and delayed posttests, the Kruskal-Wallis test showed no significant differences between the three groups at post-test, H(2) = 3.22, p = 0.20 nor at delayed posttest H(2) = 0.47, p = 0.79. Detailed information is included in Appendix K (Table K.2).

Summary

Results suggested that the provision of CF via recasts and metalinguistic information helped the development of English modals, whereas the interaction tasks alone did not.

The results also indicated a significant gain for the metalinguistic information group in the longer term, but the recast group showed gains in the short-and longer-terms.

However, when groups were compared, the results indicated that the three intervention types did not seem to lead to significant difference. This seems to contradict the finding that the metalinguistic information and the recast groups made gains on tests but the task only group did not.

7.8.6 Attitudinal Questionnaire

This section presents descriptive statistics for three constructs: learners' opinions about the content of the intervention activities; learners' opinions about error correction and accuracy generally; learners' opinions about the CF techniques used during the intervention.

Learners' opinions about the intervention activities

The descriptive statistics for the scores of learners' opinions about the intervention activities indicated that the average scores for learners' opinions towards the intervention activities ranged from 3.22 to 3.46 with the metalinguistic information group achieving the lowest score (representing less positive opinions) and recast group attaining the

highest (representing more positive opinions) as shown in Table 7.11. However, a one-way ANOVA test indicated no significant differences between these groups F (2, 35) = 0.64, p = 0.54.

Table 7.11 Opinions about the intervention activities

Group	N	Mean (k=5)	SD	Min	Max
Metalinguistic	13	3.22	0.49	2.40	3.80
Recast	13	3.46	0.73	2.20	4.60
Task only	10	3.26	0.48	2.40	4.20
Total	36	3.32	0.58	2.20	4.60

(On a 1-5 Likert scale)

Opinions about error correction generally

Table 7.12 shows the descriptive statistics for the scores of learners' opinions about error correction generally. The average scores for learners' opinions towards error correction generally ranged from 3.33 to 3.46 with the task only group achieving the lowest score (representing less positive opinions) and recast group attaining the highest (representing more positive opinions). However, a one-way ANOVA test indicated no significant differences between these groups, F (2, 35) = 0.16, p = 0.86.

Table 7.12 Opinions about error correction generally

Group	N	Mean (k=8)	SD	Min	Max
Metalinguistic	13	3.39	0.63	2.63	4.63
Recast	13	3.46	0.63	1.63	4.00
Task only	10	3.33	0.42	2.25	3.75
Total	36	3.40	0.57	1.63	4.63

(On a 1-5 Likert scale)

Opinions about corrective feedback during intervention

Table 7.13 illustrates the descriptive statistics for the scores of learners' opinions about corrective feedback during intervention. The average scores for learners' opinions towards the different CF techniques ranged from 3.61 to 3.81 with the recast information

group achieving the lowest score (representing less positive opinions) and metalinguistic information group attaining the highest (representing more positive opinions). However, a one-way ANOVA test indicated no significant differences between these groups, F (2, 35) = 0.44, p = 0.65.

Table 7.13 Opinions about corrective feedback during intervention

Group	N	Mean (k=8)	SD	Min	Max
Metalinguistic	13	3.81	0.36	3.13	4.50
Recast	13	3.61	0.66	1.88	4.38
Task only	10	3.65	0.66	2.00	3.38
Total	36	3.69	0.56	1.88	4.50

(On a 1-5 Likert scale)

Relationship between attitudes and tests scores

This section provides analyses of correlation between the test scores and the results of the attitude questions. Correlations are only provided using the post test scores as the attitude questionnaire was not carried out at pre or delayed posttest.

Relationship between attitudes and post test scores on picture description

The results of a Pearson correlation analysis, presented in Table 7.14 revealed the results for the three different sets of questions (content of activities, error correction generally, and the CF in the intervention) and different groups.

For the metalinguistic information and recast groups, there was no significant relationship between the picture description mean posttest scores and any of the three attitudes constructs.

For the task only group, the picture description mean posttest scores was positively related to the 'general opinions towards error correction' with a Pearson correlation coefficient of $r = 0.72$ and $p < 0.05$ but not with the 'content of the activities' and the 'CF techniques during the intervention'.

Table 7.14 Correlations between the mean scores on picture description and learners' attitudinal scores

Group	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 13)	Post PD	0.41	-0.32	-0.18
Recast (n = 13)		0.34	0.17	0.30
Task only (n = 10)		0.12	0.72*	-0.05
*** $p < 0.05$**				

Relationship between attitudes and post test scores on gap fill

The results of a Pearson correlation analysis presented in Table 7.15 revealed the results for the three different sets of questions (content of activities, error correction generally, and the CF in the intervention) and different groups.

For the metalinguistic information group, a negative significant association was found between the gap fill mean posttest scores and the attitude towards the 'CF techniques during the intervention' with a Pearson correlation coefficient of $r = -0.55$ and $p < 0.05$ but not with the 'general opinions towards error correction' or the 'content of the activities'.

For the recast group, there was no significant relationship between the gap fill mean posttest scores and any of the three attitudes constructs.

For the task only group, a significant positive correlation was found between the gap fill mean posttest scores and the 'general opinions towards error correction' with a Pearson correlation coefficient of $r = 0.71$ and $p < 0.05$ but not with the opinions towards the 'content of the activities' or the 'CF techniques during the intervention'.

Table 7.15 Correlations between the mean scores on gap fill and learners' attitudinal scores

Group	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 13)	Post GF	0.12	-0.23	-0.55*
Recast (n =13)		-0.31	-0.24	-0.50
Task only (n = 10)		-0.23	0.71*	0.25
*** $p < 0.05$**				

Relationship between attitudes and post test scores on timed GJT

The results of a Pearson correlation analysis presented in Table 7.16 revealed the results for the three different sets of questions (content of activities, error correction generally, and the CF in the intervention) and different groups.

For the metalinguistic information, recast and task only groups, there was no significant correlation between the timed GJT mean posttest scores and any of the three attitudes constructs.

Table 7.16 Correlations between the mean scores on timed GJT and learners' attitudinal scores

Group	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 13)	Post timed GJT	0.09	0.17	-0.03
Recast (n = 13)		-0.37	-0.32	-0.02
Task only (n = 10)		0.26	0.46	-0.32

7.8.7 Exit Questionnaire

On completion of each individual written test but not the free oral picture description, learners were given an exit questionnaire relating to their awareness of the focus of the tests. The questionnaire was administered after each test to discover if the test had helped learners identifying the target structure. No statistical descriptive test was carried out as the results of the questionnaire indicated that none of the students in the three groups, in all testing periods, was able to recognize and identify the target structure of the tests and the instructional activities (i.e., modals). A sample of the questionnaire is included in Appendix G.

EXPERIMENT 2: KINGDOM OF SAUDI ARABIA CONTEXT

CHAPTER 8: EXPERIMENT 2

Introduction

This section describes the methods of the second experiment which took place in SA. The research questions were the same as for experiment 1. See section 7.3 for a summary of the study design. The same three treatment conditions, the same intervention materials and CF techniques were used as in the UK (discussed in Chapter 7).

The battery of tests used in the SA experiment had similarities and differences with the UK experiment. As in experiment 1, there were three testing times: pre-, post- and delayed posttests. Similar gap fill and same free oral picture description tests were used in the SA experiment but an elicited imitation test was also used (discussed in 8.8.1), the GJT was changed to an untimed test, and metalinguistic questions were added to the GJT (discussed in section 8.8.2). The attitude questionnaire was identical to the one used in the UK. The exit questionnaire was basically the same, with few differences (discussed in section 8.9.2).

8.1 The EFL Research Site

An all-female Saudi university, which is a governmental institution in Saudi Arabia, was selected as the EFL site. The researcher is a staff member there, and will be working in the same department when returning home. After visiting the department several times, talking to the head of the department and the instructors, and examining the results of the pilot tests, the 'second level' was chosen as potentially broadly comparable to the pre-intermediate participants in the UK setting.

Background

Students are placed in the department of English based on a placement test as well as their grade point average in their high school certificate. Students have to complete successfully eight levels in order to graduate from the department and be a qualified teacher of English. In the first four levels, students are introduced to different English major subjects such as grammar, writing, speaking, listening and reading as well as some minor Arabic subjects classified as college requirements. Later, they are introduced to

some more specialized courses such as English literature (mainly drama, novel, and poetry), and linguistics (such as Introduction to Linguistics, Morphology, Syntax, etc.). Practical training courses are introduced in third and fourth years.

8.2 Participants

Initially, seventy-four second level students were recruited. Teachers and participants were briefed about the study and provided with the timeline. Participants were also encouraged to commit their free time to the assigned classes and be punctual once they decided to volunteer.

8.3 Design of Study

8.3.1 Timing

This experiment lasted thirteen to fourteen weeks, the same as in the UK context. A consent form had to be signed in the first week. Each test ended with an exit questionnaire to investigate whether there was any awareness of the target linguistic form. Similar exit questionnaire to the one in the UK was given after completing each testing session but the same attitude questionnaire was given after completing the post tests.

8.3.2 Selection of Participants and Randomization

The participants in SA were selected on the basis of the student's timetable slots that were free. Each student indicated her free time on the contact information sheet and all those with the same free time were then randomly allocated to the recast, metalinguistic information, or task only groups. The students completed the intervention activities in groups of between six to eight students (mainly for practical reasons).

8.3.3 Group Sizes and Background Characteristics

Initially, the recast and metalinguistic groups both consisted of twenty-five students and the task only group of twenty-four students. Unfortunately, ten students

missed some of the treatment sessions, and so were excluded from the data set. Of the remaining sixty-four students, nineteen were in the metalinguistic information group, twenty-five were in the recast group and twenty were in the task only group.

The average age of all the subjects was twenty-one. Demographic details of the participants are shown in Figure 8.1. All students came from the same language background but different regions and economical status. The participants' exposure to English was limited to the classrooms, to media and in few cases travelling. Based on the information given in the background questionnaire, 89% of the students had not travelled abroad. 48% of the students had studied English in schools for six years, 44% for seven to nine years and 8% for twelve to thirteen years. 83% of the students had studied English in schools and 17% only were learning English in schools as well as attending English courses after school.

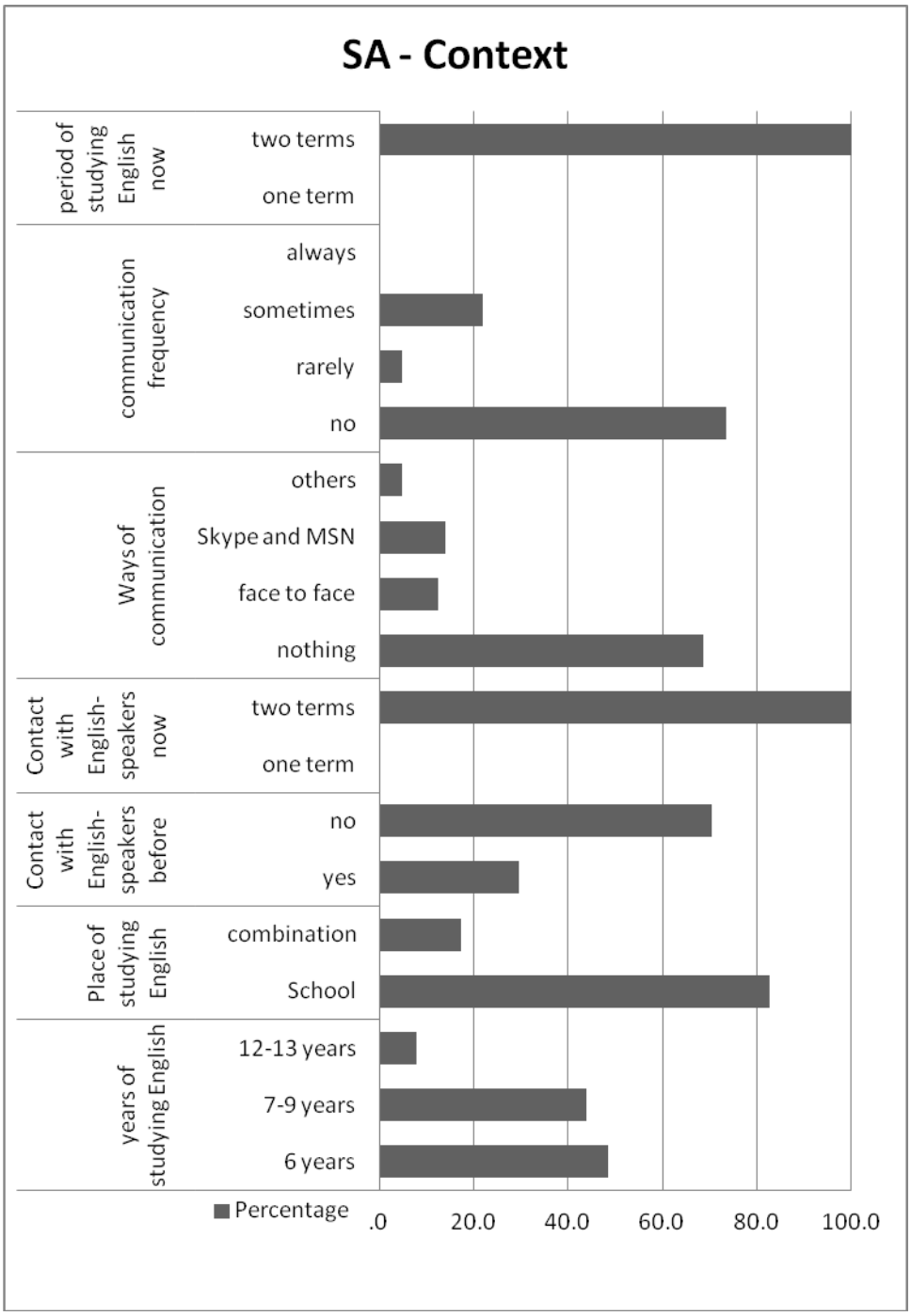


Figure 8.1 Demographic and background details of the participants in SA context

8.4 Interventional Instructional Materials

The intervention materials were the same for both contexts (ESL and EFL). Full details of the design and the different oral interaction tasks are shown in Chapter 7 section 7.4.and fully illustrated in Appendix B.

8.5 Oral Corrective Feedback Treatments and Procedures

The same CF techniques were used as in the UK described in detail in section 7.5, and full examples are included in Appendix H. Participants received either recasts or metalinguistic information following any errors in the target structure. The task only group did not receive any feedback.

8.6 Testing Instruments and Procedures

In this phase of experiment, two tests were similar to the ones used in experiment 1 and two were different. During each testing session, four tests were administered in the following order: elicited imitation, free oral picture description, gap fill and untimed grammaticality judgment including the metalinguistic questions tests as shown in Table 8.1. Each test will be discussed in the following sections.

Table 8.1 Order of tests administered in SA context

Test Type	Tests	Number of Items	Number of Participants	Frequency of Tests
Oral Implicit	Elicited Imitation	(k=60) 12 training, 33 Modals, and 15 Distracters	(n=64) one on one basis	3 Times (Pre, Post and Delayed post)
	Free Oral Picture Description	7 Pictures		
Written Explicit	Gap Fill	(K=22) 14 Modals & 8 distracters	(n=64) All participants	
	Untimed Grammaticality Judgment	(K=39) 18 Modals & 18 distracters		

8.7 Similarities to Tests Used in Experiment 1

The free oral picture description test was identical to the one used in the UK context. The gap fill was very similar with a couple of minor amendments (discussed below).

8.7.1 Free Oral Picture Description Test

As mentioned above the tasks in this test were the same as the ones in the UK. The procedure of test administration and scoring system in SA were the same as those used in the UK (discussed in 7.6.1). Detailed information is included in Appendices (D) and (I).

8.7.2 Gap Fill Test

This test was similar to the one in the UK except that three more items targeted modals (18, 21 and 22) and two more distracted items (14 and 19) were added. Same format of instructions as in the UK were used in SA. The scoring system was also the

same (discussed in section 7.6.3). More information on the test and scoring guidelines are included in Appendices (E) and (I).

8.8 Differences to Tests Used in Experiment 1

This section presents the tests that were different to the ones used in the UK experiment. Two new tests were introduced: an elicited imitation and metalinguistic knowledge questions. Another difference was the timed grammaticality judgement test was changed to untimed grammaticality judgment test in SA as the former included the provision of corrections which may tap into learners 'explicit knowledge although it was timed and learners were under time pressure.

8.8.1 Oral Elicited Imitation Test

The elicited imitation test was only undertaken in SA. It was developed following Erlam (2006, 2009) who argues that elicited imitation tests can represent implicit knowledge as awareness of grammatical rules is less likely to be used and that the correct imitation of language structure and, critically, the correction of incorrect structures during oral imitation tasks represent part of a learner's internalised (automatised) grammar (reviewed in section 2.6.1).

Two pilots of the elicited imitation were carried out. The first pilot involved six native participants and the second one involved nine non-native participants. Both pilots are discussed in the following sections.

Piloting elicited imitation test

Several factors were taken into consideration when designing the elicited imitation test, sentence complexity; the subject matter of the sentences; the inclusion of grammatically correct and incorrect items as it was hypothesized that learners acceptance or rejection of grammatical violations in spoken stimuli presented in real-time would be an indication of their internalization of targeted language structures (Erlam, 2006) and time between presentation of stimuli and repetition. The order of the test items was random for both the pilot and the main study.

Piloting EI test on native speakers

This test was first piloted on six native speakers to measure the difficulty of the test items. The design used in the pilot included eight training (practice) sentences, thirty sentences targeting modals and twelve distracters. The test was recorded by a native speaker of English. It was found that some sentences were skipped either for difficulty of repetition due to the ambiguity of some lexical items, unclear voice, or participants' lack of concentration.

The following comments were given by the native participants in response to the exit questionnaire included in Appendix C:

1- The statement on the top of the answer sheet needed to be changed to "Decide if the following statements are true, not true, or not sure."
2- Participants' background is very important in judging the sentence**.**
3- Participants might guess the answers in which case they would not be based on definite knowledge.
4- Asking participants to focus on meaning and form is a hard task to do because learners have to think of two things at the same time. They may correct the meaning and forget the form or correct the form and forget the meaning and in most cases they will do the form not the meaning, as the form of incorrect sentences would automatically be in focus.
5- Repeating part of the instruction, after each sentence in the first eight training sentences, is not necessary. It could be done at the beginning of the trial sentences and before the beginning of the actual testing items.
6- Participants might realize that the meaning of the sentence was not right after repeating it but there was no second chance.
7- It would be much easier to judge the sentence if it has been visualized.
8- The participants should be told that they need to correct the meaning and form of the sentences in case of considering the meaning as an objective of the study.

Piloting EI test on Non-native participants

This test was also piloted on non-native participants to find out the degree of difficulty and comprehensibility of the test from the students' point of view. Nine Saudi students were recruited from an English centre. In a quiet room, each participant was met individually. Each sentence was audio presented, one at a time, on a Microsoft Power Point.

A student listened to the instructions as well as the sentences. This was changed slightly for the main study, and students were allowed to listen and look at the script for the instructions only but not the test items.

In the main study, the number of training, distracters, and target sentences was increased based on the students' suggestions in response to the exit questionnaire included in Appendix C. A total of sixty belief statements distributed as follows: twelve training sentences, fifteen distracters and thirty-three sentences targeting English modals. The number of true, false and not sure sentences was taken into consideration to have equal distribution. This was true for all type of sentences, training, target, and distracter items as shown in Appendix C.

Non-native participants, in the pilot study found the test very hard to follow as some words were either not clear or difficult to listen to, understand and repeat the whole sentence back. They also suggested shorter sentences than longer ones. Thus the very long and difficult sentences were substituted with easier or shorter ones. The average length of syllables for all types of sentences, in the pilot study, was 12.5. This was reduced to an average of 11.73, in the main study.

Although learners wanted to see the sentences written this was not of course possible as one rationale for this test was to measure learners' oral implicit knowledge under time pressure. In the exit questionnaire, learners indicated the unfamiliarity of some items and thus suggested using items that are known to them.

There was only one version of the test, to maintain consistency between phases of the study. This might be a limitation to the study as discussed in Chapter 11.

Test details in the main study: Oral elicited imitation

Seven teacher assistants were selected to participate in conducting the elicited imitation test. Several meetings were held to train them on how to run the test and use the recording equipment and Laptops. Detailed instructions on the administration of the test are included in Appendix C.

The administration of the test was on a one-to-one basis. There were sixty belief statements. Sentences were randomly selected out of a hat and then quasi-randomized to assure that target items with the same target structure were not close to each other to reduce the likelihood that participants might become aware of the target structure. The statements were loosely organized around different themes (e.g., politics, driving, education, health and environment).

The test started with twelve training sentences, followed by forty-eight sentences of which thirty-three targeted modals, eighteen of which were grammatically correct and fifteen of which were grammatically incorrect sentences, and fifteen targeted different structures to distract participants' attention. There was no specific reason for the unequal number of grammatically correct and incorrect sentences in this test. Each sentence was programmed on a Microsoft PowerPoint slide after being recorded by a native speaker of English. Learners' production of each sentence was audio recorded.

Test takers were asked to take part in a "beliefs questionnaire" in which they would give their opinions about a range of topics. They were told that they would hear a statement and that they had to decide whether the statement was true or not true for them or whether they were not sure and that they were to tick their decision on the answer sheet. Focusing their attention on the meaning of each sentence is thought to reduce memory traces of the exact form of the sentence (Erlam, 2006, 2009). Participants were then required to repeat the sentence in correct English.

For the twelve training sentences learners were reminded to use their "best, most correct English". In the training phase only, participants were allowed to ask for any clarifications or a repetition for a sentence. The examiner was also allowed to give instructions or feedback concerning the student's answer at this training stage.

The sentences targeting modals were divided into eighteen grammatically correct and fifteen grammatically incorrect sentences focusing on ***can, will*** and ***must***. There was no specific rationale behind the unequal number of sentences. This could be supported by Erlam (2006) who had different number of grammatically correct (k=27) and incorrect (k=17) sentences in her EI test. During these items, participants were not allowed to ask for any elaboration and the examiner was not able to give any feedback.

Scoring

Distracters and training sentences were excluded. The accuracy of using modals was determined for each individual participant using 'suppliance in obligatory contexts scoring'. The numerator and the denominator therefore varied from one participant to another (the maximum potential denominator being thirty-three). If the target structure was supplied correctly, a score of one was given (even if the learner used a different lexical verb with the modal). A zero score was given if the student failed to supply a modal in a context where it should have been used or if a learner repeated the sentence but changed the structure so that a modal was no longer needed, then this was not counted as an 'obligatory context'. Inter-language scoring was used whereby if a learner produced an obligatory context and used a modal, but made an error such as no main verb, inflecting either the modal or the verb, inserting *be* or *do* or adding *to*, the answer was scored (0.5 mark). Initial productions were scored even if a learner self-corrected as this would provide a better measure of implicit knowledge (Ellis, 2007).

Learners' final scores were then converted to a percentage for each participant, using the formula below adapted from Ellis and Barkhuizen (2005). A copy of the test and the scoring guidelines with some genuine examples are included in Appendices (C) and (I).

$$\frac{N \quad correctly \ given \ in \ context}{Total \ obligatory \ context} \times 100 = percentage \ accuracy$$

8.8.2 Untimed Grammaticality Judgment Test

This was a pen and paper free time test, requiring learners to identify errors, provide corrections and give rules. The latter is henceforth referred to as metalinguistic questions (MQ) test. The items of the untimed grammaticality judgment test were identical to the ones used in the timed grammaticality judgment in the UK experiment, though the test was not timed. Another difference was that after each question, participants were asked to provide a rule if they reported that they had used a rule to change the sentence. The test was administered to all of the participants at one time in a large classroom.

Scoring

For untimed grammaticality judgment, the scoring used the number of target grammatical and ungrammatical items as the denominator (18). Distracters and training sentences were excluded. Learners' responses were scored as correct (1 mark), partially correct (0.5 marks) or incorrect (0 mark) as discussed in section 7.6.2.

For metalinguistic questions, the scoring used the number of target ungrammatical items as the denominator (9). If a learner stated rules strongly related to modals, the answer was scored (1 mark) even if lexical items were not spelled correctly. In case of stating part of the rules (e.g., because *ing*), the answer was scored (0.5 mark). Zero mark was given for stating different rules rather than the target (e.g., perfect tense). These proportions were then converted to a percentage for each participant. More information on scoring the MQ test is in Appendix I.

8.9 Questionnaires

8.9.1 Attitudinal questionnaire

The attitude questions and constructs were identical to the ones in the UK. The procedure of test administration and coding and scoring system in SA were the same as the ones used in the UK (discussed in 7.6.5). Detailed information is included in Appendix G. The attitudinal questionnaire was given on completion of posttests.

8.9.2 Exit questionnaire

The exit questionnaire was administered on completion of each oral and written pre-, post- and delayed posttest different to the one in the UK where it was only given after each written test. The questionnaire in SA include three questions unlike the one used in the UK. One similar question concerned learners' awareness of the target structure being tested but two more questions related to the provision of rules were added. One exit questionnaire was given on completion of both GF and untimed GJT tests due to time constrained. Please refer to Appendix G for more information.

Coding and Scoring of Exit Questionnaire

To code the exit questionnaire, same system of coding groups and participants was followed as the attitudinal questionnaire (see section 7.6.5 for more information). However, to code learners' response to the three open ended questions given at the end of each oral and written measure at the three testing time, zero was given for no response and one was given to the provision of a response.

8.10 Descriptive Statistics and Data Analysis

For all measures undertaken in SA, the mean scores and standard deviations were calculated for percentage scores from Pre-, post- and delayed posttests. A range of statistical tests were used to assess change over time and between conditions. Parametric or non-parametric tests were used based on the normality of distribution and baseline equivalence (discussed in detail in section 7.7).

This section encompasses the results of test of normality and baseline on all outcome measures, and the descriptive analysis and discussion relative to the first and third research questions (see Chap 6).

Normality of data for all measures

The results of the Shapiro-Wilk test are presented in Appendix J. The test of normality indicated that the assumption of normality was violated in the data from SA context in elicited imitation (correct and incorrect items), free oral picture description test, gap fill test, grammatically correct items and grammatically incorrect items of

untimed grammaticality judgment test, and finally metalinguistic questions test, thus non-parametric tests were carried out for those measures. The data in the overall untimed grammaticality judgment was normally distributed, thus parametric tests were carried out.

Baseline parity

The results of ANOVA suggested similar baseline across different groups for the overall untimed GJT test F (2,63)=0.25, p=0.78. The results of Kruskal-Wallis suggested similar baselines for EI test (H (2) = 1.72, p= 0.42), grammatically correct items in EI test (H (2) = 1.12, p = 0.57), grammatically incorrect items in EI (H(2) = 5.20, p = 0.10), free oral picture description test (H(2) = 2.45, p= 0.29), gap fill test (H(2) = 0.88, p= 0.64), the grammatically correct items in untimed GJT test(H(2) = 0.94, p = 0.63), the grammatically incorrect items in untimed GJT test (H(2) = 1.48, p = 0.48), and metalinguistic questions test (H(2) = 2.87, p = 0.24). For all these measures, the actual scores were analysed, rather than gain scores. Please refer to Appendix J for more details.

The data of this phase of the study will be presented in tabular and graphical form and analysed and discussed in the following order: oral elicited imitation test (with grammatically correct and incorrect sentences presented separately), the free oral picture description test, gap fill, untimed grammaticality judgment test (with grammatically correct and incorrect sentences presented separately), and metalinguistic questions test.

Test reliability

The reliability of the scoring of oral EI and PD tests, and untimed GJT, MQ and GF written measurements was carried out. Twenty five percent of the data for each measure was scored by native and native-like speakers of English. The results shown in the table below were submitted to Cronbach's alpha coefficients. This was considered sufficiently high for the researcher to independently code the remainder of the data.

Outcome measures in SA	Inter-rater reliability (r)
EI	**0.96**
PD	**0.95**
GJT	**0.98**
MQ	**0.97**
GF	**0.99**

8.10.1 Oral Elicited Imitation Test

Descriptive results on overall scores

The mean scores and standard deviations for the total overall elicited imitation scores (obligatory occasions) are shown in Table 8.2 and presented graphically in Figure 8.2.

Table 8.2 Mean scores on elicited imitation test (EI)

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	68.44	23.68	87.91	10.80	75.70	30.31
Recast	25	60.08	25.52	82.94	11.91	77.62	25.15
Task only	20	67.90	23.05	77.66	22.29	83.80	14.26

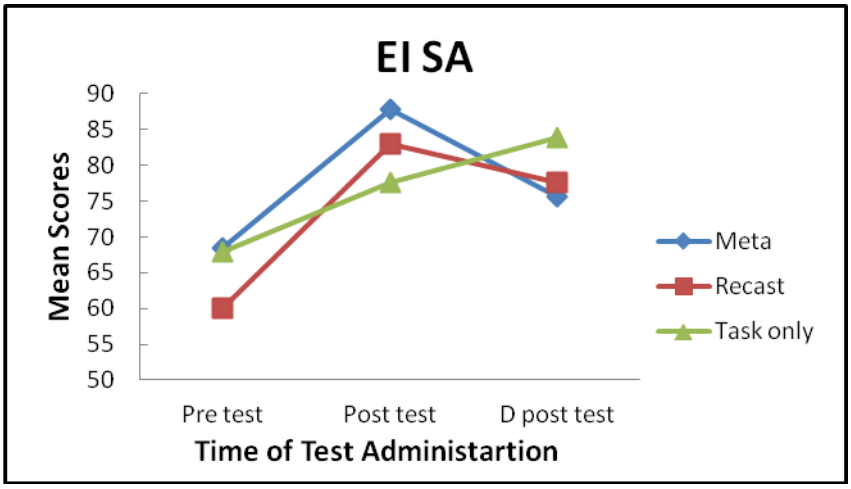


Figure 8.2 Mean scores on elicited imitation test

Analysis

The results of the Friedman test indicated a significant difference between testing times for the metalinguistic information group, χ^2 (2) = 14.14, p <0.01, the recast group, χ^2 (2) = 27.75, p <0.01and a significant difference for the task only group, χ^2 (2) =13.00, p <0.01.

For the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre- and posttests (Z = -3.46, p <0.01), and between pre- and delayed posttests (Z = -2.37, p = 0.01), but not between post- and delayed posttests (Z = -0.82, p = 0.21).

For the recast group there was a statistically significant difference between pre- and posttest, (Z = -4.16, p <0.01) and between pre- and delayed posttests, (Z = -3.44, p <0.01), but not between post- and delayed posttests, (Z = -0.21, p = 0.42).

For the task only group there was a statistically significant difference between pre- and posttest (Z =-2.76, p <0.01), a significant difference between pre- and delayed posttests (Z = -2.77, p <0.01) and between post- and delayed posttests (Z = -2.09, p = 0.02).

To find out whether there was a significant difference between the three groups at the time of post-and delayed posttests, the Kruskal-Wallis test showed no significant differences between the three groups at posttest, H (2) =3.16,p = 0.21 nor at delayed posttest, H (2) = 0.77, p = 0.68. Detailed information is included in Appendix K (Table K.6).

Summary

Results suggested that the provision of CF via recasts and metalinguistic information, and the interactional tasks alone helped the development of English modals.

The results also indicated a significant gain for the metalinguistic information and recast groups in the short-term that was maintained at delayed posttests. The task only group continued to gain significantly after the intervention had finished.

However, when groups were compared, the results indicated that the three intervention types did not seem to lead to significant differences between them.

8.10.2 Correct Items in Elicited Imitation Test

Descriptive results

The mean scores and the standard deviations for correctly repeating back the correct items in the elicited imitation test presented in Table 8.3 and in Figure 8.3.

Table 8.3 Mean scores on correct items in EI test

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	65.34	36.08	92.85	10.35	87.45	26.04
Recast	25	68.38	31.80	81.76	24.12	84.97	27.73
Task only	20	72.74	33.42	84.06	24.97	91.20	13.39

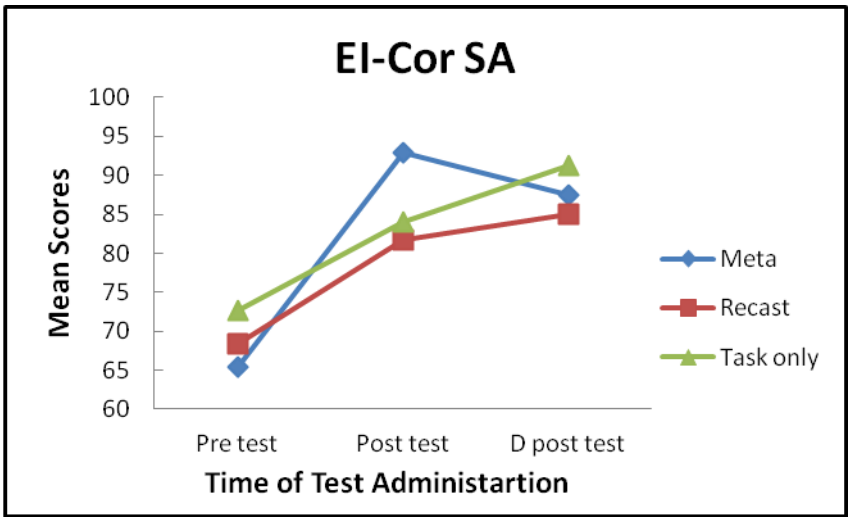


Figure 8.3 Mean scores on correct items in EI test

Analysis

The results of the Friedman test indicated a significant difference between testing times for the metalinguistic information group, χ^2 (2) = 15.13, p <0.01, recast group, χ^2 (2) = 9.29, p = 0.01, and a significant difference for the task only group, χ^2 (2) = 10.83, p <0.01.

For the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre- and posttests (Z = -3.21, p = 0.0<0.010), and between pre- and delayed posttests (Z = -2.77, p <0.01), but not between post- and delayed posttests (Z = -0.27, p = 0.41).

For the recast group there was a trend to statistically significance difference between pre- and posttests (Z = -1.87, p = 0.06) and between post- and delayed posttest (Z = -1.59, p = 0.06). A statistically significant difference was found between pre- and delayed posttests (Z = -2.47, p = 0.01).

For the task only group the Wilcoxon test showed a statistically significant difference between pre- and posttest (Z = -2.30, p = 0.02), and between pre- and delayed posttests (Z = -2.42, p = 0.01), but not between post- and delayed posttests (Z = -1.26, p = 0.11).

To find out whether there was a significant difference between the three groups at the time of post-and delayed posttests, the Kruskal-Wallis test showed no significant differences between the three groups at posttest, H(2) = 3.42, p = 0.18 nor at delayed posttest, H(2) = 0.26, p = 0.88. Detailed information is included in Appendix K (Table K.7).

Summary

Results suggested that the provision of CF via recasts and metalinguistic information and the interactional tasks alone helped the development of English modals.

The results also indicated a significant gain, for the metalinguistic information and the task only groups on short-term that was maintained at delayed posttest, but no further gain was made once the intervention finished. On the other hand, the recast group showed significant gains after the post test.

However, when groups were compared in relation to each other, the results indicated that the three intervention types did not seem to lead to different scores.

8.10.3 Incorrect Items in Elicited Imitation Test

Descriptive results

The mean scores and the standard deviations for correctly changing the incorrect items in the elicited imitation test are shown in Table 8.4 and presented graphically in Figure 8.4.

Table 8.4 Mean scores on incorrect items in EI test

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	60.66	23.06	82.61	14.01	69.05	29.75
Recast	25	46.45	24.35	77.15	17.51	70.73	20.35
Task only	20	61.15	21.37	67.84	27.35	74.02	23.89

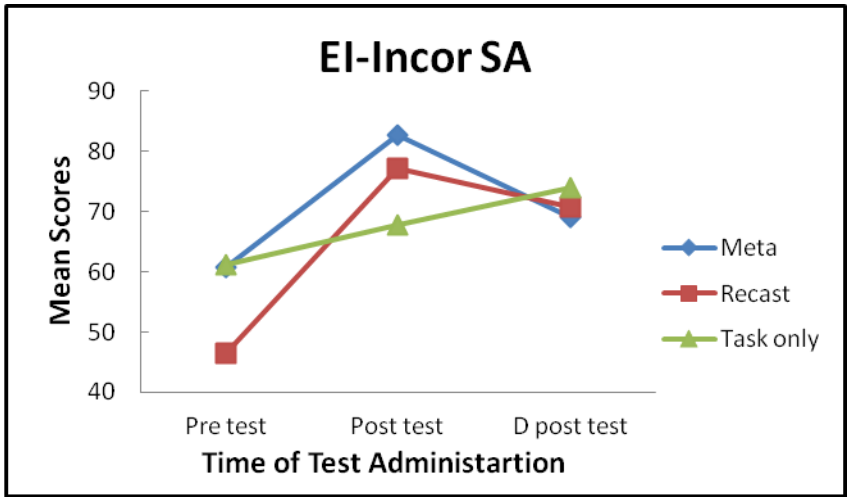


Figure 8.4 Mean scores on incorrect items in EI test

Analysis

The results of the Friedman test indicated a significant difference between testing times for the metalinguistic group, χ^2 (2) = 9.12, p = 0.01, the recast group, χ^2 (2) = 29.46, p = 0.01, and a significant difference for the task only group, χ^2 (2) = 6.63, p = 0.04.

For the metalinguistic group, the Wilcoxon test showed a statistically significant difference between pre- and posttest (Z = -3.44, p <0.01). There was a borderline statistically significant difference between pre- and delayed posttest (Z =-1.78, p = 0.08), but no significant difference between post- and delayed posttest (Z =-1.25, p = 0.12).

For the recast group there was a significant difference between pre- and posttest (Z =-4.19, p <0.01) and between pre- and delayed posttest (Z =-3.86, p <0.01). In addition, a tendency to significant difference was found between post- and delayed posttest (Z = -1.42, p = 0.08).

For the task only group, a statistically significant difference was found between pre- and delayed posttest (Z = -1.94, p = 0.05) but not between the pre and posttest (Z = -1.55, p = 0.13) or between the post and delayed posttest (Z = -1.33, p = 0.10).

To find out whether there is a significant difference between the three groups at the time of post-and delayed posttests, the Kruskal-Wallis test indicated no significant difference between the three groups at posttest H(2) =3.69, p = 0.16 nor at delayed posttest H(2)= 0.67, p = 0.72. Detailed information is included in Appendix K (Table K.8).

Summary

Results suggested that the provision of CF via metalinguistic information and recasts and the interactional tasks alone helped the development of English modals.

The results also indicated a significant gain for the metalinguistic information and recast groups immediately following the intervention which was maintained at delayed posttest, but the recast group also continued to show some gains once the intervention had finished. The task only group indicated a significant gain at delayed posttest only.

However, when groups were compared, the results indicated that the three intervention types did not seem to lead to different scores. This result seems to contradict the findings that the experimental groups (metalinguistic information and recast) made gains straight after the intervention but the task only group did not.

8.10.4 Free Oral Picture Description Test

Descriptive results

The mean scores and standard deviations are displayed in Table 8.5 and presented graphically in Figure 8.5.

Table 8.5 Mean scores on free oral picture description test (PD)

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	82.00	29.63	91.98	11.59	87.60	22.74
Recast	25	70.21	37.13	88.21	14.26	85.22	26.96
Task only	20	86.14	23.23	90.98	11.02	93.77	9.26

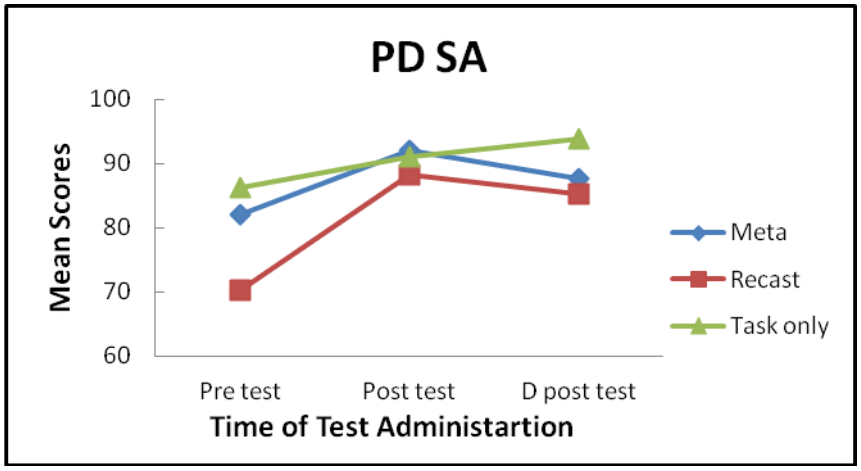


Figure 8.5 Mean scores on PD test

Analysis

The results of the Friedman test indicated no statistical significant difference between testing times for the metalinguistic information group, χ^2 (2) =4.62, p = 0.10 and the recast group, χ^2 (2) = 4.85, p = 0.09, nor for task only group, χ^2 (2) = 2.69, p = 0.26.

To find out if there was a statistical significant difference between the three groups at the time of post- and delayed posttests, the Kruskal-Wallis test showed no significant differences between the three groups at posttest, H(2) =1.30, p = 0.52 nor at delayed posttest, H(2) =1.45, p = 0.49. For more information, please refer to Appendix K (Table K.9).

Summary

Drawing on the results, it seems that the provision of CF via metalinguistic information and recast and the interactional tasks alone did not help learners' development of English modals and did not lead to different scores.

8.10.5 Gap Fill Test

Descriptive results

The mean scores and the standard deviations on gap fill (k=14) items for the three groups are illustrated in Table 8.6 and presented graphically in Figure 8.6.

Table 8.6 Mean scores on gap fill test (GF)

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	19.84	18.79	38.49	26.54	38.10	29.81
Recast	25	20.71	20.59	31.57	24.28	30.00	24.33
Task only	20	29.11	27.77	41.61	30.73	35.54	30.62

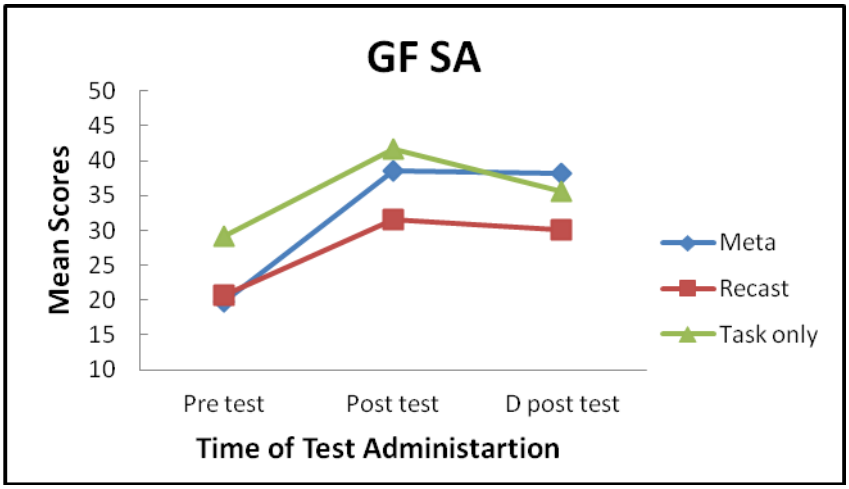


Figure 8.6 Mean scores on GF test

Analysis

The results of the Friedman test indicated a statistically significant difference between testing times for the metalinguistic information group, χ^2 (2) = 16.13, p <0.01, the recast group, χ^2 (2) =14.38, p <0.01, and a significant difference for task only group, χ^2 (2) = 12.57, p <0.01.

For metalinguistic information group, a Wilcoxon test showed a statistically significant difference between pre-and posttest (Z = -3.34, p <0.01) and between pre- and delayed posttest, (Z = -3.21, p <0.01), but no significant difference between post-and delayed posttest, (Z = -0.13, p = 0.46).

For the recast group, the results indicated a statistically significant difference between pre-and posttest, (Z = -3.62, p <0.01), and between pre- and delayed posttest (Z = -2.60, p <0.01), but there was no significant difference between post- and delayed posttest (Z = -0.60, p = 0.29).

For task only group, there was a statistically significant difference between pre- and posttest (Z = -2.95, p <0.01), between pre- and delayed posttest (Z = -2.05, p = 0.04). In addition, a tendency towards a significant gain was shown between post-and delayed posttest (Z = -1.61, p = 0.06).

To find out whether there was a significant difference between the three groups at the time of post- and delayed posttests, the Kruskal-Wallis test showed no statistically

significant differences between the three groups at posttest, H (2) =1.09, p = 0.58 nor at delayed posttest, H (2) = 0.72, p = 0.70. Detailed information is included in Appendix K (Table K.10).

Summary

Results suggested that the provision of CF via metalinguistic information, and recast, and interactional tasks alone helped the development of English modals.

The results also indicated a significant gain for the metalinguistic information group, recast group and task only group in the short-term that was maintained at delayed posttest, and the results suggested that the task-only group may have continued to make gains once the intervention had finished.

When groups were compared, the results indicated that the three intervention types did not seem to lead to different scores at post or at delayed posttest.

8.10.6 Untimed Grammaticality Judgment Test

Descriptive results on overall scores

The accuracy of the mean scores and the standard deviations for the total untimed GJT scores (k=18) provided in Table 8.7 and presented graphically in Figure 8.7.

Table 8.7 Mean scores on untimed grammaticality judgment test (UGJT)

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	46.93	15.13	63.74	19.32	49.38	20.11
Recast	25	45.67	15.17	55.89	16.16	54.56	19.09
Task only	20	49.03	17.62	57.08	19.86	60.00	21.33

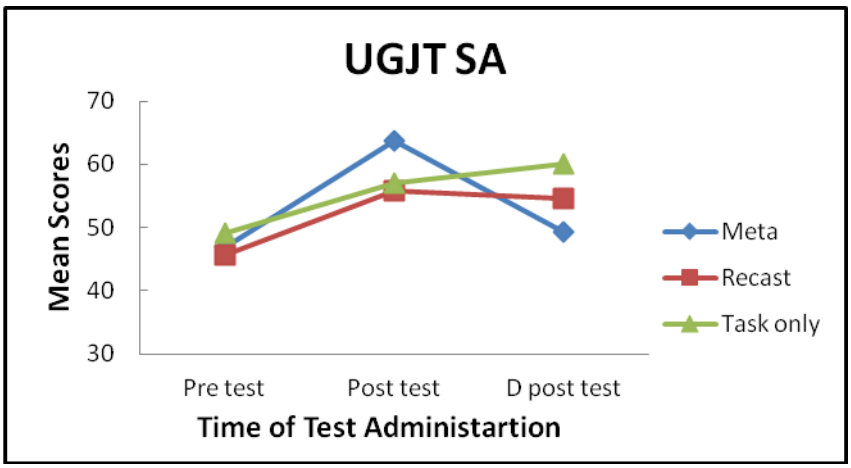


Figure 8.7 Mean scores on UGJT

Analysis

The results of mixed design ANOVA indicated a significant difference between testing times, F (2, 120) =19.00, p <0.01, but not between the treatment groups, F (2, 60) = 0.24, p = 0.79. There was a significant interaction effect between tests and treatment groups, F (4,120) = 3.41, p = 0.01.

The tests of within-subject contrasts suggested a significant difference between pre- and posttest, F (1, 60) = 38.49, p <0.01, between pre-and delayed posttest, F (1, 60) = 15.86, p <0.01, and between post- and delayed posttest, F (1, 60) = 4.50, p = 0.04. For more information, please refer to Appendix K (Table K.11).

The interaction between the test time and intervention type is likely to be due to the significant drop in scores between the post and delayed posttest by the metalinguistic group, compared to the recast and task-only group that maintained their scores at delayed posttest as indicated in the tests of within-subjects contrasts F (2, 60) = 6.23, p <0.01.

Summary

The results suggested beneficial role for all three conditions: interactional tasks alone, CF via metalinguistic information and CF via recasts. The overall scores improved significantly on all testing times for the implicit groups (recast and task only) and these gains were maintained at delayed posttest except for the metalinguistic information group which evident significant drop in the period from post to delayed post.

8.10.7 Correct Items in Untimed Grammaticality Judgment Test

Descriptive results

The mean scores on the correct items (k=9) and the standard deviation for all groups are given in Table 8.8 and Figure 8.8.

Table 8.8 Mean scores on correct items in UGJT test

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	69.59	22.48	75.44	20.81	66.67	15.71
Recast	25	64.89	17.77	76.44	19.59	76.00	24.57
Task only	20	65.00	18.48	72.78	21.77	81.11	21.66

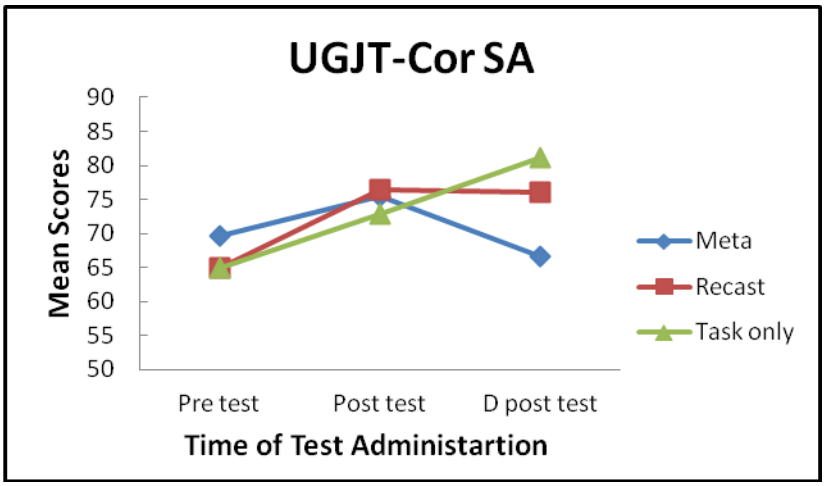


Figure 8.8 Mean scores on correct items in UGJT test

Analysis

The results of the Friedman test indicated no statistically significant difference between testing times for the metalinguistic information group, $\chi 2$ (2) = 2.84, p= 0.24. For the recast group, there was a significant difference between testing times, χ^2 (2) =

11.45, p <0.01, and a significant difference for the task only group, χ^2 (2) = 8.22, p = 0.02.

For the recast group the Wilcoxon test indicated a statistically significant difference between pre- and posttest, (Z = -2.18, p = 0.03), and between pre- and delayed posttest, (Z = -2.47, p = 0.01); however, there was no significant difference between post- and delayed posttest, (Z = -0.48, p = 0.33).

For the task only group, the Wilcoxon test showed a significant difference between pre-and delayed posttest, (Z = -2.63, p <0.01), and a significant difference between post- and delayed posttest, (Z = -1.69, p = 0.05), but no significant difference between post- and pre-test, (Z = -1.59, p = 0.12).

To find out whether there was a significant difference between groups at the time of post-and delayed posttests, Kruskal-Wallis indicated no statistically significant difference between groups at posttest, H (2) = 0.35, p= 0.84, but there was a significant difference between groups at the delayed posttest, H (2) = 7.14, p = 0.03.

Following that Kruskall-Wallis test, the results of Mann-Whitney indicated a significant difference between the metalinguistic information group and task only group at the delayed posttest, (Z = -2.44, p <0.01) in favour of task only group, but there was no significant difference between the recast group and task only group at the delayed post-testing time, (Z = -0.82, p = 0.21). However, there was a significant difference between the two experimental groups (metalinguistic information vs. recast) at delayed posttest, (Z = -2.11, p = 0.02) in favour of recast group. More information is in Appendix K (Table K.12).

Summary

The results suggested that the recast group and task only group may have made some gains in knowledge as measured by correct items on an untimed GJT test after the intervention finished as they significantly improved their test scores at delayed posttest.

For the recast group, the results indicated significant gains that were maintained at delayed posttest.

For the task only group, the results showed no significant gains in the short-term, but significant gains were made in the longer-term once the intervention had finished.

There was no significant difference between recast group and task only group at the delayed post-testing times, but there was a significant difference between recast and metalinguistic information groups in the direction of recast and between the metalinguistic and task only groups at delayed post-testing time in the direction of task only group.

8.10.8 Incorrect Items in Untimed Grammaticality Judgment Test

Descriptive results

The mean scores and the standard deviation for correctly changing the incorrect items (k=9) in the untimed grammaticality judgment test for all groups are seen in Table 8.9 and in Figure 8.9.

Table 8.9 Mean scores on incorrect items in UGJT test

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	25.62	24.93	51.85	25.28	32.10	32.09
Recast	25	26.44	18.65	35.33	19.44	33.11	22.93
Task only	20	33.06	26.70	41.39	28.70	38.89	31.74

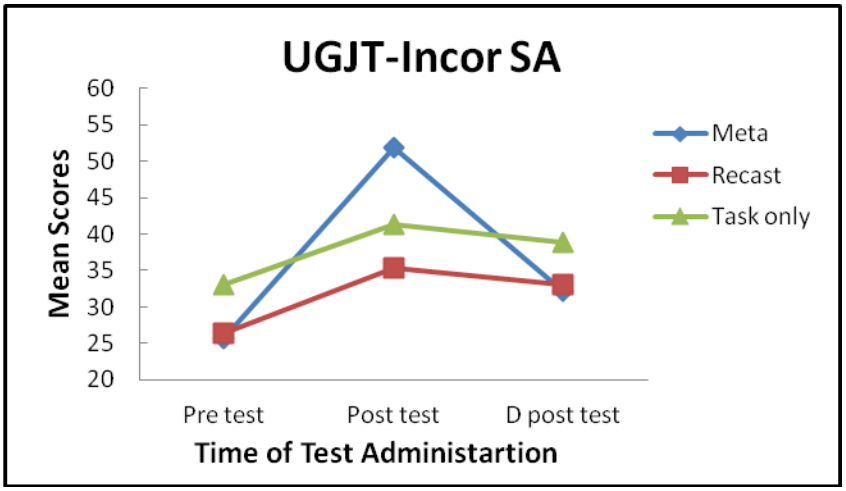


Figure 8.9 Mean scores on incorrect items in UGJT test

Analysis

The results of the Friedman test indicated significant differences between the different testing times for the metalinguistic information group, $\chi 2$ (2) =15.60, p <0.01and for the recast group, χ^2 (2) =6.27, p = 0.04, but there was no significant difference on testing times for the task only group, χ^2 (2) =3.03, p = 0.22.

For the metalinguistic information group, the Wilcoxon test showed a significant difference between pre-and posttest (Z = -3.63, p <0.01) and between post- and delayed posttest (Z = -2.64, p <0.01), but not between pre- and delayed posttest (Z = -1.07, p = 0.31).

For the recast group, there was a significant difference between pre- and posttest, (Z = -2.11, p= 0.04). However, there was no significant difference between pre-and delayed posttests (Z = -1.30, p = 0.19) or between post- and delayed posttest (Z = -0.50, p = 0.32).

To find out whether there is a significant difference between groups at the time of post-and delayed posttests, Kruskal-Wallis test showed a trend to a significant difference at post-testing time H(2)= 5.56, p = 0.06, and the graph above suggests that this was due to the higher score of the metalinguistic information group compared to the other two groups. However, there was no statistical difference between the groups at delayed post-testing time H (2) = 0.75, p = 0.69. Please refer to Appendix K (Table K.13).

Summary

The results on the incorrect items of the grammaticality judgment test suggested that the provision of CF via metalinguistic information and recasts assisted learners to develop their learning of English modals in the short term, but the oral tasks alone did not lead to learning.

The results indicated a significant gain for the metalinguistic information group in short-term only. The gains made between pre- and post-test were lost at delayed post-test.

However, when groups were compared, it was found that there was a trend to significant difference between groups at post testing time, but there were no differences evident after the delayed posttest.

8.10.9 Metalinguistic Questions Test

Descriptive results

The mean scores and standard deviation for providing rules on the metalinguistic question items (k=9) are presented in Table 8.10 and graphically in Figure 8.10.

Table 8.10 Mean scores on metalinguistic questions test (MQ)

Treatment group	N	Pre-test		Post-test		D post-test	
		M	SD	M	SD	M	SD
Metalinguistic	19	11.99	16.17	25.15	29.18	21.60	28.38
Recast	25	5.56	8.18	16.67	13.89	11.33	13.08
Task only	20	13.89	18.07	23.06	24.00	16.94	21.74

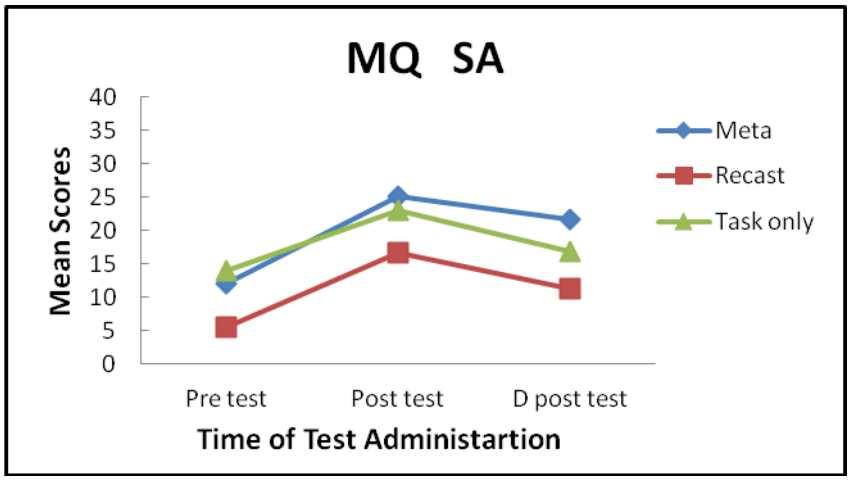


Figure 8.10 Mean scores on MQ test

Analysis

The results of the Friedman test indicated a significant difference between testing times for the metalinguistic information group χ^2 (2) = 7.14, p = 0.03, for the recast group, χ^2 (2) = 11.04, p <0.01, and for the task only group, χ^2 (2) = 7.17, p = 0.03.

For the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre-and post-test (Z = -2.67, p <0.01). However, there was no statistical difference between pre- and delayed post-test (Z = -1.65, p = 0.11), or between post- and delayed post-test (Z = -1.03, p = 0.17).

For the recast group, the results of Wilcoxon test indicated a significant difference between pre- and post-test (Z = -3.52, p <0.01), but there was no significant difference between pre- and delayed post-test (Z = -1.59, p = 0.12). There was a tendency to significant difference between post- and delayed post-test, due to the slight decrease in scores (Z = -1.55, p = 0.07).

For the task only group, there was a significant difference between pre- and post-test (Z = -2.69, p <0.01), but no significant difference between pre- and delayed post-test (Z = -0.84, p = 0.44). There was a trend toward significant difference between post- and delayed post-test (Z = -1.54, p = 0.07).

To find out whether there was a significant difference between the three groups at the time of post-test and delayed post-test, Kruskal-Wallis indicated no statistically significant differences between the three groups at posttest, H (2) = 0.35, p = 0.84 nor at the delayed posttest, H (2) = 0.63, p = 0.73. Please refer to Appendix K (Table K.14).

Summary

Results suggested that the provision of CF via recast and metalinguistic information and the oral tasks alone helped in the development of knowledge about English modals in the short term but none of these were maintained at delayed posttest.

When groups were compared, the results of Kruskal-Wallis suggested that the three intervention types did not seem to lead to different scores at post- or delayed post-tests.

Provision of correction and stating rules in the untimed GJT incorrect items

In the current study, it was found that the three groups, on the grammatically incorrect items (k=9) of untimed GJT test at all testing periods were able to correct the incorrect items more than stating rules that have been violated as shown in Tables 8.9 and 8.10.

The results in Table 8.11 demonstrate the different percentages in correcting items and stating rules at the different testing times. The results of the ungrammatical items in the untimed GJT test in the current study suggested that learner's ability to correct the ungrammatical items exceeded their ability to state rules. Learners in the metalinguistic information group provided the relevant rules for about 50% of the number of items they could actually produce corrections on all testing times but learners in the recast and task only groups provided rules for less than 50% of the number of items they could produce corrections on all testing times.

Table 8.11 Percentage of correcting errors and stating rules for incorrect items in untimed GJT

Group	Stating rules (K=9)			Correcting the incorrect items (K= 9)		
	Pre test	Post test	Delayed post test	Pre test	Post test	Delayed post test
MI	12.65	26.54	21.60	25.62	51.85	32.10
R	5.56	16.67	11.33	26.44	35.33	33.11
TO	13.89	23.06	16.94	33.06	41.39	38.89

8.10.10 Attitudinal Questionnaire

This section presents descriptive statistics for three constructs: learners' opinions about the content of the intervention activities; learners' opinions about error correction and accuracy generally; learners' opinions about the different CF techniques used in the intervention. In addition, the results of Pearson product-moment correlation (r values) analysis about the relationship between attitudes and post test scores.

Learners' opinions about the intervention activities

Table 8.12 shows the descriptive statistics for the scores of learners' opinions about the intervention activities. Learners' attitude was measured on a Likert scale (1-5). The average scores for learners' opinions towards the intervention activities ranged from 3.52 to 3.62 with the recast group achieving the lowest score (representing less positive opinions) and task only group attaining the highest (representing more positive opinions). However, a one-way ANOVA test indicated no significant differences between these groups F (2, 63) = 0.24, p = 0.79.

Table 8.12 Learners' opinions about the intervention activities

Group	N	Mean (k=5)	SD	Min	Max
Metalinguistic	19	3.56	0.61	2.20	4.60
Recast	25	3.52	0.41	2.60	4.40
Task only	20	3.62	0.44	2.80	4.40
Total	64	3.56	0.48	2.20	4.40

(On a 1-5 Likert scale)

Opinions about error correction generally

The descriptive statistics for the scores of learners' opinions about error correction generally indicated that the average scores for learners' opinions towards error correction generally ranged from 3.49 to 3.59 with the task only group achieving the lowest score (representing a less positive attitude towards error correction) and the metalinguistic group gaining the highest (representing a more positive attitude towards error correction), as illustrated in Table 8.13. However, a one-way ANOVA test indicated no significant differences between these groups F (2, 63) = 0.26, p=0.78.

Table 8.13 Opinions about error correction generally

Group	N	Mean (k=8)	SD	Min	Max
Metalinguistic	19	3.59	0.42	2.88	4.38
Recast	25	3.58	0.63	2.38	4.63
Task only	20	3.49	0.41	2.50	4.13
Total	64	3.55	0.50	2.38	4.63

(On a 1-5 Likert scale)

Opinions about corrective feedback during intervention

Table 8.14 illustrates the descriptive statistics for the scores of learners' opinions about corrective feedback during intervention. The average scores for learners' opinions towards the different CF techniques ranged from 3.49 to 3.59 with the task only group achieving the lowest scores (representing a less positive attitude towards corrective feedback) and the metalinguistic group gaining the highest scores (representing a more positive attitude towards corrective feedback). However, a one-way ANOVA test indicated no significant difference between groups, F (2, 63) = 2.32, p = 0.13.

Table 8.14 Opinions about corrective feedback during intervention

Group	N	Mean (k=8)	SD	Min	Max
Metalinguistic	19	4.13	0.42	3.25	4.88
Recast	25	3.93	0.45	2.63	4.63
Task only	20	3.86	0.33	3.13	4.50
Total	64	3.97	0.42	2.63	4.88

(On a 1-5 Likert scale)

Relationship between attitudes and test scores

This section provides analyses of correlation between the test scores and the results of the attitude questions. Correlations are only provided using the post test scores as the attitude questionnaire was not carried out at pre- or delayed post-test.

***Relationship between attitudes and post test scores on* elicited imitation**

The results of a Pearson correlation analysis, presented in Table 8.15 revealed the results for the three different sets of questions (content of activities, error correction generally and the CF during the intervention) and different groups.

For the metalinguistic information, the recast and task only groups, the results indicated no significant association between elicited imitation mean post test scores and the scores of the three different attitude questions as shown in the table below.

Table 8.15 Correlations between the mean scores on elicited imitation and learners' attitudinal scores

Group	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 19)	Post EI	-0.03	-0.27	-0.24
Recast (n = 25)		-0.01	-0.23	0.20
Task only (n = 20)		0.21	-0.18	0.37

Relationship between attitudes and post test scores on free oral picture description

The results of a Pearson correlation analyses shown in Table 8.16 reveal the results for the three different sets of questions (content of activities, error correction generally, and CF in the intervention) and different groups.

For the metalinguistic information group, the free oral picture description mean post test scores were significantly correlated with the 'opinions towards the intervention activities' with a Pearson correlation coefficient of r = 0.49 and the significance value is less than 0.05 but not with the 'general opinions towards error correction' or the 'CF techniques during the intervention'.

For the recast and task only groups, there was no significant relationship between the free oral picture description mean post test scores and any of the three attitude constructs.

Table 8.16 Correlations between the mean scores on free oral picture description and learners' attitudinal scores

Group	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 19)	Post PD	0.49*	0.05	-0.32
Recast (n = 25)		0.24	-0.05	-0.22
Task only (n = 20)		0.21	-0.15	0.05
*** $p < 0.05$**				

Relationship between attitudes and post test scores on gap fill

The Pearson correlation analyses presented in Table 8.17 revealed the results for the three different sets of questions (content of activities, error correction generally, and CF in the intervention) and different groups.

For the metalinguistic information and recast groups, there was no significant relationship between the gap fill mean post test scores and any of the three attitude constructs.

For the task only group, a highly significant positive correlation was found between the gap fill mean post test scores and the 'opinions towards the intervention activities' with a Pearson correlation coefficient of $r = 0.61$ and $p < 0.01$ but not with the 'general opinions towards error correction' or the 'CF techniques during the intervention'.

Table 8.17 Correlations between the mean scores on gap fill and learners' attitudinal scores

Group	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 19)	Post GF	0.10	-0.12	0.04
Recast (n = 25)		0.07	-0.23	0.07
Task only (n = 20)		0.61**	0.11	0.20
** $p < 0.01$				

Relationship between attitudes and post test scores on untimed GJT

The Pearson correlation analyses presented in Table 8.18 revealed the results for the three different sets of attitude questions (content of activities, error correction generally, and CF in the intervention) and different groups.

For the metalinguistic information and task only groups, there was no significant relationship between the untimed GJT mean post test scores and any of the three attitudes constructs.

For the recast group, a significant positive correlation was found between the untimed GJT mean post test scores and the opinions towards the 'intervention activities' with a Pearson correlation coefficient of $r = 0.40$, and $p < 0.05$ but not with the 'general opinions towards error correction' or the 'CF techniques during the intervention'.

Table 8.18 Correlations between the mean scores on untimed GJT and learners' attitudinal scores

Groups	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 19)	Post UGJT	-0.12	-0.25	0.01
Recast (n = 25)		0.40*	-0.29	-0.01
Task only (n = 20)		0.41	-0.03	-0.16
*** $p < 0.05$**				

Relationship between attitudes and post test scores on metalinguistic questions

The results of a Pearson correlation analyses presented in Table 8.19 revealed the results for the three different sets of attitude questions (content of activities, error correction generally, and CF in the intervention) and different groups.

For the metalinguistic information, the recast and task only groups, the Pearson correlation coefficient (r values) indicated no significant association between the metalinguistic questions mean post test scores and any of the three attitudes constructs.

Table 8.19 Correlations between the mean scores on metalinguistic questions and learners' attitudinal scores

Group	Attitudinal scores			
	Test	Activities	EC general	CF intervention
Metalinguistic (n = 19)	Post MQ	0.14	-0.11	0.06
Recast (n = 25)		0.36	-0.36	0.09
Task only (n = 20)		0.36	-0.06	-0.04

8.10.11 Exit Questionnaire

Learners were given, on completion of each test, at the three testing sessions, an exit questionnaire asking three questions relating to their awareness about the target structure of the tests. As one exit questionnaire was given on completion of both written gap fill and untimed GJT tests, the results of the questionnaire relate to both these tests were put together.

For free oral picture description and elicited imitation tests, a questionnaire was given on completion of each measure separately.

Test of normality was carried out for the data from PD, EI and both GF and untimed GJT, Shapiro-Wilk indicated that the distribution was not normal for all data. To find out if the baseline was the same on all measures, Kruskal-Wallis Test was carried out. It was found that participants started with similar baseline in PD, χ^2 (2) = 0.00, p = 1.00, EI, χ^2 (2) = 0.00, p = 1.00 and both GF and untimed GJT, χ^2 (2) = 1.36, p = 0.51. Based on these results, non-parametric tests were carried out for data analyses.

Chi-square analysis and results of the exit questionnaire data on each measure will be presented in the following sections.

Exit questionnaire on free oral picture description test

The students' responses to the exit questionnaire in free oral picture description test were cross tabulated and the results are presented in Table 8.20. Figure 8.11 shows the percentage of learners' responses to the exit questionnaire in the different testing times.

Table 8.20 Students' responses to the exit questionnaire on PD

Group	N	Group Cross Tabulation in PD					
		Pre score		Post score		Delayed post score	
		0	**1**	**0**	**1**	**0**	**1**
Metalinguistic	19	19	00	12	7	12	7
Recast	25	25	00	19	6	16	9
Task Only	20	20	00	19	1	16	4
Total	64	64	00	50	14	44	20
Definition: 0= no response 1= correct response							

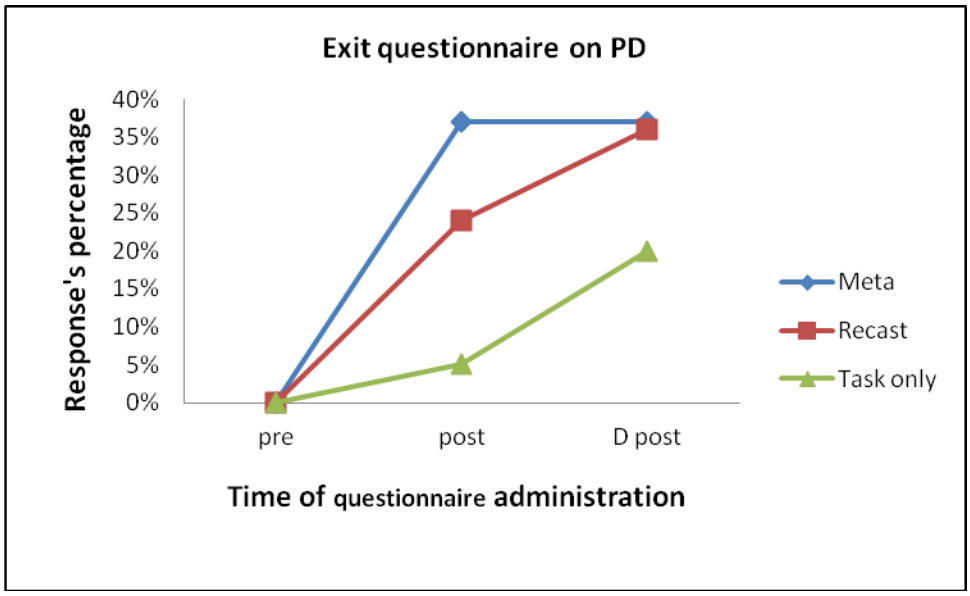


Figure 8.11 Percentage of students' responses to the exit questionnaire on PD

Analysis

Given that the dependent variable in the exit questionnaire was categorical, Chi square statistical test was run on the questionnaire given on completion of **free oral picture description** test on the three testing sessions. The results for the three groups

indicated no scores on learners 'responses thus statistics were not computed on pre exit questionnaire. The questionnaire on post free oral PD test indicated a borderline significant difference χ^2 (2) = 5.89, p = 0.05 whereas, the result of the questionnaire on delayed posttest indicated no significant difference, χ^2 (2) = 1.72, p = 0.42.

The results of the Friedman test for each individual group indicated a significant difference between testing times for the metalinguistic information group, χ^2 (2) = 10.89, p <0.01, the recast group, χ^2 (2) = 12.60, p <0.01and for the task only group, χ^2 (2) = 6,500, p = 0.04.

For the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre- and post-tests (Z = -2.65, p <0.01), and between pre- and delayed post-tests (Z = -2.65, p <0.01) but not between post- and delayed post-tests Z = -0.00, p = 1.00).

For the recast group, the Wilcoxon test showed a statistically significant difference between pre- and post-tests (Z = -2.45, p = 0.01) and between pre-and delayed post-tests (Z = -3.00, p <0.01) but not between post- and delayed post-tests (Z = -1.34, p = 0.18).

For the task only group, the Wilcoxon test showed no statistically significant difference between pre- and post-tests (Z = -1.00, p = 0.32) but a significant difference was found between pre-and delayed post-tests (Z = -2.00, p = 0.05) and a trend to significant between post- and delayed post-tests (Z = -1.73, p = 0.08).

To find out whether there was a significant difference between the three groups at the time of post-and delayed post-tests, the Kruskal-Wallis test showed a trend to significant differences between the three groups at posttest, H (2) = 5.80, p = 0.06 but not at delayed posttest, H (2) = 1.69, p = 0.43.

Exit questionnaire on elicited imitation test

The students' responses to the exit questionnaire in elicited imitation test were cross tabulated and the results are shown in Table 8.21. Figure 8.12 presents the percentage of learners' response to the questionnaire in three testing times.

Table 8.21 Students' responses to the exit questionnaire on EI

Group	N	Group Cross Tabulation in EI					
		Pre score		Post score		D post score	
		0	**1**	**0**	**1**	**0**	**1**
Metalinguistic	19	19	00	17	2	12	7
Recast	25	25	00	20	5	19	6
Task Only	20	20	00	19	1	19	1
Total	64	64	00	56	8	50	14
Definition: 0= no response 1= correct response							

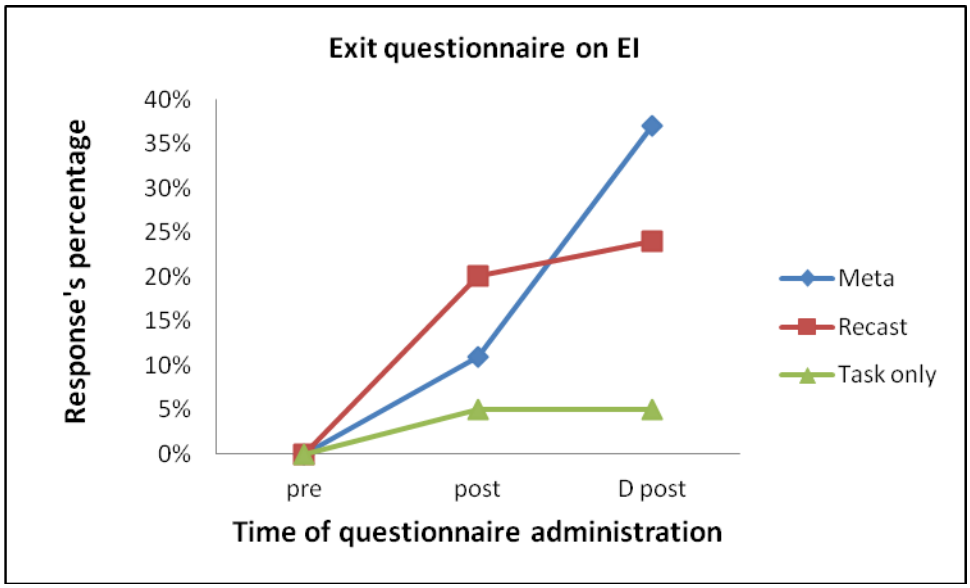


Figure 8.12 Percentage of students' responses to the exit questionnaire on EI test

Analysis

Chi square statistical test was run on the questionnaire given on completion of **elicited imitation** test in the three testing sessions. The results for the three groups in pre testing time indicated no learners 'responses thus statistics were not computed on the exit questionnaire. The results of the questionnaire on post EI test indicated no significant

difference χ^2 (2) = 2.38, p = 0.30 whereas, the result of the questionnaire on delayed posttest indicated a tendency to significant difference, χ^2 (2) = 5.89, p = 0.05.

The results of the Friedman test for each individual group indicated a significant difference between testing times for the metalinguistic information group, χ^2 (2) = 9.75, p = 0.01, the recast group, χ^2 (2) = 7.75, p = 0.02 but no significant difference for the task only group, χ^2 (2) = 1.00, p = 0.61.

For the metalinguistic information group, the Wilcoxon test showed no statistically significant difference between pre- and post-tests (Z = -1.41, p = 0.16) but a significant difference was found between pre-and delayed post-tests (Z = -2.45, p <0.01) and a trend to significant between post- and delayed post-tests (Z = -1.89, p = 0.06).

For the recast group, the Wilcoxon test showed a statistically significant difference between pre- and post-tests (Z = -2.34, p = 0.03) and between pre-and delayed post-tests (Z = -2.45, p = 0.01) but no significant difference between post- and delayed post-tests (Z = -0.45, p = 0.66).

For the task only group, the Wilcoxon test showed no statistically significant difference between pre- and post-tests (Z = -1.00, p = 0.32), between pre-and delayed post-tests (Z = -1.00, p = 0.32) and between post- and delayed post-tests (Z = -0.00, p = 1.00).

To find out whether there was a significant difference between the three groups in the elicited imitation test at the time of post-and delayed post-tests, the Kruskal-Wallis test showed no significant differences between the three groups at posttest, H (2) = 2.35, p = 0.31 but a trend to significant difference at delayed posttest, H (2) = 5.80, p = 0.06.

Exit questionnaire on both GF and untimed GJT

The students' responses to the exit questionnaire in both gap fill and untimed GJT tests were cross tabulated as shown in Table 8.22. Figure 8.13 presents the percentage of learners' response to the questionnaire on both GF and untimed GJT in the three testing times.

Table 8.22 Students' responses to the exit questionnaire on both GF & untimed GJT

Group	N	Group Cross Tabulation in GF & untimed GJT					
		Pre score		post score		D post score	
		0	**1**	**0**	**1**	**0**	**1**
Metalinguistic	19	17	2	11	8	10	9
Recast	25	21	4	20	5	17	8
Task Only	20	19	1	20	0	17	3
Total	64	57	7	51	13	44	20
Definition: 0= no response 1= correct response							

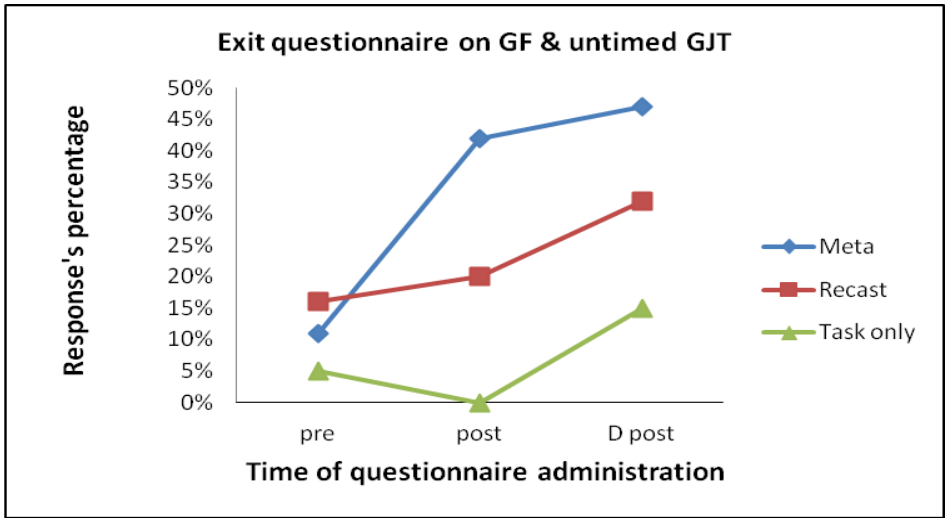


Figure 8.13 Percentage of students' responses to the exit questionnaire on GF and untimed GJT

Analysis

The result of Chi-square statistical test for the exit questionnaire given on completion of both **gap fill and untimed GJT** tests on the three testing sessions indicated no significant difference at the pre testing time, χ^2 (2) = 1.39, p = 0.50 whereas a significant difference was shown in post testing time, χ^2 (2) = 10.67, p = 0.01. There was no significant difference on learners' responses on the exit questionnaire at delayed posttest, χ^2 (2) = 4.76, p = 0.09.

To find out if a significant difference was observed in each individual group, the results of the Friedman test indicated a significant difference for the metalinguistic information group, χ^2 (2) = 8.60, p = 0.01 but no significant difference for the recast group, χ^2 (2) = 2.36, p = 0.31 and the task only group, χ^2 (2) = 3.50, p = 0.17.

In both gap fill and untimed GJT, for the metalinguistic information group, the Wilcoxon test showed a statistically significant difference between pre- and post-tests (Z = -2.21, p = 00) and a significant difference was found between pre-and delayed post-tests (Z = -2.65, p <0.01) but no significant difference between post- and delayed post-tests (Z = -0.45, p = 0.66).

For the recast group, the Wilcoxon test showed no statistically significant difference between pre- and post-tests (Z = -0.38, p = 0.71), between pre-and delayed post-tests (Z = -1.27, p = 0.21) and between post- and delayed post-tests (Z = -1.34, p = 0.18).

For the task only group, the Wilcoxon test showed no statistically significant difference between pre- and post-tests (Z = -1.00, p = 0.32), between pre-and delayed post-tests (Z = -1.00, p = 0.32) but a borderline significant difference was found between post- and delayed post-tests (Z = -1.73, p = 0.08).

To find out whether there was a significant difference between the three groups in both gap fill and untimed GJT at the time of post-and delayed post-tests, the Kruskal-Wallis test showed significant differences between the three groups at post-test, H (2) = 10.51, p = 0.01, but no significant difference at delayed post-test, H (2) = 4.67, p = 0.10.

Given a significant difference was found at delayed posttest, the result of Mann-Whitney test indicated a significant difference between the metalinguistic information group vs. task only group (Z=-2.16, p=0.03) in the direction of metalinguistic information group.

CHAPTER 9: COMPARATIVE DATA ANALYSIS AND DESCRIPTIVE STATISTICS IN THE TWO CONTEXTS (UK&SA)

This chapter presents the results and statistical analysis to answer the following research question:

Q2) Are results observed differentially in different contexts: EFL in SA, and ESL in the UK?

Note, only two tests were carried out in both the UK and SA context: the PD and the GF, therefore this chapter focuses only on these measures. It is also important to note that for a compatible study between the UK and SA contexts, the new items in the GF test were excluded remained eleven items identical to the ones used in the UK context.

The chapter pinpoints the normality and baseline of the data in the two outcome measures. Comparative descriptive results for the three conditions in the UK and SA on free oral picture description will be presented in sections 9.1, and 9.2, followed by section 9.3 which highlights the descriptive results on gap fill test.

Normality of data for PD

The results of Shapiro-Wilk test are presented in Appendix J. The tests of normality indicated that the assumption of normality was violated in the data from PD in the UK and SA contexts for metalinguistic information, recast and task only groups, thus non parametric tests were carried out for that measure.

When the groups in SA and UK were joined together, the tests of normality indicted non normal distribution for the data from PD.

Normality of data for GF

The test of normality indicated that the assumption of normality was not violated in the data from GF for metalinguistic information groups in SA and UK contexts, thus the parametric test ANCOVA (see below for why ANCOVA rather than ANOVA was chosen) was carried out.

The data from gap fill for the recast and task only groups indicated that the assumption of normality was violated, thus non parametric tests were carried out.

Baseline parity

Free oral picture description test

The results of Kruskal-Wallis suggested similar baseline across the different groups in SA and UK. For the metalinguistic information groups (H (1) =0.05, p=0.82), the recast groups (H (1) =0.69, p=0.41) and task only groups (H (1) =0.97, p=0.76). Thus the actual scores were analysed rather than the gain scores for all groups.

The results of Kruskal Wallis test indicated similar baseline for joined groups in SA and UK on picture description data H (1) = 0.03, p = 0.87.

Given that the above results suggested similar baseline across the three different groups in SA and UK for the picture description test, the population of each condition for the two contexts were joined and analysed as shown in section 9.1. To draw on the source of gains, if there were any, separate analysis for the picture description test on each condition was undertaken as discussed in section 9.2.

Gap fill test

The results of ANOVA suggested different baseline and normal distribution for the data from gap fill across the UK and SA contexts for the metalinguistic information groups, F (1,31)= 20.62, p<0.01. To take account of this difference analysis of covariance (ANCOVA, with pre-test scores as the covariate and posttest and delayed posttest as the dependent variables) was computed (e.g., Ellis et al. 2006).

The results of Kruskal-Wallis, from gap fill data, suggested different baseline and not normal distribution for the recast groups (H (1) = 14.47, p<0.01) and task only groups (H (1) = 6.03, p= 0.01). The gain scores, for this measure, were analysed rather than raw scores. Please refer to Appendix J (Tables J.9-J.14) for more information.

Given that the above results suggested different baseline across the three different groups in SA and UK for the gap fill test, the population of each condition for the two contexts were analysed separately.

9.1 Free Oral Picture Description Test for Joined Groups in UK and SA

Descriptive results

Given this chapter suggests comparative results for the two contexts, it might be of value to separately present the scores for each group in the UK and SA contexts as shown in Table 9.1-A. The mean scores and standard deviations, for the joined conditions in the two contexts, are shown in Table 9.1-B, and presented graphically in Figure 9.1. In addition, separate descriptive results for each condition will be presented in the following sections.

Note, in order to elicit the target structure (e.g., English modals) in this measure, each participant was given seven pictures to describe. The score was given on suppliance in obligatory contexts. The numerator and denominator therefore varied from one participant to another.

Table 9.1-A Comparative mean scores on oral PD test for each group in UK & SA

Country	Treatment group	N	Free oral picture description test in SA and UK					
			Pre-test		Post-test		D post-test	
			Mean	SD	Mean	SD	Mean	SD
UK	Metalinguistic	13	85.34	14.96	94.77	6.59	93.52	9.35
	Recast	13	88.19	8.90	91.22	6.63	95.84	4.41
	Task only	10	84.93	30.25	94.15	5.74	96.14	2.67
SA	Metalinguistic	19	81.98	29.63	91.98	11.59	87.60	22.74
	Recast	25	73.95	40.95	88.21	14.26	85.22	26.96
	Task only	20	86.14	23.23	90.98	11.02	93.77	9.26

Table 9.1-B Comparative mean scores on oral PD test for joined groups in UK & SA

Treatment group	N	Pre-test		Post-test		D post-test	
		Mean	SD	Mean	SD	Mean	SD
Metalinguistic **UK-SA**	32	83.34	24.48	93.11	9.84	90.00	18.51
Recast **UK-SA**	38	76.36	31.54	89.24	12.18	88.85	22.45
Task only **UK-SA**	30	85.74	25.25	92.03	9.60	94.56	7.73

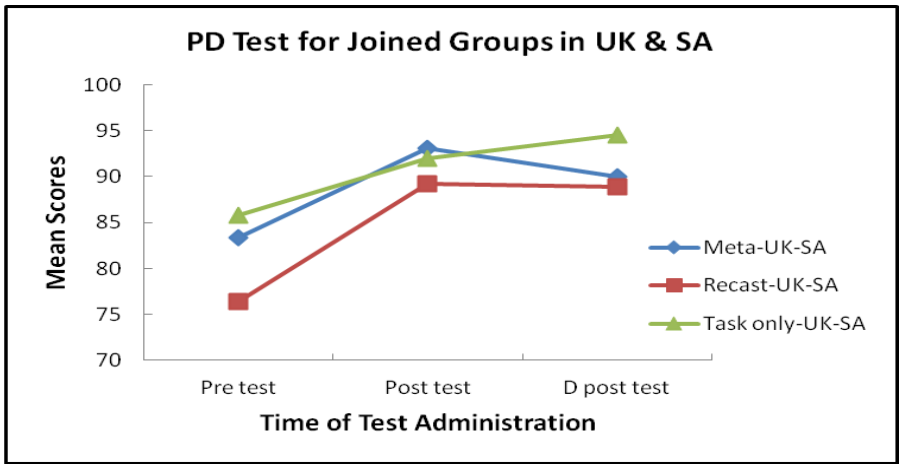


Figure 9.1 Comparative mean scores on free oral picture description test (All Groups)

Analysis

The results of the Friedman test indicated a significant difference between testing times for the joined metalinguistic information group, χ^2 (2) = 14.13, p <0.01, joined recast group, χ^2 (2) = 13.92, p <0.01, but not for the joined task only group, χ^2 (2) = 3.45, p= 0.18.

For the joined metalinguistic information groups the Wilcoxon test showed a statistically significant difference between pre- and post-test (Z = -3.63, p <0.01), and

between pre-and delayed post-test (Z = -2.77, p = 0.01), but not between post-and delayed post-test (Z = -0.88, p = 0.38).

For the joined recast groups there was a statistically significant difference between pre-and post-test (Z = -2.63, p = 0. 01), between pre- and delayed post-test (Z = -3.36, p <0.01), and there was a trend to statistically significant difference between post-and delayed post-tests (Z = -1.90, p = 0.06).

For the joined task only groups there was no statistically significant difference between pre-post-test (Z = -0.20, p = 0.85), pre-delayed post (Z = -1.64, p = 0.10) but significant difference between post-delayed post-test (Z = -2.16, p = 0.03).

To find out whether there was a significant difference between the three groups across the two contexts at the time of post and delayed post-tests, the Kruskal-Wallis test showed no differences between the three groups at post-test, H (2) = 3.75, p = 0.15 nor at delayed post-test, H (2) = 0.84, p = 0.66. Please refer to Appendix K (Table K.21).

Summary

The results suggested that the provision of CF via recasts and metalinguistic information helped the development of English modals in the two contexts in the short—term and this was maintained in the longer term. The delayed post test results indicated no further gains were made after the intervention. The provision of interaction tasks alone did not seem to significantly help the learning of English modals as shown in Table 9.2.

However, when joined groups were compared at post and delayed post-tests the results indicated that the three intervention types did not seem to lead to significantly different scores.

Table 9.2 Summary of statistically significant differences for joined groups in PD test

Test	Change over time	Joined Groups		
		Joined MI	Joined R	Joined TO
PD	Pre-post	Sig	Sig	Not sig
	Pre-delayed post	Sig	Sig	Not sig
	Post-delayed post	Not sig	Trend	Sig

9.2 Comparison of Free Oral Picture Description Test scores between UK and SA context

Descriptive results for metalinguistic information groups

The mean scores and the standard deviations for the metalinguistic information groups across SA and UK contexts are shown in Table 9.3 and presented graphically in Figure 9.2.

Table 9.3 Comparative mean scores on free oral picture description tests for MI groups

Treatment group	N	Pre-test		Post-test		D post-test	
		Mean	SD	Mean	SD	Mean	SD
Metalinguistic **UK**	13	85.34	14.96	94.77	6.59	93.52	9.35
Metalinguistic **SA**	19	82.00	29.63	91.98	11.59	87.6	22.74

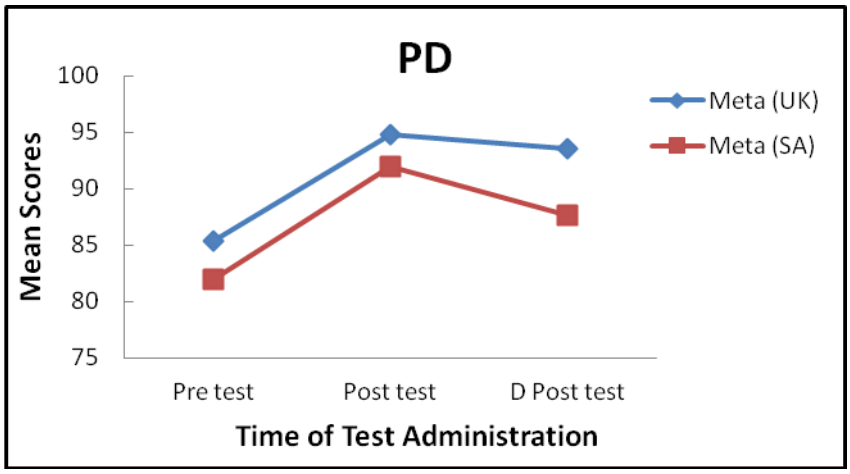


Figure 9.2 Comparative mean scores on free oral picture description test (MI groups)

Analysis

The results of the Friedman test indicated a significant difference between testing times for the **metalinguistic information groups** in the two contexts. The result for the metalinguistic information group in SA showed no significant difference between tests, χ^2 (2) = 4.62, p = 0.09 but the metalinguistic information group in the UK indicated significant difference between tests χ^2 (2) = 10.92, p <0.01.

The Wilcoxon test showed a statistically significant difference in the UK group between pre-and post-test (Z = -2.76, p <0.01) and between pre- and delayed post-tests (Z = -2.82, p <0.01) but not between post- and delayed post-tests (Z = -0.94, p = 0.35).

To find out whether there was a significant difference between the metalinguistic groups across SA and UK contexts at the time of post and delayed post-tests, the Kruskal-Wallis test indicated no significant difference at post-test, H (1) =0.31, p=0.58 nor at the delayed post-test (1) =0.54, p=0.46. Please refer to Appendix K (Table K.15) for more information.

Descriptive results for recast groups

The mean scores and the standard deviations for the recast groups across SA and UK contexts are shown in Table 9. 4 and presented graphically in Figure 9.3.

Table 9.4 Comparative mean scores on oral PD test for recast groups

Treatment group	N	Pre-test		Post-test		D post-test	
		Mean	SD	Mean	SD	Mean	SD
Recasts - **UK**	13	88.19	8.90	91.22	6.63	95.84	4.41
Recasts - **SA**	25	70.21	37.13	88.21	14.26	85.22	26.96

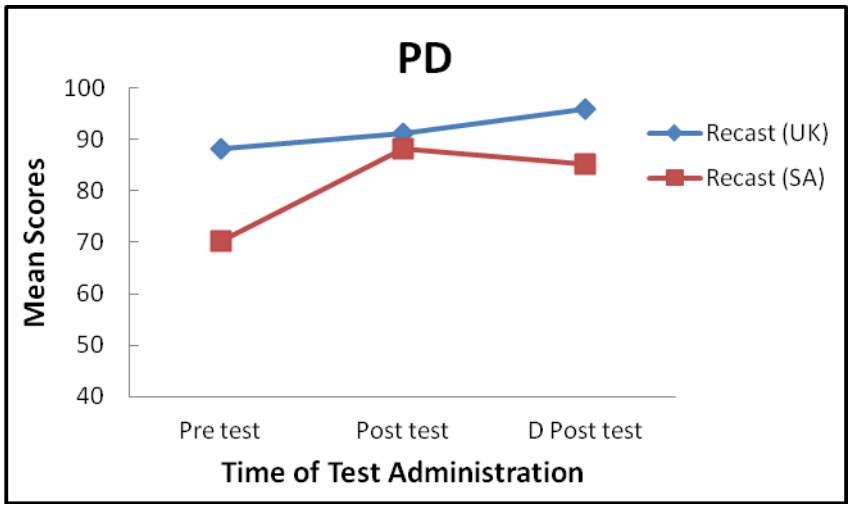


Figure 9.3 Comparative mean scores on free oral picture description test (R groups)

Analysis

For the **recast groups** in SA and UK, the results of Friedman test indicated no statistically significant difference in testing times for the recast group in SA, χ^2 (2) = 4.85, p = 0.09, but there was a significant difference in testing times for the recast group in the UK, χ^2 (2) = 12.67, p <0.01.

To find out which testing time for the recast group in the UK was statistically different, the Wilcoxon test indicated a significant difference between pre-and delayed post-test (Z = -2.67, p = 0.01), and between post- and delayed post-test (Z = -3.06, p <0.01) but no significant difference between pre-and post-test (Z = -1.49, p = 0. 14).

To find out whether there was a significant difference between the two groups across SA and UK at the time of post-test and delayed post-tests, a Kruskal-Wallis test showed no significant differences between the groups at post-test, H (1) = 0.06, p = 0.81 nor at delayed post-test, H (1) = 1.10, p = 0.22. Please refer to Appendix K (Table K.16) for more information.

Descriptive results for task only groups

The mean scores and the standard deviations for the task only groups across SA and UK contexts are shown in Table 9.5 and presented graphically in Figure 9.4.

Table 9.5 Comparative mean scores on oral PD test for task only groups

Treatment group	N	Pre-test		Post-test		D post-test	
		Mean	SD	Mean	SD	Mean	SD
Task Only **UK**	10	84.93	30.25	94.15	5.74	96.14	2.67
Task Only **SA**	20	86.14	23.23	90.98	11.02	93.77	9.26

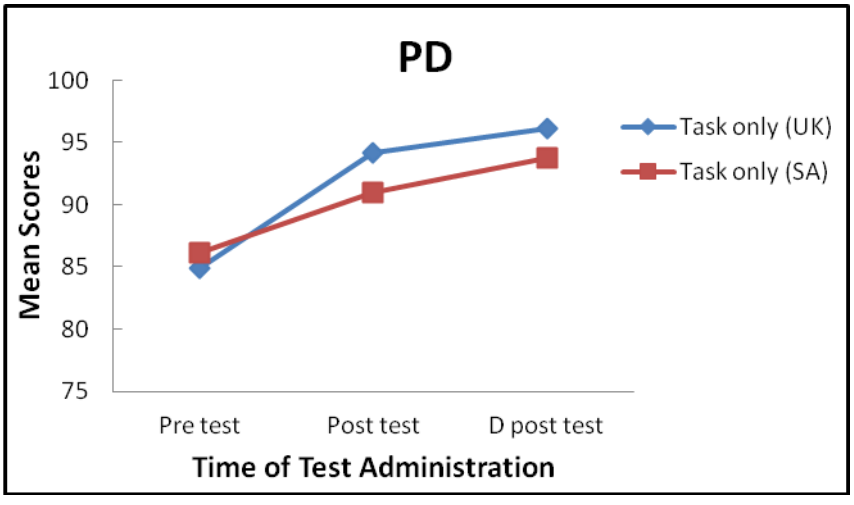


Figure 9.4 Comparative mean scores on free oral picture description test (TO groups)

Analysis

The results for the **task only groups** in SA and UK indicated no significant differences at all levels of statistical tests. The Friedman test indicated no significant difference at testing times for the group neither in SA χ^2 (2) = 2.69, p = 0.26 nor in UK, χ^2 (2) = 4.20, p = 0.12. Please refer to Appendix K (Table K.17).

Summary

The results on oral PD suggested that the provision of the CF via recasts and metalinguistic information helped the development of English modals in the UK context but not SA context. It also indicated that the provision of the instruction tasks alone did not lead to the development of English modals as measured by a PD task neither in SA nor in UK as summarized in Table 9.6.

Table 9.6 Summary of statistically significant differences across context on PD

Groups	Change over time	Change across context	
		SA	UK
Metalinguistic information	Pre-post	Not sig	Sig
	Pre-delayed		
	Post-delayed	Not sig	Not sig
Recast	Pre-post	Not sig	Not Sig
	Pre-delayed	Not sig	Sig
	Post-delayed		
Task only	Pre-post	Not sig	Not sig
	Pre-delayed		
	Post-delayed		

9.3 Gap Fill Test – Comparisons between UK and SA Contexts

Given that this section suggests comparative results for the different groups on GF test administered in the UK and SA contexts, Table 9.7 presents separate score for each group in each context. In addition, the descriptive results for each joined group will be discussed later in this section.

Table 9.7 Comparative mean scores on gap fill test for each group in UK & SA

Gap Fill Test in UK and SA								
Country	Group	N	Pre test		Post test		D post test	
			Mean	SD	Mean	SD	Mean	SD
UK	Metalinguistic	13	52.10	20.38	65.38	19.77	72.73	21.88
	Recast	13	59.44	20.62	78.67	12.77	76.22	18.10
	Task only	10	61.36	24.64	66.82	28.91	69.55	21.65
SA	Metalinguistic	19	20.81	18.26	37.80	25.94	40.66	30.01
	Recast	25	23.09	22.99	35.27	27.00	34.18	25.92
	Task only	20	32.05	28.99	44.32	31.28	36.36	30.01

Descriptive results for metalinguistic information groups

The mean scores and the standard deviations for the metalinguistic information groups across SA and UK contexts are shown in Table 9.8 and presented graphically in Figure 9.5.

Table 9.8 Comparative mean scores on gap fill test for MI groups

Treatment group	N	Pre-test (k=11)		Post-test (k=11)		D post-test (k=11)	
		Mean	SD	Mean	SD	Mean	SD
Metalinguistic UK	13	52.10	20.38	65.38	19.77	72.73	21.88
Metalinguistic SA	19	20.81	18.26	37.80	25.94	40.66	30.01

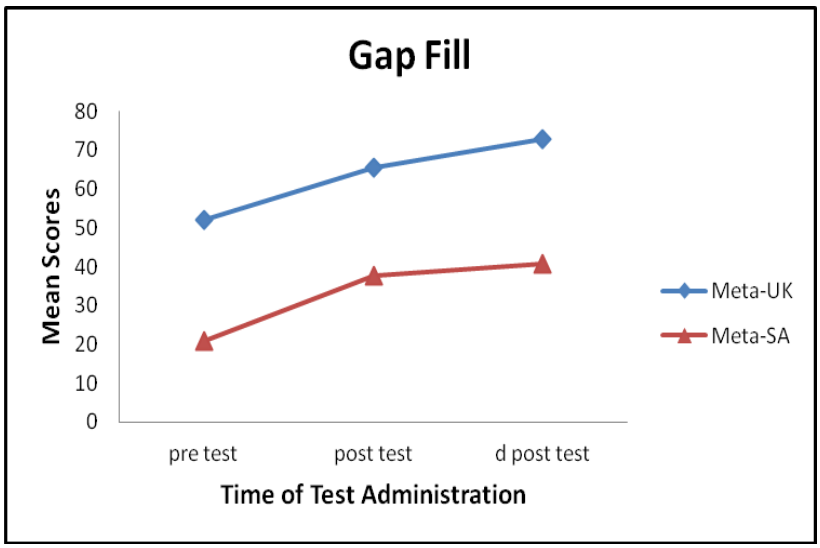


Figure 9.5 Comparative mean scores on gap fill test (MI groups)

Analysis

The analysis of the changes over testing time for the **metalinguistic information** groups in SA and the UK contexts, the results of the ANCOVAs suggested a significant difference on pre-posttest, F (1, 30) =29.47, p <0.01, on pre-delayed posttest, F (1, 30) =38.21, p<0.01and on post-delayed posttest F (1, 30) =84.39, p <0.01.

When context was used as a between-subject variable, the results of ANCOVA for the metalinguistic information group, suggested a significant difference for the group in SA on pre-posttest, F (1, 17) =21.85, p <0.01but no significant difference for the UK group, F (1, 11) =0.97, p = 0.35. There was a significant difference on pre-delayed posttest, for the metalinguistic information group in SA, F (1, 17) =21.96, p <0.01but not in the UK, F (1, 11) =2.53, p = 0.14. However, there was a significant difference on post-delayed posttest in SA, F (1, 17) =51.67, p <0.01, and UK, F (1, 11) =7.66, p = 0.02. For more information, please refer to Appendix K (Table K.18).

Descriptive results for recast groups

The descriptive results on gain scores and the standard deviations for the recast groups across SA and UK contexts are shown in Table 9.9 and presented graphically in Figure 9.6. ***Note,*** the gain scores for the recast groups were used because the baseline was

not the same across contexts. The assumption of normality was violated and so non-parametric tests were used.

Table 9.9 Comparative gain scores on gap fill test for recast groups

Treatment group	N	Gain Scores		
		Post-pre	D post- pre	D post-post
Recasts UK	13	19.23	16.78	-2.45
Recasts SA	25	12.18	11.09	-1.09

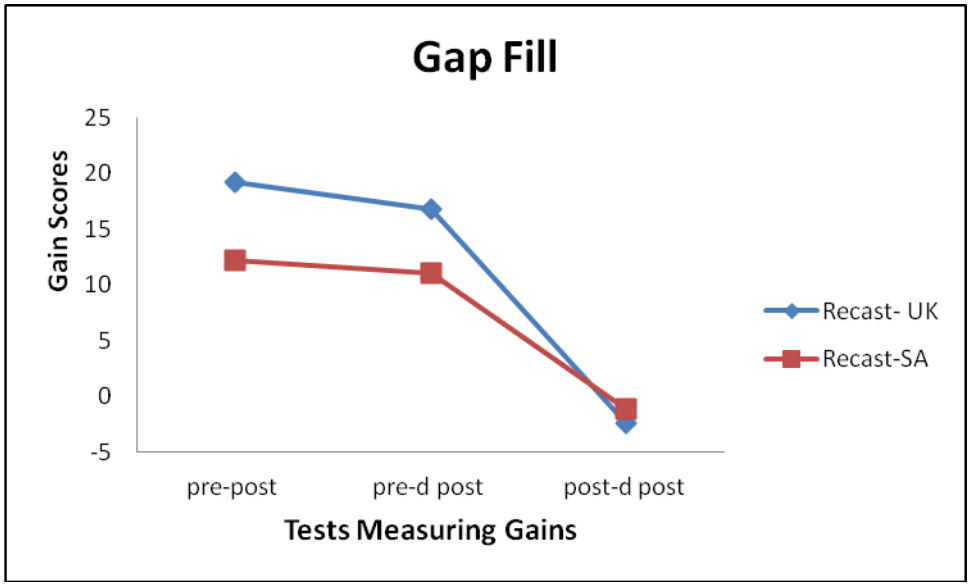


Figure 9.6 Comparative gain scores on gap fill test (R groups)

Analysis

For the **recast groups** in SA and UK contexts, the results of the Kruskal-Wallis suggested no significant difference between groups on pre-post gains (H (1) =1.55, p=0.21), pre-delayed post gains (H (1) =0.62, p=0.43), and on post-delayed post gains (H (1) =0.01, p=0.97). Detailed information is included in Appendix K (Table K.19).

Descriptive results for task only groups

The descriptive results on gain scores and the standard deviations for task only groups across SA and UK contexts are shown in Table 9.10 and presented graphically in Figures 9.7. ***Note,*** the gain scores for the task only groups were used because the baseline was not the same across contexts. The assumption of normality was violated and so non-parametric tests were used.

Table 9.10 Comparative gain scores on gap fill test for task only groups

Treatment group	N	Gain Scores		
		Post - pre	D post- pre	D post-post
Task Only UK	10	5.46	8.19	2.73
Task Only SA	20	12.27	4.31	-7.96

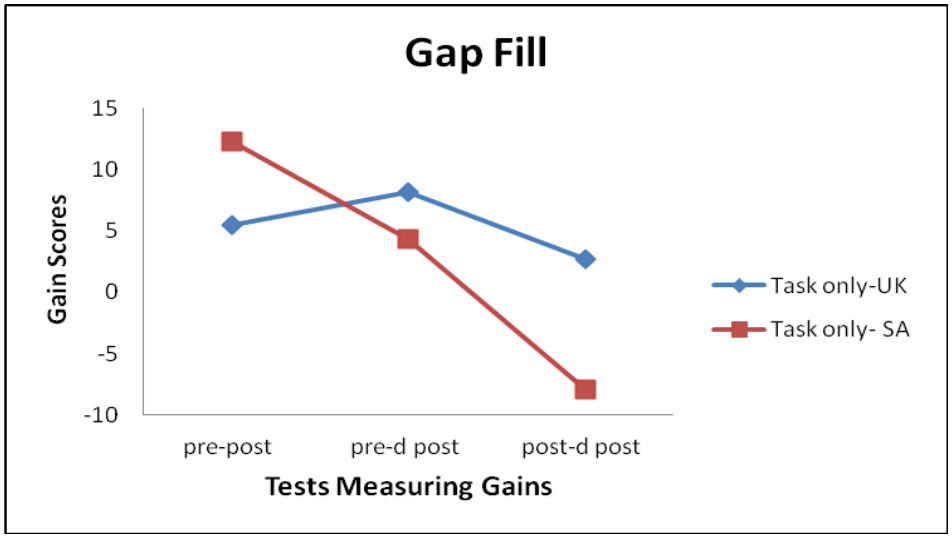


Figure 9.7 Comparative gain scores on gap fill test (TO groups)

Analysis

For **task only** groups in SA and UK contexts, the results of Kruskal-Wallis indicated no significant differences between groups on pre-post gains (H (1) =1.58, p=0.21), pre-delayed post gains (H (1) =0.33, p=0.56) or on post- delayed post gains (H (1) =2.29, p=0.13). More information is included in Appendix K (Table K.20).

Summary

The results of the gap fill test suggested that the provision of CF via metalinguistic information helped the development of English modals for the metalinguistic information groups in the UK and SA contexts.

It also indicated that the provision of recasts and the instruction tasks alone did not lead to significant development of English modals as measured by a GF task neither in SA nor in UK as summarized in Table 9.11.

Table 9.11 Summary of statistically significant differences across SA and UK on Gap Fill

Groups	Change over time	Change across context	
		SA	UK
Metalinguistic information	Pre-post	Sig	Not sig
	Pre-delayed		
	Post-delayed	Sig	Sig
Recast	Pre-post	Not sig	Not sig
	Pre-delayed		
	Post-delayed		
Task only	Pre-post	Not sig	Not sig
	Pre-delayed		
	Post-delayed		

CHAPTER 10: OVERALL DISCUSSION

This chapter discusses the findings presented in Chapters 7, 8 and 9. Generally, the findings of the current study are relative to the different research questions. They seemed very complicated and I had to break it down for the study was a classroom based experiment not laboratory, it was carried out in 2 different contexts (UK and SA) and at different time, and different battery of oral and written outcome measures were administered.

The results of the study might be in line or different to other reviewed studies. Of course, the results from the reviewed studies cannot be directly compared to that of the current study because of a fundamental difference in the way the studies were operationalized for example, the population, the type of measures and the operationalization of the interventional feedback.

The discussion of the different outcome measures (broadly intended to tap into more implicit and more explicit knowledge) used in the UK is presented in section 10.1. Discussion of outcome measures administered in SA will be presented in section 10.2. The possible effects of tests will be discussed in section 10.3. A discussion of the possible effects of the different contexts (ESL and EFL) on the effectiveness of metalinguistic information and recasts will be presented in section 10.4. The relationship between the types of CF techniques (i.e. metalinguistic information and recast) and learners' attitudes in both contexts will be discussed in section 10.5.

10.1 Overall Discussion of the Results from the UK

Based on the data analysis for the different outcome measures (thought of as more implicit and more explicit) administered in the UK contexts, a summary of the statistical significant overtime changes at the different testing times are presented in Tables 10.1& 10.3. A summary of gains and directions observed (no statistical significant indication) on the outcome measures for the three groups are demonstrated in Tables 10.2 & 10.4. Detailed discussion on the results of each outcome measure relative to the reviewed

previous research on the effectiveness of CF on learning a language will be presented separately in the following sections.

10.1.1 Effectiveness of CF on Implicit Knowledge in the UK

The term 'implicit knowledge' is used to refer to the knowledge elicited by the free oral PD task and the correct items in timed GJT (see Ellis, 2005 where correct items were found to correlate highly with implicit knowledge). The term 'explicit knowledge' is used to refer to the knowledge elicited by the gap fill and the incorrect items in timed GJT. It is acknowledged that these measures may elicit both explicit and implicit knowledge at different times, and that the distinction between the two knowledge types is not entirely clear cut.

Table 10.1 Summary of statistically significant differences within the group in outcome measures (more implicit) in the UK using Wilcoxon statistical test

Type of measure	Tests	Change over time	Group		
			MI	R	TO
Implicit Measures	PD	Pre - post	Sig	Trend	Not sig
		Pre - delayed	Sig	Sig	Not sig
		Post- delayed	Not sig	Sig	Not sig
	Timed GJT-Correct	Pre - post	Not sig	Not sig	Not sig
		Pre - delayed	Not sig	Not sig	Sig
		Post - delayed	Not sig	Not sig	Not sig
Sig: significant					

Table 10.2 Summary of gains and directions for all groups on outcome measures (more implicit) in the UK but statistical significant is not indicated

Type of measure	Tests	Pre - post gain	Pre-delayed post gain	Post-delayed post gain
Implicit Measures	PD UK	9.43=9.22>3.03	11.21>8.18>7.65	4.62>1.99> -1.25
		MI=TO>R	**TO>MI>R**	**R>TO>MI**
	Timed GJT overall-UK	9.83>8.88>1.71	12.77=12.4>8.12	10.69>3.89> -1.71
		MI>TO>R	**TO=R>MI**	**R>TO>MI**
	Timed GJT - Correct-UK	7.77>2.57>-5.13	21.11>2.56>-6.84	13.34>7.69> -9.41
		TO>MI>R	**TO>R>MI**	**TO>R>MI**

Effects of CF and oral tasks alone on free oral picture description test

The results suggested a pre-post and pre-delayed post significant development for the target forms in the metalinguistic information group, whereas for the recast group pre-delayed post and post-delayed post significant gains were observed as well as a trend to significant pre-post gains.

Although the results from the current study cannot be directly compared to that from Ammar's (2006) because of a fundamental difference in the operation of the explicit CF, similar results were found in Ammar's suggesting significant outperformance for the recast and prompt groups over the control group in the picture description tasks on the immediate and delayed post-tests.

The usefulness of recast observed in the current study is different from Lyster's (2004) which implied the ambiguity and difficulties of recast in noticing errors. The

recast group might not have perceived this type of corrective feedback as error correction for they provided implicitly during the interaction activities and the flow was not interrupted and thus could be seen as positive evidence which might have contributed to the longer term gains. The delayed effect for the recast CF suggests that implicit knowledge takes longer process to be automatized.

The significant results for the two CF groups might support the claim that "embedding CF within communicative activities is more effective than participation in such activities without CF" (Ammar & Spada, 2006, p.562).

Effects of CF and oral tasks alone on grammatically correct items in timed GJT

Given that grammatical items in timed grammaticality judgment test may tap into learners' implicit knowledge, the descriptive results on gain scores for the three groups (see Table 10.2) showed that task only group scored higher than the experimental groups (the metalinguistic information and the recast) at all testing times. This finding is in line with the arguments made by Schwartz (1993) and Truscott (1996) suggesting that positive evidence is sufficient for learners to acquire L2 but different from Mackey and Philp (1998); Long and Robinson (1998); Nicholas, et al (2001); Long, (2006), and Révész (2009) suggesting that for the CF to be effective focus on form and meaning should be provided together in a classroom context and that interactional feedback might benefit learning, enhance comprehensibility and facilitate L2 development.

When groups were compared to each other via the MW statistical test with reference to the descriptive results, metalinguistic information group significantly outperformed the recast and task only groups in post testing time as shown in Table 10.5. This result is in line with Ellis et al (2006) and Sheen (2006) whose studies indicated the useful role of metalinguistic information feedback over recast and control groups (although fundamental difference could be seen in the operation of metalinguistic information feedback and the outcome measures in their studies), but different to Loewen and Erlam (2006) who found no significant effects for group or time in timed GJT test (see Chapter 2 for more details on this study). Conversely, the insignificant role for the metalinguistic information group in the delayed post-test lends support to Krashen's theory suggesting that learning via error correction and explicit teaching of rules is not

relevant to language acquisition as there is no interface between the two (Krashen, 1982, p.11).

Final comments on outcome measures (more implicit)

As measured by the oral free response picture description and correct items of timed GJT used here, the results suggest a useful role of metalinguistic information and recast CF particularly on the oral outcome measure.

Metalinguistic information feedback had some positive impact on the 'implicit' measures. This could be attributed to the fact that 'implicit measures' did not actually tap into implicit knowledge i.e. participants used some explicit knowledge.

It might also be argued that learners in the recast group benefited from the repeated exposure to positive models and opportunities which enable them to infer negative evidence i.e. the recast treatment led to explicit knowledge about the target structure.

The short term progression for the metalinguistic information group rather than long term supports Krashen's strong interface position indicating that learning cannot be converted to acquisition whereas the short and long term progression for recast group is in line Krashen's input hypothesis signifying the useful role of positive comprehensible input.

10.1.2 Effectiveness of CF on Explicit Knowledge in the UK

'Explicit knowledge' was operationalized in the UK context via a gap fill test and the grammatically incorrect items in timed grammaticality judgment tests which could tap into learners' explicit knowledge (Ellis, 2005). In timed GJT it is likely that learners access their explicit knowledge even if they are under time pressure (DeKeyser 2003). Note also that in the current study learners were asked to provide corrections on the ungrammatical items, making it particularly likely that these items tapped explicit knowledge.

However, it is acknowledged, as above, that these tests do not solely elicit explicit knowledge. Table 10.3 summarises the statistical significant overtime changes within each individual group on the different outcome measures (thought of as explicit).

Table 10.4 entails summary of gains and directions (statistical significant is not indicated) for the three groups on the different outcome measures in the UK.

Table 10.3 Summary of statistically significant differences within the group on outcome measures (more explicit) in the UK using Wilcoxon statistical test

Type of measure	Tests	Change over time	Group		
			MI	R	TO
Explicit Measures	Timed GJT-Incorrect	Pre ─ post	Sig	Not sig	Not sig
		Pre - delayed	Sig	Sig	Not sig
		Post - delayed	Not sig	Not sig	Not sig
	GF	Pre - post	Not sig	Sig	Trend
		Pre - delayed	Sig	Sig	Not sig
		Post - delayed	Not sig	Not sig	Not sig
Sig: significant					

Table 10.4 Summary of gains and directions for all groups on outcome measures (more explicit) in the UK but statistical significant is not indicated

Type of measure	Tests	Pre - post gain	Pre-delayed post gain	Post-delayed post gain
Explicit Measures	Timed GJT-Incorrect-UK	17.09>10>8.55	23.07>22.22>4.45	13.67>5.98> -5.55
		MI>TO>R	**MI>R>TO**	**R>MI>TO**
	GF- UK	19.23>13.28>5.46	20.63>16.78>8.19	7.35>2.73> -2.45
		R>MI>TO	**MI>R>TO**	**MI>TO>R**

Effects of CF and oral tasks alone on gap fill test

The results suggest a significant development for the recast group on pre-post and pre-delayed post time but not on post-delayed post time. The significant gains evident for the recast group could be attributed to the provision of positive evidence which might have helped learners to locate the errors and notice the difference between their erroneous production and the correct form of the target and so induce their own grammatical rules.

The pre-delayed post significant gains for the metalinguistic information group could suggest that this type of CF had raised learners' awareness to the linguistic problems in their production and the target information provided and so induce their own interlanguage.

When groups were compared to each other in post and delayed post testing times, the MW statistical test indicated that on post-test time a more implicit condition (recasts) had a greater significant effect on a gap fill test than a more explicit condition (metalinguistic information as shown in Table 10.5. This result is different to DeKeyser (1995) who found better performance for the explicit deductive condition (traditional rule teaching) than the implicit inductive (going from examples to rules) condition in the fill-in-the-blank test.

Effects of CF and oral tasks alone on grammatically incorrect items in timed GJT test

The results on the incorrect items of timed GJT test showed internal significant overtime changes for the metalinguistic information group in pre-post and pre-delayed post but not on post-delayed post time. This finding suggests that the explicit type of feedback might have helped learners' conscious attempts to derive and test hypotheses related to language structure, which, in turn, could contribute either directly or indirectly to the interlanguage development (Ellis, 1993; Ellis, et al 2006; Sheen 2006).

Significant internal overtime change was also found for the recast group on pre-delayed post time only. For example, the gains on the ungrammatical items of timed GJT test for the recast group suggested that rules might have been induced via the examples in the input.

The results indicated no significant internal overtime changes on the incorrect items of timed GJT test for the interaction alone group. This result is different to Erlam and Loewen (2010) who found interaction alone effective.

The different result for the two CF conditions is not in line with Loewen and Nabei (2007) whose study showed that the "three feedback groups outperformed the non-feedback groups but did not differ from each other"(p.374).

Final comments on the outcome measures (more explicit)

A possible explanation of the recognition evident in the grammatically incorrect items of timed GJT test in short-and longer-terms particularly for the metalinguistic information group could be attributed to the fact that timed GJT test required participants to focus attention primarily on forms through judging the correctness of sentences. This process may push learners to notice and visualize the incorrect elements which, later, became a source learners referred to (though learners were asked not to go back to the previous pages and were under time pressure). However, such a test effect is unlikely to explain the gains made in the CF groups as the task only group was recruited from the same language school population but no significant gains observed (i.e. a test effect did not lead to changes in their scores).

The gains observed at delayed post-test on these 'explicit' measures (grammatically incorrect items of timed GJT and GF tests) for the metalinguistic information group could have been due to the metalinguistic information being converted to implicit knowledge. On the other hand, they may be due to participants remembering the metalinguistic information at delayed post-test.

The delayed gains observed on the explicit measures (grammatically incorrect items of timed GJT and GF tests) for the recast group could be attributed to the fact that positive input helped in converting explicit knowledge to implicit knowledge which in turn takes longer process to be atomized.

Summary

The experiment undertaken in the UK context compared the effectiveness of two different types of corrective feedback and oral task alone on learners' performance in free oral PD, timed GJT (grammatically correct and incorrect) and GF tests. Learners' achievement was measured in three testing times; pre-, post- and delayed- posttests.

Generally, based on the progression made in the three testing times on the different outcome measures for the three groups, the results indicated significant gains for the two experimental groups on certain tests and at certain time. However, the significant effects for the two CF types lend support to previous research findings (e.g., Carroll and Swain, 1993; Erlam & Loewen, 2010; Loewen and Erlam, 2006; Loewen and Nabei, 2007; Sauro, 2009).

The task only group in the three testing times on all outcome measures indicated no significant gains with the exception of significant and a trend to significant gains observed in pre-delayed post-test in correct items of timed GJT and pre-post-test of GF test respectively. This result is different to Erlam and Loewen (2010) who found significant effects for oral interaction.

When groups were compared via the MW statistical test and the descriptive results on post-and delayed-post-tests, the metalinguistic information group outperformed the task only and the recast groups in post grammatically correct items of timed GJT test. On contrary, the recast group outperformed the metalinguistic information group in post-test of GF as shown in Table 5.

Table 10.5 Statistical significant difference between groups at post and delayed posttests in the UK using MW statistical test and the descriptive results

Tests	Post	Delayed post	Statistical result
Timed GJT - Correct- UK	**MI>TO**		**Sig**
	MI>R		**Sig**
GF- UK	**R>MI**		**Sig**

10.2 Overall Discussion in SA

In Saudi context, implicit knowledge was gauged via two oral tests: oral elicited imitation and free oral picture description as discussed in 10.2.1. Explicit knowledge was measured via gap fill, untimed grammaticality judgment, and metalinguistic questions. The relationship of each measure to the different effects of the CFs will be discussed separately in section 10.2.2.

10.2.1 Effectiveness of CF on Implicit Knowledge in SA

The following discussion presents the effects of two types of CF on the different outcome measures (assumed as more implicit and more explicit) used in SA.

A summary of internal statistical significant overtime changes for the three groups on the different oral outcome measures (more implicit) at three different times of tests is shown in Table 10.6 and the different gains and directions (no statistical significant indication) are shown in Table 10.7.

Table 10.6 Summary of statistically significant differences within the group on outcome measures (more implicit) in SA using Wilcoxon statistical test

Type of measure	Tests	change over time	Groups		
			MI	R	TO
Implicit Measures	PD	Pre – post	Sig	Sig	Not sig
		Pre – delayed	Not sig	Sig	Not sig
		Post – delayed	Not sig	Not sig	Sig
	EI Overall	Pre – post	Sig	Sig	Sig
		Pre – delayed	Sig	Sig	Sig
		Post – delayed	Not sig	Not sig	Sig
	EI-Correct	Pre – post	Sig	Trend	Sig
		Pre – delayed	Sig	Sig	Sig
		Post-delayed	Not sig	Trend	Not sig
	EI-Incorrect	Pre – post	Sig	Sig	Not sig
		Pre – delayed	Trend	Sig	Sig
		Post – delayed	Not sig	Trend	Not sig

Table 10.7 Summary of gains and directions for all groups on outcome measures (more implicit) in SA but statistical significant is not indicated

Type of measures	Tests	Pre - post gain	Pre-delayed post gain	Post-delayed post gain
Implicit Measures	EI overall-SA	22.86>19.47>9.76	17.54>15.9>7.26	6.14> -5.32> -12.21
		R>MI>TO	**R>TO>MI**	**TO>R>MI**
	EI -Correct -SA	27.51>13.38>11.32	22.11>18.46>16.59	7.14> 3.21> -5.40
		MI>R>TO	**MI>TO>R**	**TO>R>MI**
	EI - Incorrect-SA	30.70>21.95>6.69	24.28>12.87>8.39	6.18> -6.42> -13.56
		R>MI>TO	**R>TO>MI**	**TO>R>MI**
	PD - SA	18.00>9.98>4.84	15.01>7.63>5.60	2.79 > -2.99> -4.38
		R>MI>TO	**R>TO>MI**	**TO>R>MI**

Effects of CF and oral tasks alone on oral elicited imitation test

The oral elicited imitation test was designed to measure learners' knowledge (more implicit) as it intends to focus learners' attention on meaning under time pressure (discussed in section 2.6.1). The descriptive statistics shown in Table 8.2 indicated an increase for the three groups over time.

Statistical results indicated a significant gain for each group on testing times. Each of the experimental group and the task only group gained significantly on pre-post

and pre-delayed on the EI overall accuracy as shown in Table 10.6. This finding could be attributed to different factors: 1) the CF types; 2) the interaction activities alone; 3) extraneous factors such as maturation; and 4) the tests effects. Each of these factors will be raised in the upcoming sections. The useful role of the CF types and the interaction alone indicated in the current study is similar to Erlam and Loewen (2010)'s finding.

When groups were compared using the MW statistical test, there was a trend to statistical significant group differences on overall scores of EI post testing time with the metalinguistic information group over the task only group as shown in Table 10.12.

Effects of CF and oral tasks alone on grammatically correct items of EI test

The descriptive statistics on the grammatically correct items of EI test showed that the three groups developed their knowledge of English modals over time.

The statistical results suggested significant internal overtime changes for all three groups. The significant gains made by the metalinguistic information group suggest that metalinguistic information can benefit performance on an EI test. This could run counter to the notion that EI tests elicit implicit knowledge or it could suggest that the learners in the metalinguistic group made some gains in implicit knowledge. It is hard to identify which explanation is the best.

The gains evident for the recast group in the grammatically correct items of EI test may support the beneficial role of interactional implicit CF as it tells learners that they have said something incorrect, allowing them to correct their utterances towards a more comprehensible or native-like use of the target language.

The significant gains evident for the task only group may lend support to the argument that interaction with feedback may not necessarily be more facilitative of L2 development than interaction alone (Mackey and Goo, 2007).

In regards to group differences, although a trend to sig difference was evident for the metalinguistic information group over the recast group in the post grammatically correct items of the EI test, the result is different to Ellis et al (2006) whose study showed no significant group differences on the immediate post-test for the grammatically correct items in the EI test and also different to Ellis et al (2006) in delayed post-test time who found group differences with the metalinguistic information group differed significantly

from the control and recast groups on the grammatically correct items of EI test as shown in Table 10.12.

Effects of CF and oral tasks alone on grammatically incorrect items in EI test

The descriptive results on the grammatically incorrect items suggested that the three groups developed their knowledge of English modals over time.

The statistical results, on the grammatically incorrect items of EI test, suggested significant over time changes for the metalinguistic information group on pre-post and a trend toward significant on pre-delayed post. Similarly, the recast group evident significant over time changes on the grammatically incorrect items of EI test at the pre-post and a pre-delayed post time but a trend to significant change evident in post-delayed post time.

Apparently a significant over time changes was noticed for the task only group in pre-delayed post time.

The significant over time changes and gains evident for the experimental groups as well as the task only group in the oral EI test could be attributed to the novelty of this type of activities for this particular group of learners.

When groups were compared to each other (shown in Table 10.12) on post grammatically incorrect items of EI test by using the MW statistical test, a trend to significant difference was found for the metalinguistic information group over task only group. This result is different to Ellis et al (2006) whose study showed no significant group differences on the immediate post-test for the grammatically incorrect items in the EI test. The non-significant group difference on the delayed post-test is also different to Ellis et al (2006) who found group differences on the delayed post-test with the metalinguistic information group differ significantly from the control and a trend toward significance for metalinguistic group over recast group on the grammatically incorrect items.

Although the results on the overall scores and grammatically correct and incorrect items of the EI test (discussed above) for the metalinguistic information group evident a trend to significant group differences compared to recast and task only groups in post testing time, it lends support to Carroll and Swain (1993) who reported that the group

received direct metalinguistic feedback outperformed all of the other groups in the production of the target structure.

Correlation of grammatical and ungrammatical sentences in the EI test

Test takers were told that they should repeat the statements in correct English. An important issue relative to the EI test is the relationship between participants' ability to repeat grammatical sentences correctly and their ability to spontaneously correct ungrammatical sentences. Learners' responses on the grammatically correct and incorrect short and long sentences were computed. The result suggests that learners' scores for repeating short and long grammatically correct sentences were significantly correlated with scores for repeating grammatically incorrect sentences correctly on the three periods of testing time as shown in Table 10.8. This result is in line with Erlam (2006) who found significant positive correlation ($r=0.73$, $p<0.01$, $n=95$) for L2 learners' scores in repeating grammatical sentences with scores for correcting ungrammatical sentences. The result for the current study also suggests that the aim of establishing a reconstructive EI test rather than rote imitation was met.

Table 10.8 Relationship of grammatically correct and incorrect sentences

Type of sentence	N	Pre test		Post test		D post test	
		Mean	Correlation	Mean	correlation	Mean	correlation
Short correct	64	69	r=0.44, p<0.01	76	r=0.42,p<0.01	86	r=0.28,p=0.02
Short incorrect		51		67		66	
Long correct		70	r=0.40, p<0.01	85	r=0.35,p<0.01	84	r=0.49,p<0.01
Long incorrect		56		79		72	

Length of sentences in the EI test

Length of sentences in the EI test has been reviewed in previous research discussed in section 2.6.1 and the general range of sentence length was found between six and nineteen syllables (Graham, McGhee & Millard 2010).

In the current study, the statements in the EI test (including training and distracters) varied between 7 and 17 syllables in length, with the mean length being 11.45. There were thirty three statements targeted English modals, twelve were short syllables (ranged between 7-11 syllables) with the average length 9.4 and twenty one were long syllables (ranged between 12-17 syllables) with the average length 13.1.

To find out the performance of all participants (n=64) in both short and long syllable sentences, the scores were computed for the three testing times. The results shown in Table 10.8 indicate that all participants improved their performance on short and long syllable sentences from pre to post and from pre to delayed posttests but slight decrease or increase in their performance on post to delayed posttest. It was also found that participants were slightly better at repeating short and long grammatically correct than repeating short and long grammatically incorrect sentences.

The result also suggests that participants were slightly better at repeating long syllable sentences than short syllable sentences. This result is different to Bley-Vroman and Chaudron (1994)'s suggestion that "because memory limitations are crucially involved, we expect accuracy when length is short" (p. 252).

Focus on meaning and learners' response in the EI test

To find out that participants had focused on meaning of the statements that they heard, four statements were created with the intention that participants would be likely to consider 'true' (1, 2, 3 and 4 in Table 10.9), and four which they would be more likely to consider 'not true' or which would elicit the response 'not sure' (5, 6, 7, and 8).

Table 10.9 presents the results for all participants (n=64) demonstrating that the 'belief' responses to the eight statements were indeed focusing on meaning as intended in the design of the test. (The first sentence could be a source of confusion for the learners as the first lexical item was not clearly produced by the native speaker).

Table 10.9 Learners' Responses

	Statements	True**	Not true / not sure**	no response**
1	*Muslims don't must serve Alcohol.	28	36	
2	*Women cannot driving in Saudi Arabia.	57	7	
3	*To get a better job, you must to work very hard.	42	20	2
4	Everyone must breathe oxygen to live.	46	18	
5	Students must pass an entrance exam to get an American visa.	15	45	4
6	*To stay healthy, you must not drinking water every day.	20	44	
7	King AbdulAllah will go to London next week.	14	50	
8	*The weather will remains hot forever.	10	52	2
** **number of participants indicating their feelings**				

***Effects of CF and oral tasks alone on free oral picture description* test**

Another outcome measure thought to tap into implicit knowledge is the free oral picture description test. The descriptive results presented in Table 8.5 indicated a development for all three groups over time.

Table 10.6 illustrates the overtime statistical changes within each individual group in all testing time. The scores for the metalinguistic information group significantly changed from pre-post-testing time but not on pre-delayed and post-delayed post-testing times. However, the scores for the recast group significantly changed from pre-post and pre-delayed post-tests but not from post -delayed post testing time. The scores for the task only group significantly changed on post-delayed post testing time only.

The above results suggest that the type of feedback students received along with the interaction tasks alone have some effects on learners' fluency of the target structure.

10.2.2 Effectiveness of CF on Explicit Knowledge in SA

A range of outcome measures were assumed to elicit more explicit knowledge such as gap fill, untimed grammaticality judgment and metalinguistic questions tests. Each outcome measure will be discussed separately. A summary of statistically

significant differences for the three groups on the outcome measures (more explicit) is shown in Table 10.10 and the different gains and direction (no statistical significant indication) are shown in Table 10.11.

Table 10.10 Summary of statistically significant differences within the group on outcome measures (more explicit) in SA using Wilcoxon statistical test

Type of measure	Tests	change over time	Groups		
			MI	R	TO
Explicit Measures	GF	Pre – post	Sig	Sig	Sig
		Pre – delayed	Sig	Sig	Sig
		Post – delayed	Not sig	Not sig	Trend
	Untimed GJT-Correct	Pre – post	Not sig	Sig	Not sig
		Pre – delayed	Not sig	Sig	Sig
		Post – delayed	Sig	Not sig	Sig
	Untimed GJT-Incorrect	Pre – post	Sig	Sig	Sig
		Pre – delayed	Not sig	Not sig	Not sig
		Post – delayed	Sig	Not sig	Not sig
	MQ	Pre – post	Sig	Sig	Sig
		Pre – delayed	Not sig	Not sig	Not sig
		Post – delayed	Not sig	Trend	Trend

Table 10.11 Summary of gains and directions for all groups on outcome measures (more explicit) in SA but statistical significant is not indicated

Type of measures	Tests	Pre - post gain	Pre-delayed post gain	Post-delayed post gain
Explicit Measures	GF- SA	18.65>12.5>10.86	18.26>9.29>6.43	-0.39> -1.57> -6.07
		MI>TO>R	**MI>R>TO**	**MI>R>TO**
	Untimed GJT-overall -SA	16.81>10.22>8.05	10.97>8.89>2.45	2.92>-1.33>-14.36
		MI>R>TO	**TO>R>MI**	**TO>R>MI**
	Untimed GJT-Correct- SA	11.55>7.78>5.85	16.11>11.11>-2.92	8.33>-0.44> -8.77
		R>TO>MI	**TO>R>MI**	**TO>R>MI**
	Untimed GJT - Incorrect-SA	26.23>8.89=8.33	6.48=6.67>5.83	- 2.50= - 2.22 > -19.75
		MI>R=TO	**MI=R>TO**	**TO=R>MI**
	MQ-SA	13.16>11.11>9.17	9.61>5.77>3.05	-3.55> -5.34>-6.12
		MI>R>TO	**MI>R>TO**	**MI>R>TO**

Effects of CF and oral tasks alone on gap fill test

The statistical results suggested significant overtime changes for all three groups in the three testing times indicating that both types of CF (metalinguistic information and recast) and the interactional tasks alone might have had positive effects on learners' achievement. The significant development for the three groups is similar to Erlam and Loewen (2010) who found both the CF types a long with the interaction activities alone have similar effects in developing the target structure. The significant overtime changes in this outcome measure for all groups could be attributed to the explicitness of this type of test or the explicitness of the different intervention. Although recast could be seen ambiguous (Egi 2007) the significant over time progression observed for the recast group might give the indication that this type of intervention was salient and thus helped learners to develop their own interlanguage. It is also possible that recasts result in explicit knowledge, as demonstrated in Long, Inagaki, and Ortega (1998) in which students learned the target structure 'Spanish adverb word order' through recasts and then were able to explicitly and correctly formulate an explanation of the rule. An explanation for the significant over time changes for the task only group could be attributed to the provision of the comprehensible input as suggested by Krashen (1982).

Effects of CF and oral tasks alone on overall results in untimed GJT test

The overall mean scores in the descriptive results on the untimed GJT test indicated that all groups have improved over time as shown in Table 8.7 suggesting beneficial role of both CF types and the interactional tasks alone. The mixed design statistical test ANOVA indicated a significant overtime changes for the three groups with a significant drop in scores witnessed for the metalinguistic information group between the post and delayed post-test whereas recast and task only groups maintained their scores.

The progression made by the three groups could be due to the beneficial role of the three types of intervention. However, the significant drop for the metalinguistic information group in the period between post and delayed post-test might be due to the fact that negative evidence (e.g., metalinguistic information feedback) does not help

learning and that structures learned through error correction cannot become part of internal grammar (Krashen, 1982). In contrary, the significant overtime progression for the two interactional groups (recast and task only) could lend support to Krashen (1982) Schwartz (1993) and Truscott (1996) who propose the sufficient role of positive evidence and comprehensible input in the acquisition of L2.

Effects of CF and oral tasks alone on grammatically correct items in untimed GJT

The results in Table 10.10 on the grammatically correct items in untimed GJT test indicated significant group differences for the three groups in the different testing time.

The statistical results showed no significant overtime changes for the metalinguistic information group on the period between pre-post- and pre-delayed post times but significant change observed in post-delayed post testing time. In contrary, the recast group had significant overtime changes between pre-post- and pre-delayed post testing times but no significant change observed between post-delayed post testing times. Opposite to recast, task only group had significant overtime changes between pre-delayed and post-delayed post testing times but not on pre-post testing time.

Although significant overtime changes witnessed in certain times for certain groups, the results pinpoint the usefulness of the three types of intervention; metalinguistic information, recast and interaction tasks alone.

When groups were compared, in the grammatically correct items of untimed GJT test, via the MW statistical test in post and delayed post testing times, there was no significant difference between the groups at post-test but at delayed post testing times as shown in Table 10.12. This finding is in line with Ellis et al (2006) whose study showed significant group differences evident in delayed posttest but not in post testing time.

Given that group differences were observed in the delayed post testing time on the grammatically correct items of untimed GJT test, the result indicated that task only group differed significantly from the metalinguistic information group.

The result of the two experimental groups indicated a significant group differences on the delayed post-test for the grammatically correct items of untimed GJT

where the recast group differed significantly from metalinguistic information group. This finding is different to Ellis et al (2006) who found metalinguistic information group significantly outperformed the recast group on delayed post-test for the grammatically correct items of untimed GJT.

Generally, the significant outperformance for the two interaction groups (recast and task only), in the grammatically correct items of untimed GJT test, over the metalinguistic information group in the delayed post-test time may support the sufficient role of interactional feedback along with the interactional activities. It might be in line with previous research suggesting the productivity of recasts in foreign language classrooms or laboratory contexts (Nicholas, Lightbown & Spada, 2001; Sheen, 2006; Li, 2010). The above result may also lend support to Mackey and Goo (2007)'s suggestion that interaction with feedback may not necessarily be more facilitative of L2 development than interaction alone.

Effects of CF and oral tasks alone on grammatically incorrect items in untimed GJT test

The statistical results shown in Table 10.10 for the grammatically incorrect items of the untimed GJT test reveal significant overtime changes for the three groups particularly the period between pre-post times. Although the metalinguistic information group evident an increase in the period from pre-post, a significant loss evident in the period between post-delayed post-test indicating that explicit type of feedback may help increasing knowledge temporarily followed by significant decrease. In other words, this result could suggest that information may have temporarily raised learners' awareness to detect and correct their erroneous productions. The finding may be in line with Schwartz (1993) and Truscott (1996) who suggest that structures learned through error correction cannot become part of internal grammar and so will be rapidly forgotten. It could also be argued that metalinguistic corrective feedback is more effective in short term.

The MW statistical result of the two experimental groups (shown in Table 10.12) indicated a significant group difference on the post-test for the grammatically incorrect items of untimed GJT where the metalinguistic information group differed significantly

from the recast group. This finding is different to Ellis et al (2006) who found no group differences neither on post-test nor on delayed post-testing times for the grammatically incorrect items of untimed GJT.

The insignificant group differences on the delayed post-test for the grammatically incorrect items of untimed GJT test for three groups is similar to Ellis's et al (2006) who also found no group differences on the delayed post-test for the grammatically incorrect GJT items.

Effects of CF on grammatical and ungrammatical items in the untimed GJT test

Based on the above results of the grammatically correct and incorrect items of the untimed GJT test, group differences were found (indicated in Table 10.12). In respect to the grammatically correct items of untimed GJT test, it was found that the interactional groups (task only and recast) significantly outperformed metalinguistic information group in delayed post-testing time. However, in the grammatically incorrect items of the untimed GJT, a significant outperformance for the metalinguistic information group over recast group in post-test time was found. The finding for this test is different to Loewen and Nabei (2007) who found an overall benefit to the incorporation of feedback over no feedback, but little differences between recasts, metalinguistic clues and clarification requests. They also found that learners did better on the ungrammatical items than the grammatical items on short-term indicating the validity of this test in measuring the explicit knowledge.

Effects of CF and oral tasks alone on metalinguistic questions test

The statistical results on Table 10.10 for the MQ test suggest a significant overtime development for all three groups in their explicit knowledge of English modals particularly on pre-post testing time. There were no significant internal overtime changes on pre-post and pre-delayed post but a trend toward significant loss evident on post to delayed post-test for the recast and task only groups.

It might be argued that the CF types and the interaction tasks alone might have provided an indication to learners that there is a gap between their production and the TL,

which in turn assisted them to work out rules (inductively in the case of recast and task-only groups). This argument could have been supported if group differences were found via MW statistical test.

However, in the current classroom study, the amount of practice needed to turn explicit knowledge into the state of automaticity, may not have been sufficient for gains to be maintained at delayed post-test (DeKeyser 2007).

Correcting errors and providing rules

The metalinguistic questions test required learners to provide corrections and state rules regarding the ungrammatical sentences. Table 8.11 displays that to some extent learners in all groups were able to correct the ungrammatical items and state rules.

For the metalinguistic information group, 52% of the sentences were corrected and 27% of them were given rules. The recast group corrected 35% of the sentences and stated rules for 17%. The task only group provided 41% correct forms for the incorrect sentences and were able to state rules for 23%.

This result is similar to Green and Hetch (1992)'s who gave three hundred German students a set of sentences containing grammatical errors and asked them to correct sentences and state the rules that were violated. They found that learners were able to correct 78% of the wrong sentences but state rules for 46% of the cases. An explanation for this discrepancy was given by Green and Hetch (1992) suggesting that learners' explicit rules form only a subset of their available implicit knowledge and that learner's ability to correct the errors exceeded their ability to explain rules.

Final comments

In this classroom study, clearly all students were exposed to CF, even though it may only have been directed to one individual. It is therefore possible that this 'passive feedback' may have influenced the findings, for example by priming the target structure or inducing explicit knowledge.

The usefulness of the recasts is not in line with arguments that recasts can be perceived ambiguously by learners (Egi, 2007). Instead, learners in the current study may have noticed the corrective functions of recasts (Lyster & Mori, 2006).

Learners' performance in task only group suggests that it was not the feedback alone that facilitated learning; the tasks that the students completed in the interaction sessions seem to have resulted in learning.

On a more general note, participants' significant performance, in the current study, with English modals runs against Saeed's (2009) whose university upper intermediate Arab learners of English (in the Department of English) had low performance with English modals though they had been previously exposed to English language for 12-14 years. However, potentially, Saeed's rationale and measures were different to those used in the current study and so this may account for the different performances observed (see Chapter 5 for more information on this study).

Summary

In summary, the experiment undertaken in SA compared the effectiveness of recast, metalinguistic information and interaction tasks alone on learners' performance on different outcome measures (broadly tap into more implicit and more explicit) in the form of EI, PD, GF, UGJT, and MQ tests.

Drawing on the above results (shown in Tables 10.6 and 10.10), it seems that all groups significantly developed their knowledge of English modals overtime in all measures thought to tap into more implicit and more explicit knowledge. The results suggested the productivity of the metalinguistic information and recast feedback on learning accuracy as well as engaging in interaction tasks alone.

One should bear in mind that the progression evident regardless of group and time in the experiment undertaken in SA might be of the novelty of these types of intervention.

When groups were compared to each other, there were no statistically significant group differences on post and delayed posttests testing times in most oral and written outcome measures but a few exceptions indicated in the MW statistical test as shown in Table 10.12 and inevitably discussed earlier in each outcome measure.

Given that measures of statistical significance do not necessarily inform the researcher about the importance or magnitude of the effect, the effect size (ES) represents a way to measure or quantify the effectiveness of an intervention, treatment or a program (Ledesma et al, 2009). Thus it might be of a great value for this thesis to pinpoint the

effect sizes of all outcome measures administered in both contexts (UK and SA). A brief discussion and a summary table will be presented in the following section.

Table 10.12 Statistical significant difference between groups at post and delayed posttests in SA using MW statistical test and the descriptive results

Tests	Post	D post	Statistical result
EI overall - SA	**MI>TO**		**Trend to sig**
EI -Correct - SA	**MI>R**		**Trend to sig**
EI -Incorrect - SA	**MI>TO**		**Trend to sig**
Untimed GJT - Correct- SA		**TO> MI**	**Sig**
		R>MI	**Sig**
Untimed GJT - Incorrect - SA	**MI>R**		**Sig**

Effect sizes on all outcome measures in the UK and SA

The effect size (ES) is a more precise way of summarizing the data (Wolf, 1986). It is a better indicator of the impact of the new teaching activity and can be obtained through a standardized measure of the difference between the means of the experimental groups and a control group (task only in the current study). For example the correlation coefficient Cohen´s *d* (Cohen, 1988) was applied in the current study for post and delayed post oral and written outcome measures administered in both UK and SA contexts, where *d* = (mean of the experimental group minus mean of control group) / (pooled standard deviation) as illustrated in the following equation:

$$d = \frac{mean_e - mean_c}{S_w}$$

Besides these statistical criteria, some practical rules for interpreting effect size have been suggested. For example, Cohen (1988) describes an ES value of approximately 0.2 as "small"; an ES value of 0.5 as "medium" and "large enough to be visible to the naked eye"; and an ES value of 0.8 as "completely noticeable and therefore large". Tables 10.13 and 10.14 summarise the results of the ES for each outcome measure administered in the UK and SA contexts.

Table 10.13 Effect sizes in the United Kingdom

	OUTCOME MEASURES	**Effect Sizes Cohen´s *d***	
		Metalinguistic Group	**Recast Group**
1	Picture description (Post)	0.10	* - 0.47
	Picture description (Delayed post)	* - 0.40	*-0.08
2	Gap Fill (Post)	* - 0.06	0.60
	Gap fill (Delayed post)	0.15	0.34
3	Grammaticality judgment (Post)	0.59	*- 0.25
	Grammaticality judgment test (Delayed post)	0.26	0.10
	Grammaticality judgment Correct Items (Post)	1.24	0.02
	Grammaticality judgment Correct Items (Delayed post)	*- 0.18	*- 0.26
	Grammaticality judgment Incorrect items (Post)	* -0.03	* - 0.26
	Grammaticality judgment Incorrect items (Delayed post)	*- 0.2	* - 0.22

Definition:

Trivial = 0.00 **-** 0.20, **small**=0.20 **-** 0.50, **medium**=0.50 **-** 0.80 and **large**= 0.80 **-** 2.00.

*****= **Negative Effects**

Table 10.14 Effect sizes in Saudi Arabia

	OUTCOME MEASURES	Effect Sizes Cohen´s *d*	
		Metalinguistic Group	**Recast Group**
1	Overall elicited imitation (Post)	0.60	0.30
	Overall elicited imitation (Delayed Post)	*-0.37	*- 0.30
	Elicited imitation correct items (Post)	0.50	*- 0.09
	Elicited imitation correct items (Delayed Post)	*- 0.19	*- 0.29
	Elicited imitation incorrect items (Post)	0.71	0.43
	Elicited imitation incorrect items (Delayed Post)	*- 0.19	*- 0.15
2	Picture description (Post)	0.09	*- 0.22
	Picture description (Delayed post)	*- 0.39	* -0.44
3	Gap Fill (Post)	*- 0.11	*- 0.37
	Gap fill (Delayed post)	0.08	*- 0.20
4	Overall untimed Grammaticality judgment (Post)	0.34	*- 0.07
	Overall untimed Grammaticality judgment test (Delayed post)	*- 0.51	* - 0.27
	Grammatically correct items GJT (Post)	0.12	0.20
	Grammatically correct items GJT (Delayed post)	*- 0.77	*- 0.22
	Grammatically incorrect items GJT (Post)	0.39	*- 0.26
	Grammatically incorrect items GJT (Delayed post)	*- 0.21	*- 0.22
5	Metalinguistic Knowledge (Post)	0.08	*- 0.35
	Metalinguistic Knowledge (Delayed post)	0.20	*- 0.33

Definition:

Trivial = 0.00 - 0.20, **small**=0.20 - 0.50, **medium**=0.50 - 0.80 and **large**= 0.80 - 2.00.

***= Negative Effects**

10.3 Effects of Tests

One of the limitations in the current study is not having a test only group. Evidence that test effect is not solely the cause of gains observed in the different groups can be justified in the following factors.

10.3.1 Test effects in the UK

For the UK groups, it is true that tests were repeatedly administered in three different testing sessions. The significant gains for the CF groups could not be due to a test effect because there were no significant gains for the task only group in all explicit measures and some implicit measures as shown in Table 10.1 and 10.3.

A strong test effect may also have been observed on the delayed post-test, yet significant decrease was observed on the delayed post-tests scores.

If gains were the effects of tests practice, learners would probably have been able to identify the target structure being tested on the exit questionnaires.

10.3.2 Test effects in SA

For the three groups in SA, the significant development in the experimental and task only groups in some outcome measures could arguably have been due to a test effect as there was no test-only control group.

However, in the exit questionnaire administered after each oral and written test on the three testing sessions, all learners were not able to identify the target structure being tested on the pre-test. On post and delayed post-tests, the CF groups scored higher percentage on the exit questionnaire than the task only group as shown in Tables 8.20-8.22.

If the gains were due to test effect, one might expect the 'third' test (delayed post-test) to produce the best results. However, in most measures in most groups gains were only observed between pre and post, and scores either stayed the same or decreased at delayed post-test.

10.4 Effects of Context

One of the rationales of this study is to examine the relative relationship of context and the types of CF. Two tests were replicated in the ESL and EFL contexts; free oral picture description and gap fill tests

Some research suggests that CF may be more salient in a foreign language setting, where the primary focus is on learning rather than on communication or content (Sheen, 2004; Ellis & Sheen, 2006; Loewen & Nabei, 2007; Lyster & Ranta 1997; Li, 2010).

Li (2010) found that studies conducted in foreign language contexts produced larger effect sizes than those in second language contexts.

It has also been argued that learners in foreign language contexts may have positive attitudes toward error correction more than learners in second language contexts (Loewen et al., 2009), which may make it more likely for the effects of feedback to be integrated.

The descriptive results in Chapter 9 (Tables 9.3-9.5) for the EFL and ESL contexts suggested slightly better production of English modals on **picture description** test in the UK compared to SA. The summary of the statistically significant differences across context shown in Table 9.6 suggested significant context difference for the metalinguistic information groups on pre-post and pre-delayed post-tests in the direction of the UK context. The ESL recast group was significantly better than the EFL recast group on pre-delayed and post-delayed post-tests. No significant context difference for the task only groups.

The descriptive results in the **gap fill** test (shown in Tables 9.7-9.9 and Figures 9.5-9.7) suggested context difference for the three groups. Table 9.10 showed significant context difference for the metalinguistic information groups in the direction of the UK on pre-post and pre-delayed post-tests but there was no significant context difference for the recast and task only groups.

Environment effects

The learners in the UK were exposed to English in schools, streets, host families, media and native friends, but the significant gains in the CF groups cannot be attributed

to the environment effects (e.g. maturation, or learning from outside the intervention) as if this was the case, it would have been happened for the task only group as well.

Generally, although minor differences in the replicated outcome measures were found between the groups in the UK and SA context there were no convincing patterns of difference between the effects of the CF and oral tasks in the two different contexts. Instead, the results indicate the positive contribution of the type of CFs to language development regardless of which context the experiment was in. However, these results should be taken with cautious as the number of participants in the ESL context was small.

10.5 Comparison of the Results of the Attitude Questionnaire in the UK and SA

Learners' positive attitudes could be a relative factor to the significant effects of the three types of intervention. The present study sought students' opinions concerning the different activities, CF techniques and error correction in the classroom and correlations between tests and these three constructs (presented in Chapters 7 (UK context) and 8 (SA context).

10.5.1 Learners' Attitudes

The descriptive results on the attitudinal questionnaire in UK and SA, sections 7.8.6, and 8.9.1 showed that all groups were equally in favour of the intervention activities, error correction and the different type of CF techniques and there was no significant group difference. The following sections discuss the correlations between tests' scores and scores of the three attitudinal constructs.

Learners' attitudes towards interactive activities

UK: The results of the questionnaire in the UK context indicated that 73 % of the participants found the activities interesting and 58 % of the participants pointing to the usefulness of the activities and the CF in improving their English.

SA: The results of the questionnaire in SA context suggested the usefulness of the interactive activities for 94 % of the participants found the activities interesting.

Learners' attitudes towards error correction

UK: As for the questions concerning the necessity of correction, 83.4% of learners preferred having their errors corrected, and 88.9 % think that error correction is absolutely the best way to learn English. This finding lends support to Schulz (1996) indicating that the majority of the students (90%) had a positive attitude towards error correction.

SA: As for learners' attitudes towards error corrections, 88% of the participants preferred having their errors corrected and 92 % of them indicated that error correction could be the best way to learn English.

Correction techniques preferred by students

UK: It seems that learners' preference for implicit CF 92% is higher than that for explicit type of CF 72%, the results on the relationship between tests scores and CF scores discussed below suggest a significant positive correlation for the metalinguistic information group on the gap fill test. This result might refer to the fact that preference does not always lead to gains.

SA: The relative association of the type of CF and learners' attitudes indicated that 80% of the participants preferred having rules given to them in accordance to their errors similar preference was evident for the implicit type of CF.

The high preference for all the three constructs among the three groups suggested the usefulness and the efficient role of these types of intervention in the EFL context as they were new for these groups of learners. Learners' positive preference could be supported by the significant gains for the three groups on the different outcome measures.

The correlation between tests' scores and learners' attitudes

The relationship between learners' attitudes and post-tests' scores for all outcome measures was presented in Chapters 7 and 8. The following discussion points to the correlation between tests' scores and learners' attitudes in both contexts.

Oral PD test and learners' attitudes

UK: There was no significant association between the scores of the PD test and the three attitudes constructs for the recast and metalinguistic information groups. For the task only group there was a significant correlation between attitudes towards error correction generally and PD test but no significant association between scores and attitudes towards the activities and the types of CF.

Of course, the non-significant association between test's scores and the type of CF for the task only group is likely to be due to the fact that there was no CF during the intervention sessions.

SA: Gains on pre-post PD test for the metalinguistic information was significantly associated with learners' opinions about the intervention activities indicating that learners were in favour of the interventional activities that include grammatical rules in a meaningful context rather than mere representation. Table 8.16 indicated no significant relationship between test's score, for the recast and task only groups, on the three attitude constructs. This might be arguably due to the ambiguity of the implicit CF and learners' disability to discover the target structure from the interactive activities or it might refer to the state of response (free time constrained) in this type of measure.

GF test and learners' attitudes

UK: Significant correlations suggested between the attitudes towards the CF scores and pre-post gains on gap fill for the metalinguistic information group but no significant association was found between the scores of any of the three attitude constructs and the pre-post gains for the recast group as shown in Table 7.15.

An explanation for the significant association between test and CF scores for the metalinguistic information suggested that learners were aware they were being corrected and thus they may have been able to relay on the rules they have learned throughout the

intervention tasks. The significant relationship between the explicit type of measure and the explicit type of CF suggested learners' **preference** for metalinguistic information (though learners' **preference** for implicit was higher); particularly they were used to be instructed via traditional teaching methods in their home town. In addition, the significant association between the explicit outcome measure and the explicit type of CF might speak of the strong validity of this test in tapping into more explicit knowledge.

There was a significant correlation between attitudes towards error correction and pre-post gains on gap fill test for task only group. The similar significant associations between pre-post gains on GF and PD tests and the scores on attitudes toward error correction for task only group suggest that learners may have realized that they were making errors and thus might have preferred being corrected as no CF was provided.

This conclusion could be supported by the results of the implicit type of CF(recast) in which no mediating effect was evident indicating that learners might not have been aware of being corrected.

SA: The results on Table 8.17 suggested no significant association between pre-post gains on GF for the experimental groups (metalinguistic information and recast) and learners' opinions about the three attitudes constructs although significant pre-post gains evident for these groups in the GF test. These gains might not be the results of learners' preference towards the CF, error correction or the interactive tasks. It might be argued that the teacher and the materials were crucial factors in learners' learning as learners valued the whole procedure they went through in the three months.

A significant association on pre-post GF gains and the scores on learners' attitudes towards the intervention activities for task only group might predict the usefulness of those activities especially that this group did not receive any CF during the four intervention sessions.

GJT tests and learners' attitudes

UK timed GJT: The results of the Pearson correlation analysis indicated no significant relationship between pre-post gains on GJT test and the scores of the three

attitudes constructs for the experimental groups (metalinguistic information and recast) and task only group as shown in Table 7.16.

It might be argued that learners, on testing sessions, might have benefited from the interactive activities and the types of CF but because of time pressure they were not able to rely on what they have learned. Time pressure was indeed a problem for learners as indicated in the exit questionnaire.

SA untimed GJT: Gains on pre-post GJT test for the metalinguistic information and task only groups indicated no significant association between tests' scores and any of the three attitudes constructs but there was a significant association of test's scores and the scores on learners' attitude towards the intervention activities for the recast group as shown in Table 8.18.

It might be assumed that the positive attitude for the recast group to the different activities helped learners to deduce their own rules from the recast models or the input.

EI test and learners' attitudes

SA: This test was only administered in SA. The correlation analysis for attitudes scores and pre-post gains indicated no significant association for all groups on any of the attitudes constructs as shown in Table 8.15 although the deceptive results indicated a development on the target structure for the three groups from pre-post-test. This result might possibly suggest that learners' performance does not always correlate with their attitudes.

MQ test and learners' attitudes

SA: The MQ test was part of the untimed GJT test which took place in SA only. The results for the three intervention groups indicated no significant correlation between the test gains and any of the three attitudes constructs for the metalinguistic information, recast and task only groups as shown in Table 8.19.

By and large, the descriptive results on pre-post gains for the three groups suggested learners' ability to apply what they learned to their tests regardless of their

attitudes towards the type of activities, the CFs, and error correction. This result lends support to Truscott's (1996).

10.5.2 Attitudinal disparity

The results of the questionnaires suggested some differences between participants' attitudes in the UK and SA relative to the different content areas.

Generally, results concerning the significant association between the battery of implicit and explicit tests administered in the UK and learners' preference were found for error correction and the CF intervention.

The association between the battery of implicit and explicit tests administered in SA and learners' preference was found for the interaction activities at most. This result suggests that EFL learners were in favour of focus on form language activities as this type of interaction is missing in their regular teacher-centred classrooms.

Given the fact the EFL learners were used to being introduced to grammatical rules in their regular language classes, learners were still in favour of metalinguistic information though equal preference for both type of CFs evident in this study.

Similar finding was displayed in Loewen et al, (2009) who found that EFL (e.g., Arabic and Japanese) were in favour of grammar instruction and error correction. They also found very few ESL learners like practicing or speaking in grammar instruction, but they were keener to improve communicative skills than were foreign language learners.

CHAPTER 11: CONCLUSION

The chapter summarizes the relative findings of the study. It is important to note that for readers ease, the research questions are reiterated in this chapter. The chapter points to the theoretical and pedagogical implications, limitations of the current study and several directions for future research.

11.1 Summary of findings

Drawing on the substantial amount of research that has established the effects, the frequency and the facilitative role of recasts in the classroom (e.g., Loewen & Philp, 2006; Nabei & Swain, 2002; Nicholas, Lightbown, & Spada, 2001; Philp, 2003; Lyster & Ranta, 1997; Sheen, 2004; Han, 2002; Mackey & Philp, 1998), and on the empirical research on the impact of explicit feedback over implicit feedback (e.g., Ellis et al., 2006; Sheen, 2007), the present study set out to investigate the effects of these two types of corrective feedback on a grammatical structure that is considered difficult for Arab learners of English in ESL and EFL contexts.

The findings in the current classroom study provided empirical support for the benefits of both implicit (recast) and explicit (metalinguistic information) CF techniques in L2 development, though the results in both contexts (UK and SA) have some differences and might be in line or different to other reviewed studies as discussed earlier in Chapter 10.

Although there were mixed statistical results in the different outcome measures (broadly more implicit and more explicit) administered in the UK and SA, a clear summary of the statistically significant differences and gains for all groups on all outcome measures reported in Tables (10.1-10.4) and Tables (10.6, 10.7, 10.10 and 10.11).

In general, the measures in the UK study indicated the usefulness of metalinguistic information and recasts for learning accuracy in pre-post and pre-delayed post-tests but not for oral tasks alone. The results also indicated significant overtime changes for each experimental group in most of the measures but not for task only group.

In SA context, generally, the three types of intervention were found beneficial for learning accuracy in the three testing times within each individual group. When the different testing times were compared within each group using the Wilcoxon statistical test, significant differences were found in certain measures and at certain time as shown in Table 11.1. The results also suggested a significant role of the interaction tasks alone for the learners in the SA context.

Although there has been some debate as to whether recasts are ambiguous (i.e., Lyster, 1998; Egi, 2007), and could be interpreted, for example, as a simple repetition of what was said rather than a correction (i.e., Mackey et al 2000). However, the learners' significant gains in the UK and SA do not support this view, and instead suggest that learners were able to use the recasts to improve accuracy. This result is compatible with Mackey and Philp 1998; Mackey and Oliver, 2002; Révész 2009 who's findings indicated the useful role of interactional recast.

The apparent effects of recast in the current study could be attributed to different factors; it could be because learners perceived recast as a reaction to the form, not the content, of their sentences (Long et al, 1998) and thus became salient for them so they were able to notice the mismatch between their erroneous utterances and the correct forms provided. It could be attributed to the repeated comprehensible input which provide a target reformulation and thus simultaneously offer positive evidence (e.g., Leeman, 2003). It could be due to the novelty of this type of technique particularly for this group of learners.

The results in both contexts suggest that metalinguistic information feedback (in line with Ellis et al, 2006; Sheen, 2006) made a significant contribution to the development of the target structure. This type of feedback might have assisted learners to locate the source of error in their production which in turn helped them to carry out the cognitive comparison and/or noticing the gap between their errors and target forms. Such a cognitive comparison is believed to be crucial for L2 acquisition (Ellis, 1994; Schmidt, 1990).

The effectiveness of the oral interaction tasks was observed in the SA (EFL) context, possibly because learners were in favour of this type of interactive activities for they were novel for them. The effective role for the interaction alone in SA is different to

other studies (e.g., Mackey & Philp 1998; Mackey & Oliver, 2002; Révész 2009; Yang & Lyster, 2010). In contrast, the task-only group in the UK made no significant gains possibly indicating that this type of intervention was normal for this group as it was part of their classroom communication and daily life. The insignificant development for the UK task only group in the different time of tests is different to Erlam and Loewen (2010) who found significant effects of oral interaction as well as the other types of intervention.

In regards to learners' preference in the different contexts, equal preference was found for metalinguistic information and recast feedback in the EFL context (80%), whereas a preference for recasts over metalinguistic information was found in the ESL (92%).

In terms of learners' attitudes towards the interaction tasks alone there were different preferences as the ESL learners scored lower (73%) than EFL learners (94%). This could be because practising and using the language in a native environment may have made the tasks familiar to the ESL learners (note also, there were no significant gains evident for the task only group in the UK context). Meanwhile, the absence of the oral interactive activities in regular classes in SA may have raised learners' desire to interact freely and made the experimental tasks more enjoyable due to their novelty.

Regarding error correction, learners in both contexts (UK and SA) indicated equal preference (89%) for their errors to be corrected.

The following table summarises the findings (discussed in Chapter 10) in relation to the research questions.

Table 11.1 Summary of findings

RQ1:	• Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic?
Findings 1	• Yes, all Saudi learners improved their learning accuracy of English modals over time on all outcome measures.
RQ1a:	• What is the effectiveness of these three intervention types relative to each other?

Findings 1.a

- In the **UK** context, the results indicated significant overtime changes for both experimental groups in most of the measures but not for the task only group.
- When groups were compared to each other in post and delayed post testing times, the **MW statistical** test indicated no sig group differences **except in two case**s:

a. In **posttest** of grammatically correct items of timed GJT test, metalinguistic information group differed significantly from recast and task only groups. This result is different to Loewen and Erlam (2006) whose study indicated no group differences in timed GJT. But it is in line with Sheen (2006) who found oral metalinguistic conflated with recast significantly outperformed the oral recast and the control groups in the immediate and delayed posttests although the study is different in the operation of the metalinguistic information feedback, the outcome measures and the function of the control group.

b. In **posttest** of GF test, the recast group differed significantly from the metalinguistic information group. This result is different to DeKeyser (1995) who found better performance for the explicit deductive condition than the implicit inductive condition in the fill-in-the-blank test.

- In **SA** context, the three types of intervention were found beneficial for learning accuracy in the three testing times within each individual group.
- When groups were compared to each other there were no statistically significant differences on post and delayed posttest in all measures but a few exceptions indicated in the **MW statistical** tests as follow:

a. In **posttests** of the overall and incorrect items of EI test, a trend to significant difference for metalinguistic information

group over task only group.

b. In **posttests** of the grammatically correct items of EI test**,** a trend to sig difference for metalinguistic information group over recast group. This result is different to Ellis et al (2006) who found no significant group differences on immediate posttest for either grammatical or ungrammatical items of EI test.

c. In **posttest** of grammatically incorrect items of the untimed GJT test, a significant difference evident for metalinguistic information group over recast group different to Ellis et al (2006) who revealed no group differences on immediate grammatical or ungrammatical items of untimed GJT test.

d. In **delayed posttest** of the grammatically correct items of untimed GJT test, task only group was significantly better than metalinguistic information group.

e. In **delayed posttest** of the grammatically correct items of untimed GJT test, recast group was significantly better than metalinguistic information group unlike Ellis et al (2006) who found significant differences for metalinguistic information group over recast on the delayed posttest for the grammatical items of untimed GJT.

RQ1.b

- Are any gains maintained after a delay of about 7 weeks?

Findings

- In response to the question, the study found that generally learners in the **UK** study did not significantly retain their knowledge in the period from post to delayed post-test with the one exception:

a. Recast group had significantly sustained knowledge of English modals in PD.

- In **SA** learners did not significantly retain their knowledge

but a few exceptions are indicated in the results:

a. Metalinguistic information group lost their knowledge significantly in the period from post to delayed post in untimed GJT (grammatically correct & incorrect items).
b. Recast group had a trend to significant increase in the grammatically correct items of EI but a trend to significant decrease in the grammatically incorrect items of EI and MQ tests.
c. Task only group had a trend to significant loss in GF and MQ whereas, a significant increase observed in PD, the overall score of the EI, and the grammatically correct items of the untimed GJT.

RQ1c

- Are gains observed differentially on different outcome measures?

Findings 1c

- Yes, there are gain differences on the different outcome measures (thought to tap into more implicit and more explicit knowledge).
- In the **UK**, gains of implicit outcome measures were less than gains of explicit outcome measures in all testing times.
- In **SA**, gains of implicit measures were slightly more than gains of explicit measures in all testing times.

RQ2

- Are results observed differentially in different contexts: EFL in SA, and ESL in the UK?
- The results in the **UK** and **SA** were found different for the metalinguistic information group in **PD** and **GF** in the period from pre-post and pre-delayed posttests.
- The recast groups in the **UK** and **SA** were also found significantly different in **PD** from pre-delayed post and from post-delayed posttests.
- Task only groups in the **UK** and **SA** had no significant

difference either on **PD** or on **GF** at all testing time.

RQ3

- What are learners' opinions about the different feedback techniques?

a. Do opinions differ according to the context in which the study was done?

Findings 3

- In terms of learners' perception towards error correction, learners in both contexts (**UK** and **SA**) had equal preference (89%) for their errors to be corrected.
- In terms of learners' preference for the type of CF, participants in the **UK** favored recasts but participants in **SA** liked both types of CF (metalinguistic information and recast) equally.

11.2 Implications

In Chapters 2 and 3, a number of theoretical and pedagogical issues, concerning the effective roles of CF in facilitating L2 development, were presented. In the light of the current findings, this section will first consider selected theoretical implications then the implications for language pedagogy.

11.2.1 Theoretical Implications

Again, it is emphasised that these studies did not aim to test any one particular learning theory. The design of the experiments can only suggest where findings are compatible or not with general theories. For example, the focused interaction activities, the types of corrective feedback, the opportunities and practice, and the output could not be teased apart from each other as they are essential components in a robust language environment.

The results relate to Long's interaction hypothesis in the following ways: In the SA context, the task only group made gains, suggesting that tasks which offer opportunities for negotiation and interaction aid learning. However, this was not the case

in the UK context. Long (1996, 2006) argued that recasts facilitate acquisition by drawing learners' attention to form throughout a conversational exchange that keep learners focused on meaning. The gains made by the recast groups in the UK and SA are in line with this (though this could be because the recasts provided implicit positive evidence and/or because the learners' construed their own, explicit, grammatical rules).

The comprehensibility and usefulness of the interventional materials as proposed by Krashen's Input Hypotheses may have helped learners to engage in the activities successfully and consequently improved their language fluency and accuracy.

Krashen claims that explicit correction of grammar would only improve explicit knowledge, which would not be accessible during certain tasks. The study suggested different results as the metalinguistic information feedback were found beneficial in certain tasks for both implicit and explicit knowledge in pre-post and pre-delayed post testing times.

In the current study, the beneficial role of the explicit feedback (metalinguistic information) is in line with Schmidt's (1995, 2001) noticing hypothesis.

The learners' gains observed in all groups in SA and in the CF groups in the UK were in line with the output hypothesis (Swain, 2005) which argues for the developmental benefits of pushed output. The learners' productions and the CF may have helped them to reformulate their initial utterances, monitor their production and hence produce accurate output.

The findings may also be compatible with skill acquisition theory (DeKeyser, 2007). The opportunities for automatisation provided during the intervention sessions might have been enough to convert some explicit knowledge into implicit knowledge (the strong interface position), as observed on some of the more implicit measures in which learners did not identify the target of the test.

Lastly, this study demonstrates the importance of examining the effectiveness of different types of CF in relation to language context. It was found that explicit and implicit CF can be equally effective in both ESL and EFL contexts. This differs from the trend reported by Li, (2010) whose meta-analysis research suggested a more effective role for the CF in EFL contexts than that in SL contexts. However, "the difference was not significant" (p. 338).

11.2.2 Pedagogical Implications

The current study investigated the relative effectiveness of implicit and explicit type of CF on the development of English modals which are difficult for EFL/ESL learners. The following implications might be adaptable and relevant to other language features, though further research would be needed to corroborate this.

It has been argued (e.g., Carpenter et al 2006) that recast might not be of value when learners fail to recognize it as a correction of an error. The significant results shown in the current study in both contexts for the recast groups are not in line with this suggestion. In this study, recasts were found beneficial for short- and longer-term learning in the UK and SA. Thus, language teachers could use this type of CF.

This study suggests that metalinguistic CF technique proved to be effective for EFL/ESL learners. This could run against the opinions of some teachers' dis-preference for direct and overt negative feedback believing that explicit feedback may result in embarrassment and demotivation of the learners (e.g., Seedhouse, 2001). The results in the current study suggest the importance of providing explanation in meaningful interactive activities. Further, in the current study, some learners expressed their cheerfulness in participating in the study; some came to my office in their spare time looking for more activities.

The positive attitudes towards the materials created for this project suggest that meaningful activities that suit the students' needs can be more effective.

Trainee teachers could be informed about these different feedback types in their methodological courses. The attitude and achievement data presented here suggests that interactional materials that focus on learners' interests and give them the chance to use the language freely in their communication can benefit accuracy. This is particularly important for EFL learners with which oral communication is almost absent. However it is acknowledged that classrooms in SA are big, thus the opportunity to talk and express ideas can be difficult to integrate into lessons. Whether such tasks and CF can be used effectively in larger classes is a matter for further research.

11.3 Limitations

It is important to acknowledge some limitations of this study.

- The sample sizes are quite small.
- The picture description test may have been more successful in eliciting implicit knowledge if it had been time constrained.
- The timed grammaticality judgment test (UK experiment) was originally intended to elicit implicit knowledge; unfortunately, because learners were asked to correct incorrect language, they probably referred to their explicit knowledge (Ellis, 2005). Even though participants indicated, in the exit questionnaire, their unfamiliarity of the target, their need for more time and their dislike of that type of test, the data from this test were used in the analysis of more explicit knowledge, meaning that there was only two measures of implicit knowledge in the UK experiment (correct items and the PD)
- Another limitation in this study is having the same outcome measures repeated in three different testing sessions. Even though learners' were not aware of the target structure in the exit questionnaire, and scores in delayed post tests for almost all groups were not significant, creating different 'versions' of the 'same' test would have been even more rigorous. Nevertheless, it is noted that an advantage of having identical versions of test in the three testing sessions is that the same level of difficulty is assured.
- It is regrettable that a test only control group was not used, as this would have helped to confirm the effectiveness of the interactional activities.
- Since the target structure selected in this study is considered difficult for Arab learners of English, more activities could have been included to elicit more productions. Consequently, more time might be needed for the treatment sessions.

11.4 Future Directions

Despite the limitations, this study may contribute to our understanding of the effect of CF and oral tasks in relation to different language contexts and learners' attitudes.

This dissertation aimed to examine the relative effects of CF to second language acquisition by beginning to address the following somewhat neglected areas: (1) using batteries of tests, thought to elicit different knowledge types; (2) it was administered in two different contexts (ESL vs. EFL); (3) implicit (recast only) and explicit (metalinguistic information only) CF techniques were provided and these were isolated ;(4) the use of a task only group with no CF; (5) a target structure that is considered difficult hitherto neglected in classroom CF research; and (6) learners' attitudes towards CF in relation to different contexts and test scores.

Some potentially fruitful directions for future research are as follows:

- Analysing learners' uptake during the interaction tasks to investigate the relationship between how learners respond to feedback and learning.
- Investigating the relationship between syntactic priming in response to recast or metalinguistic information.
- The inclusion of a test only group.
- An analysis of oral and written modes of production and correction.
- Further research is needed to document how Saudi learners of English use modals and how this progresses at different proficiency levels.
- Any relationship between task complexity and CF efficacy.

BIBLIOGRAPHY

Abunowara, A. (1996). *Modality in English and Arabic: A Study of the Difficulties Encountered by Arabic-Speaking Learners of English, Using A Systemic Functional Approach.* Cardiff: University of Wales.

Abunowara, A. (2005). Sa- and Sawfa as Modal Forms: New Evidence from Informant Testing. *International Journal of Arabic-English Studies, 6,* 85-92.

Alahmadi, B. (2007). The viability of computer-assisted classroom discussion (CACD) as a facilitator of communicative interaction. *The Jalt Call Journal, 3*(3), 3-32.

Al-Awadh, K. (2000). *The problem of teaching English in intermediate stage of the Qassim province as perceived by teachers.* MA thesis, King Saud University.

Alharbi, A. (2002). Verbal Modals. *Majalat Aljamia, 14*(9), 1-27.

Amador, Y. A. (2008). Learner Attitudes Toward Error Correction in A Beginners English Class. *Revista Comunicacion, 29*(1), 18-28.

Ammar, A. (2008). Prompts and recasts: Differential effects on second language morphosyntax. *Language Teaching Research, 12*(2), 183–210.

Ammar, A., & Spada, N. (2006). Size Fits All? Recasts, Prompts and L2 Learning. *Studies in Second Language Acquisition, 28,* 543-574.

Ancker, W. (2000). Errors and corrective feedback: Updated theory and classroom practice. *English Teaching Forum, 8*(4), 20-25.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Reviw, 89,* 369-406.

Anderson, J., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & & Qin, Y. (2004). An integratd theory of the mind. *Psychological Review, 111,* 1036-1060.

Anghelesca, N. (1999). Modalities and Grammaticalization in Arabic. In Y. Suleiman (Ed.), *Arabic Grammar and Linguistics.* Surrey: Curzon.

Bahns, J. (1982). Naturalistic L2 acquisition of the English modal auxiliaries. *TESOL Quarterly, 16*(4), 555-559.

Bargiel-Matusiewicz, K., & Bargiel-Firlit, A. (2009). Correction techniques preferred by students during the process of learning a foreign language. *Psychology of Language and Communication,* 39-52.

Batstone, R. (1994). *Grammar*. Oxford: Oxford University Press.

Bedell, D., & Oxford, R. (1996). Cross-cultural comparisons of language learning strategies in the People's Republic of China and other countries. In R. Oxford (Ed), *Language learner strategies around the world: Cross cultural perspectives.* Honolu:University of Hawai'i Press: Second Language Teaching & Curriculum Center.

Bialystok, E. (1978). A Theoretical Model of Second Language Learning. *Language Learning, 28*(1), 69-83.

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning, 29*(1), 81-103.

Bialystok, E. (1982). On the relationship between knowing and using forms. *Applied Linguistics, 3,* 181-206.

Bialystok, E. (1985). The compatibility of teaching and learning strategies. *Applied Linguistics, 6,* 255-262.

Bialystok, E. (1994). Representation and ways of knowing: Three issues in second language acquisition. In N. C. Ellis (Ed), *Implicit and explicit learning of languages* (pp. 549–569). San Diego: Academic Press.

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language. In E. Tarone, S. Gass, & A. Cohen(eds), *Research Methodology in Second-language* (pp. 245–61). New Jersey: Lawrence Erlbaum.

Bock, J. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355-387.

Bohannon, J., & Stanowicz, L. (1988). The issue of negative evidence: adult responses to children's language errors. *Developmental Psychology, 24*, 684-689.

Bowen, J., & McCreary, C. (1977). Teaching the English modal perfect. *TESOL Quarterly, 11*(3), 283-301.

Branigan, H. P., Pickering, M., Stewart, A., & and Mclean, J. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition, 28*(8), 1297-1302.

Butler, C. S. (1985). *Statistics in linguistics.* Oxford: Blackwell.

Bygate, M., Skehan, P., & Swain, M. (Eds.).(2001). *Researching Pedagogic Tasks, Second Language Learning, Teaching and Testing.* London: Longman.

Byrne, D. (1986). *Teaching Oral English.* Longman.

Carpenter, H., Jeon, K., MacGregor, D., & Mackey, A. (2006). Learners' interpretations of recasts. *Studies in Second Language Acquisition, 28*, 209–236.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition.* Amsterdam: John Benjamin.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: an empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition, 15*, 357-386.

Carroll, S., Roberge, Y., & Swain, M. (1992). The role of feedback in second language acquisition:Error correction and morphological generalization. *Applied Psycholinguistics, 13*, 173-198.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course.* New York: Heinle.

Chaudron, C. (1977). A descriptive model of discourse in the corrective treatment of. *Language Learning, 27*, 29-46.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.

Corder, S. (1974). 'Idiosyncratic Dialects and Error Analysis'. In J. Richard (Ed.), *Error Analysis: Perspectives on second Language Acquisition* (pp. 158-171). Essex: Longman.

Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161-170.

DeKeyser, R. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *Modern Language Journal, 77*(4), 501-514.

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17,* 379-410.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language. In C. Doughty, & J. Williams (Eds), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge: Cambridge University Press.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty, & M. H. Long (Eds), *Handbook of second language learning* (pp. 313–348). Oxford: Blackwell.

DeKeyser, R. M. (2007). Situating the concept of practice. In R. DeKeyser (Ed), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 1-18). New York: Cambridge University Press.

DeKeyser, R. M., & Juffs, A. (2005). Cognitive Considerations in L2 Learning. In E. Hinkel (Ed), *Handbook of Research in Second Language Teaching and Learning.* (p. 437). New Jersey: Lawrence Erlbaum Association, Inc.Mahwah.

Dilans, G. (2010). Corrective Feedback and L2 Vocabulary Development Prompts and Recasts in the Adult ESL Classroom. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes, 66*(6), 787–815.

Doughty, C. (2001). Cognitive underpinnings of focus on form. In Robinson. In P. Robinson (Ed), *Georgetown University roundtable on language and linguistics* (pp. 206-257). Cambridge University Press.

Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty, & M. H. Long(Eds.), *The Handbook of Second Language Acquisition* (pp. 256-310). Oxford: Blackwell.

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty, & J. Williams (Eds), *Focus on form in Classroom Second Language Acquisition.* Cambridge: Cambridge University Press.

Dunn, R., & Griggs, S. A. (1995). *Multiculturalism and learning style: Teaching and counseling adolescents.* Connecticut: Praeger Publishers.

Egi, T. (2007). Interpreting recasts as linguistic evidence:The roles of length and degree of change. *Studies in Second Language Acquisition, 29,* 511–537.

Egi, T. (2010). Uptake, modified output, and learner perceptions of recasts: learner responses as language awareness. *The Modern Language Journal, 94,* 1-22.

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition, 13*, 161–186.

Ellis, R. (1993). Second language acquisition and the structural syllabus. *TESOL Quarterly, 27*, 91–113.

Ellis, R. (2001). *Form-Focussed Instruction and Second Language Learning (ed.). Special issue of Language Learning.* Oxford: Blackwell.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*, 227-275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of second language: A psychometric study. *Studies in Second Language Acquisition, 27,* 141 – 172.

Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. In Z. Dornyei, & K. Bardovi-Harlig (Eds), *Themes in SLA Research. AILA* (Vol. 19, pp. 18-41). Amsterdam: John Benjamins.

Ellis, R. (2007). The differential effects of corrective feedback on two grammatical structures. In A. Mackey (Ed), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 339 – 360). Oxford: Oxford University Press.

Ellis, R. (2009). Corrective Feedback and Teacher Development. *L2 Journal, 1* , 3-18.

Ellis, R., & Barkhuizen, G. (2005). *Analyzing Learner Language.* Oxford: Oxford University Press.

Ellis, R., & Sheen, Y. (2006). Reexamining the Role of Recasts in Second Language Acquisition. *SSLA, 28*, 575-600.

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner uptake in communicative ESL lessons. *Language Learning, 51,* 281 – 318.

Ellis, R., Loewen, S., & Erlam, R. (2006). Implicit and explicit corrective feedback and the acquisition of L2 grammar. *Studies in Second Language Acquisition, 28*, 339-368.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension and the acquisition of L2 word meanings. *Language Learning, 44*, 449-491.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*, 464 – 491.

Erlam, R. (2009). The Elicited Oral Imitation Test as a Measure of implicit Knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 65-93). Clevedon: Multilingual Matters.

Erlam, R., & Loewen, S. (2010). Implicit and Explicit Recasts in L2 Oral French Interaction. *Canadian Modern Language Review*.

Farrar, M. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology , 28*(1), 90-98.

Field, A. (2009). *Discovering Statistics Using SPSS.* London: SAGE Publications Ltd.

Fitts, P., & Posner, M. (1967). *Learning and skilled performance in human performance.* . Belmont CA: Brock-Cole.

Gardner, R. (1985). *Social Psychology and Second Language Learning: The role of attitudes and motivation.* London: Edward Arnold.

Gass, S. (1988). Integrating research areas: a framework for second language studies. *Applied Linguistics, 9*, 198-217.

Gass, S. (2003). Input and Interaction. In C. Doughty, & M. Long (Eds), *The Handbook of Second Language Acquisition.* Oxford: Blackwell Publishing.

Gass, S., & Selinker, L. (2008). *Second Language Acquisition: An introductory course* . New York: Routledge.

Gass, S., Mackey, A., & Ross-Feldman, L. (2005). Task-based interactions in classroom and laboratory settings. *Language Learning, 55*, 575-611.

George, H. (1972). Common errors in language learning. *Rowley, Mass*.

Gibbs, D. (1990). Second language acquisition of the English modal auxiliaries can, could, may, and might. *Applied Linguistics, 11*(3), 297-314.

Graham, R., McGhee, J., & Millard, B. (2010). The Role of Lexical Choice in Elicited Imitation Item Difficulty. *Selected Proceedings of the 2008 Second Language Research Forum* (pp. 57-72). Somerville, MA: Cascadilla Proceedings Project.

Green, P., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics, 13*, 168–184.

Hall, J. K. (2007). Redressing the Roles of Correction and Repair in Research on Second and Foreign Language Learning. *The Modern Language Journal, 91*(4), 511-526.

Hamayan, E., Saegert, J., & Laraudee, P. (1977). Elicited imitation in second language learners. *Language and Speech, 20*, 86–97.

Hameyer, K. (1980). Testing oral proficiency via elicited imitation. *Revue de phonetique applique, 53*, 11–24.

Han, Y. (2000). Grammaticality judgment tests: How reliable are they? *Applied Language Learning, 11*(1), 177-204.

Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output . *TESOL Quarterly, 36* , 543 – 572.

Havranek, G., & Cesnik, H. (2001). Factors affecting the success of corrective feedback. In S. Foster-Cohen, & A. Nizegorodzew (Eds), *EUROSLA Yearbook* (Vol. 1, pp. 99-122).

Hendrickson, J. (1978). Error correction in foreign language teaching: Recent theory, research, and practice. *Modern Language Journal, 62*, 387-398.

Hinkel, E. (1995). The use of modal verbs as a reflection of cultural values. . *TESOL Quarterly, 29*(2), 325- 343.

Hu, G. (2002). Psychological constraints on the utility of metalinguistic knowledge in second language production. *Studies in Second Language Acquisition, 24*, 347–386.

Hulstijn, J. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. : . *Second Language Research, 18*, 193–223.

Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form of –te I (ru) by learners of Japanese as a foreign language. *Language Learning, 54*, 311–394.

Izumi, S. (2003). Comprehension and production processes in second language learning: In search of the psycholinguistic rationale of the output hypothesis. *Applied Linguistics, 24*, 168-196.

Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korea TESOL Journal, 4*(1), 1-15.

Kim, Y. (2009). *The role of task complexity and pair grouping on the occurrence of learning opportunities and L2 development.* A Dissertation, Northern Arizona University, Applied Linguistics, Arizona.

Kim, Y., & McDonough, K. (2008). Learners' production of passives during syntactic priming activities. *Applied Linguistics, 39*, 149-154.

Krashen, S. (1980). The input hypothesis. In J.E. Alatis (Ed), *Current Issues in Bilingual Education* . Washington, DC: Georgetown University Press.

Krashen, S. (1982). *Principles and practice in second language acquisition* . Oxford: Pergamon.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* Oxford: Pergamon Press.

Kreidler, C. (1999). *Introducing English Semantics.* London: Routledge.

Kubota, S. (2000). *Input Enhancement in Classroom Second Language Acquisition of Japanese.* Purdue University.

Larson-Hall, J. (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS.* New York: Routledge.

Ledesma, R. D., Macbeth, G., & Kohan, N. C. (2009). Computing Effect Size Measures with ViSta - The Visual Statistics System. *Tutorials in Quantitative Methods for Psychology, 5*(1), 25-34.

Leeman, J. (2003). Recasts and L2 development. *Studies in Second Language Acquisition, 25*, 25-37.

Lennane, B. M. (2007). *Cross-cultural influences on corrective feedback preferences in English language instruction.* Montreal: McGill University.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-Analysis. *Language Learning.* , 309–365.

Lightbown, P. M. (1998). Great Expectations: Second -Language Acquisition Research and Classroom Teaching Concordia University. *Applied Linguistics, 6*(2).

Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in Communicative Language Teaching: Effects on second language acquisition. *Studies in Second Language Acquisition, 12*, 429-448.

Lightbown, P., & Spada, N. (1993). *How languages are learned.* Oxford: Oxford University Press.

Liu, J. (2007). The place of methods in teaching English around the world. In J. Liu (Ed), *English language teaching in China: New approaches, perspectives, and standards* (pp. 13–41). New York: Continuum International Publishing Group.

Lochtman, K. (2002). Oral corrective feedback in the foreign language classroom: How it affects interaction in analytic foreign language teaching. *International Journal of Educational Research, 37*, 271–283.

Loewen, S. (2005). Incidental focus on form and second language learning. . *Studies in Second Language Acquisition, 27*(3), 361-386.

Loewen, S. (2009). Grammaticality judgment tests and the measurment of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip, & H. Reinders(Eds.), *Implicit and Explicit Knowledge in Second Language Testing and Teaching* (pp. 94-112). Clevedon, UK: Multilingual Matters.

Loewen, S., & Erlam, R. (2006). Corrective feedback in the chatroom: An experimental study. *Computer Assisted Language Learning, 19*(1), 1 – 14.

Loewen, S., & Nabei, T. (2007). The effect of oral corrective feedback on implicit and explicit L2 knowledge. In A. Mackey (Ed), *Conversational interaction and second language acquisition: A series of empirical studies.* (pp. 361-378). Oxford: Oxford University Press.

Loewen, S., & Philp, J. (2006). Recasts in the adult L2 classroom: characteristics, explicitness and effectiveness. *Modern Language Journal, 90*(4), 536-556.

Loewen, S., Li, S., Fei, F., Thompson, A., Nakatsukasa, K., Ahn, S., et al. (2009). L2 learners' beliefs about grammar instruction and error correction. . *Modern Language Journal, 93*(1), 91-104.

Long, M. (1991). Focus on form: a design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds), *Foreign Language Research in Cross-Cultural Perspective* (pp. 39-52). Amsterdam: John Benjamin.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie, & T. Bhatia, *Handbook of Second Language Acquisition* (pp. 413–68). New York: Academic Press.

Long, M. (2006). The story so far . In M. Long(Ed), *Problems in SLA.* Mahwah, NJ: Lawrence Erlbaum.

Long, M. H. (1980). Input, interaction and second language acquisition. . *Unpublished Ph.D. dissertation, UCLA.*

Long, M. H. (2007). *Problems in SLA.* Lawrence Erlbaum Associates.

Long, M., & Robinson, P. (1998). Focus on form: Theory, research and practice. In C. J. Doughty, & J. Williams(Eds), *Focus on form in classroom second language acquisition* (pp. 15-41). New York: Cambridge University Press.

Long, M., Inagaki, S., & Ortega, L. (1998). The role of negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal, 82*, 357–371.

Loschky, L. (1994). Comprehensible input and second language acquisition: what is the relationship ? *Studies in Second Language Acquisition, 16*, 303-323.

Lynch, T. (2001). Seeing what they mean: Transcribing as a route to noticing. *ELTJ, 55*(2), 124-132.

Lyster, R. (1994). The effect of functional-analytic teaching on aspects of French immersion students' sociolinguistic competence. *Applied Linguistics, 15*, 263–287.

Lyster, R. (1998). Recasts, repetition, and ambiguity in L2 classroom discourse. *Studies in Second Language Acquisition, 20*, 51-81.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *SSLA, 26*, 399–432.

Lyster, R., & Izquierdo, J. (2009). Prompts Versus Recasts in Dyadic Interaction. *Language Learning, 59*(2), 453–498.

Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition, 28,* 269 – 300.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition, 19,* 37-66.

Lyster, R., & Satio, K. (2010). Oral feedback in classroom in SLA: A meta-Analysis. *Studies in Second Language Acquisition, 32,* 265 – 302.

Mackey, A. (1999). Input, interaction and second language development: an empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21*, 557-588.

Mackey, A. (2006). Feedback, Noticing and Instructed Second Language Learning. *Applied Linguistics*, 405–430.

Mackey, A. (2007). *Conversational interaction in second language acquisition: A collection of empirical studies.* Oxford: Oxford University Press.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In A. Mackey (Ed), *Conversational interaction in second language acquisition: A collection of empirical studies* (pp. 407 – 452). Oxford: Oxford University Press.

Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System, 30*, 459-477.

Mackey, A., & Philp, J. (1998). Conversational interacion and second language development: Recasts, responses, and red herrings? *Modern Language Journal, 82*, 338-356.

Mackey, A., Gass, S., & McDonough, K. (2000). How do learners perceive interactional feedback? *Studies in Second Language Acquisition, 22*, 471-497.

Macrory, G., & Stone, V. (2000). Pupil progress in the acquisition of the perfect tense in French: the relationship between knowledge and use. *Language Teaching Research, 4*(1), 55–82.

Mandell, P. B. (1999). On the reliability of grammaticality judgment tests in second language acquisition research. *Second Language Research, 15*, 73.

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition, 46*, 53–85.

Marsden, E. (2005). Input-based grammar pedagogy: a comparison of two possibilities. *Language Learning Journal, 31*, 9-20.

Marsden, E. (2006). Exploring Input Processing in the Classroom: An Experimental Comparison of Processing Instruction and Enriched Input. *Language Learning, 56*(3), 507–566.

Masny, D. (1987). The role of language and cognition in second language metalinguistic awareness. In J. Lantolf, & A. Labarca(Eds.), *Research in Second Language Learning: Focus on the Classroom* (pp. 59–73). Norwood NJ: Ablex.

McDonough, K. (2001). Promoting Self-regulation in foreign language learners. *Clearing House, 74*(6), 323-326.

McDonough, K. (2006). Interaction and syntactic priming: English L2 speakers' production of dative constructions. *Studies in Second Language Acquisition, 28*, 179-207.

McDonough, K. (2007). Interactional feedback and the emergence of simple past activity verbs in L2 English. In A. Mackey (Ed), *Conversational Interaction in Second Language Acquisition* (pp. 323-338). New York: Oxford University Press.

McDonough, K., & Mackey, A. (2006). Responses to recasts: Repetitions, primed production, and linguistic development. *Language leaning, 56*(4), 693-720.

McDonough, K., & Trofimovich, P. (2009). *Using Priming Methods in Second Language Research.* New York: Routledge.

Mitchell, R., & Myles, F. (2004). *Second Language Learning Theories*. New York: Oxford University Press.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Morris, F. A., & Tarone, E. E. (2003). Impact of classroom dynamics on the effectiveness of recasts in second language acquisition. *Language Learning, 53*(2), 325–368.

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited Imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. Tarone, S. Gass, & A. Cohen (Eds), *Research Methodology in Second -language Acquisition* (pp. 227 - 245). Mahwah: NJ.

Nabei, T., & Swain, M. (2002). Learner awareness of recasts in classroom interaction:A case study of an adult EFL students' second language learning. *Language Awareness, 11*(1), 43-63.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *The Modern Language Journal, 77*(3), 330-339.

Naiman, N. (1974). The use of elicited imitation in second language acquisition research. *Working Papers on Bilingualism, 2*, 1–37.

Nassaji, H. (2007). Elicitation and reformulation and their relationship with learner repair in dyadic interaction. *Language Learning, 57*(4), 511-548.

Nassaji, H. (2009). Effects of Recasts and Elicitations in Dyadic Interaction and the Role of Feedback Explicitness. *Language Learning Research Club, University of Michigan, 59*(2), 411-452.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as feedback to language learners. *Language Learning, 51*, 719-758.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*, 417–528.

Nunan, D. (2004). *Task-Based Language Teaching.* Cambridge: Cambridge University Press.

Oliver, R. (1995). Negative feedback in child NS-NNS dyads. *Studies in Second Language Acquisition, 17*, 459-481.

Oliver, R. (2000). Age differences in negotiation and feedback in classroom and pair work. *Language Learning, 50*(1), 119-151.

Oliver, R., & Mackey, A. (2003). Interactional context and feedback in child ESL classrooms. *Modern Language Journal, 87*, 519–533.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and uptake in an adult ESL classroom. *TESOL Quarterly, 36*, 573–595.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism and second language acquisition. In N. Ellis (Ed), *Implicit and Explicit Language Learning* (pp. 393–419). London: Academic Press.

Perkins, K., Brutten, S. R., & Angelis, P. J. (1986). Derivational complexity and item difficulty in a sentence repetition task. *Language Learning, 36*, 125–141.

Philp, J. (2003). Constraints on "noticing the gap": Nonnative speaker's noticing of recasts in NS-NNS interaction. *Studies in Second Language Acquisition, 25*, 99-126.

Pica, T. (1994). Questions from the Language Classroom: Research Perspectives. *TESOL Quarterly, 28*(1), 49-79.

Pica, T., L. Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition, 11*, 63–90.

Rasinger, S. M. (2008). *Quantitative Research in Linguistics An Introduction.* New York : Continuum Internatinal Publishing Group.

Reid, J. (1987). The learning style preferences of ESL students. , 21/1, . *TESOL Quarterly, 21*(1), 87-111.

Révész, A. (2009). Task Complexiy, Focus on Form, and Second Language Development. *SSLA, 31*, 437– 470.

Révész, A. (2012). Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures. *Language Learning, 62*(1), 93–132.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *second language acquisition, 17*, 368–392.

Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for second language acquisition: A meta-analysis of the research. In J. Norris, & L. Ortega(Eds.), *Synthesizing research on language learning and teaching* (pp. 131–164). Amsterdam: Benjamins.

Sachs, R., & Suh, B. (2007). Textually enhanced recasts, learner awareness, and L2 outcomes in synchronous computer-mediated interaction. In . In A. Mackey (Ed.), *Conversational interaction in second language acquisition* (pp. 324–338). New York: Oxford University Press.

Saeed, A. (2009). Arab EFL Learners' Acquisition of Modals. *Research in Language, 7*, 75-98.

Samuda, V., & Bygate, M. (2007). *Tasks in second language learning.* New York: Palgrave Macmillan.

Sanz, C. (2003). Computer delivered implicit vs. explicit feedback in processing instruction. In B. VanPatten (Ed), *Processing Instruction: Theory, research, and commentary (second language acquisition research).* Mahwah, NJ: Lawrence Erlbaum.

Sauro, S. (2009). Computer-Mediated corrective feedback and the development of L2 grammar. *Language Learning & Technology, 13*(1), 96-120.

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

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed), *Attention and awareness in foreign language learning.* (pp. 1-63). Honolulu: University of Hawaii Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed), *Cognition and second language instruction* (pp. 1-32). New York: Cambridge university press.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language. A case study of an adult learner of Portuguese. In R. Day(Ed), *Talking to Learn: Conversation in Second Language Acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Schulz, R. (1996). Focus on form in the foreign language classroom: Students and teachers views on error correction and the role of grammar. *Foreign Language Annals, 29*, 343-364.

Schulz, R. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corretiv Feedback. *Modern Language Journal, 85*(2), 244-258.

Schwartz, B. (1993). Explicit and negative data effecting and affecting competence and linguistic behavior. *Studies in Second Language Acquisition, 15,* 147-163.

Seedhouse, P. (2001). The case of the missing "no" The relationship between pedagogy and interaction. *Language Learning, 51*(Supplement 1), 347-385.

Shaffer, B. (2004). Information ordering and speaker subjectivity: modality in ASL. *Cognitive Linguistics, 15*(2), 175-195.

Sharwood Smith, M. (1988). On the Role of Linguistic Theory in Explanations of Second Language Developmental Grammars. In S. Flynn, & W. O'neil (Eds), *Linguistic Theory in Second Language Acquisition* (p. 173). The Netherlands: Kluwer Academic.

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research, 8*(3), 263–300.

Sheen, Y. (2006). Exploring the relationship between characteristics of recasts and learner uptake. *Language Teaching Research , 10*, 361 – 392.

Sheen, Y. (2007). The Effect of Corrective Feedback, Language Aptitude, and Learner Attitudes on the Acquisition of English Articles. In A. Mackey (Ed), *Conversational Interaction in Second Language Acquisition.* (pp. 301-322). Oxford: Oxford: University Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Spada, N. (1997). Form-focused instruction and second language acquisition: a review of classroom and laboratory research. *Language Teaching, 30*, 73–87.

Spada, N., & Lightbown, P. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition, 15*, 205-224.

Spada, N., & Lightbown, P. (2009). Interaction in second/foreign language classrooms. In A. Mackey, & C. Polio(Eds.), *Multiple Perspectives on Interaction: Second Language Research in Honor of Susan M. Gass* (pp. 157 – 174). London : Routledge.

Stevenson, M. (1987). *English Syntax.* Boston and Toronto: Little Brown and Company.

Suzuki, M. (2004). Corrective feedback and learner uptake in adult ESL classrooms. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics, 4*(2), 1-21.

Swain, M. (1985). Communicative competence: some rules of comprehensible input and comprehensible output in its development. In S. Madden, & C. Gass (Eds), *Input and second language acquisition* (pp. 91-103). Rowley, MA: Newbury House.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review, 50*, 158-164.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds), *Principle and Practice in applied Linguistics* (pp. 125-144). Oxford: Oxford University Press.

Swain, M. (2000). The output hypothesis and beyond: mediating acquisition through collaborative dialogue. In J. Lantolf (Ed), *Sociocultural Theory and Second Language Learning* (pp. 97-114). Oxford: Oxford University Press.

Swain, M. (2005). The Output Hypothesis: Theory and research. In E. Hinkel (Ed), *Hand book of research in second language teaching and learning* (pp. 97-114). Mahwah, NJ.

Truscott, J. (1996). The case against grammar correction in L2 writing class. *Language Learning, 46*, 327-369.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Acquisition Research, 14*, 103-135.

Van den Branden, K. (2006). Training teachers: Task-based as well? In K. V. Branden(Ed.), *Task-Based Language Education: From Theory to Practice* (pp. 217–248).

Van Lier, L. (1988). *The Classroom and the Language Learner. Ethnography and Second-Language Classroom Research.* Harlow: Longman.

VanPatten, B. (2004). *Processing instruction: theory, research, and commentary*. Mahwah, NJ: Lawrence Erlbaum Association.

VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten, & J. Williams(Eds), *Theories in second language acquisition: An introduction* (pp. 115-135). Mahwah, NJ: Erlbaum.

Vethamani, M. E., Abd Manaf, U. K., & Akbari, O. (2008). Students' Use of Modals in Narrative Compositions: Forms and Functions. *English Language Teaching, 1*(1).

White, L. (1987). Against comprehensible input: The input hypothesis and the development of second language competence. *Applied Linguistics, 8*, 95-110.

Wolf, F. (1986). *Meta-Analysis: Quantitative Methods for Research Synthesis.* Newbury Park, CA: Sage.

Yang, Y. (2008). *Corrective feedback and Chinese learners' acquisition of English past tense.* Montreal: McGill University.

Yang, Y., & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners' acquisition of regular and irregular past tense forms. *Studies in Second Language Acquisition, 32*, 235– 263.

APPENDICES

Appendix A: Experimental Timetable

Week1	Consent Form

Pre-Tests

Week2	Oral Production Test One-to-one
	Elicited Imitation Test (SA only) One-to-one
	Gap Fill Test All participants
	Grammaticality Judgment All participants

Treatment Sessions

	Group	Day	Time
Week 3	Recast	Saturday	45 minutes
	Metalinguistic	Sunday	
	Task only	Monday	
Week 4	Recast	Saturday	
	Metalinguistic	Sunday	
	Task only	Monday	
Week 5	Recast	Saturday	
	Metalinguistic	Sunday	
	Task only	Monday	
Week 6	Recast	Saturday	
	Metalinguistic	Sunday	
	Task only	Monday	

Post-Tests

Week7	Oral Production Test One-to-one
	Elicited Imitation Test (SA only) One-to-one
	Gap Fill Test All participants
	Grammaticality Judgment & Attitudinal Questionnaire All participants
Week 8-13	Interval

Delayed Post-Tests

Week14	Oral Production Test One-to-one
	Elicited Imitation Test (SA only) One-to-one
	Gap Fill Test All participants
	Grammaticality Judgment All participants

Appendix B: Oral Communicative Tasks

B.1 First Session

First Day Session Includes Five Activities.

Activity (1)
Researcher: we will all listen to a conversation between a boy and a lady. The boy is telling the lady some of the things that he can and cannot do.
Researcher: Ok, we need to listen carefully to the conversation because you will answer some questions.

A conversation between a teacher and a student

Teacher: Can you use a computer, Josh?

Josh: Yes, of course I can. All my friends can. I use a computer at school and at home.

Teacher: That's very good. What other things can you do?

Josh: Well, I can run fast, very fast, and I can draw a bit. I can draw planes and cars very well but I can't drive a car of course. When I'm big I want to be a farmer and drive a tractor.

Teacher: And I know you can speak French.

Josh: Yes, I can. I can speak French very well because my dad's French. We sometimes speak French at home.

Teacher: Can you speak any other languages?

Josh: No, I can't. I can't speak German or Spanish, just French - and English of course! and I can cook! I can make cakes. My grandma makes lovely cakes and I sometimes help her. Yesterday we made a big chocolate cake.

Researcher: Ok, now please answer the following questions:

- Where do you think this conversation is taking place?
- Who are the main characters?
- Can you guess how old Josh is?
- Can Josh use a computer?
- Can his friend use a computer?
- What other things can Josh do?
- Can Josh drive a car?
- Can Josh speak Arabic?

- What languages can he speak?
- Can he cook? What can he make?

Activity (2)
Each participant will be given a package of words (in an envelope) and asked to use these words to tell his next partner what he can or cannot do. They need to listen carefully to each other. Then, each one will take turn to retell the rest of the class what his friend can/cannot do. For example, each package consists of the following words:

Song	**Swim Fast**	**Bicycle**	 **Flip Backwards**
 **Ski**	**Carrots Cake**	**Karate**	**Airplane**
 **Tennis**	**Train**	**Song**	**Rice**
Car	 **Golf**	**Run**	**Chicken**

Jet Ski	**Meat**	**Horse**	 **Hockey**
Bicycle	**Run Fast**	**Cake**	**Basketball** 
Chinese Food	 **Flip Forwards**	 **Tractor**	**Indian Song**

Activity (3)
In this excersise, each participant will be given a package of pictures (2 pictures at least) as seen in the following pages.

Researcher: **"**This week, Linda has been chosen the main character of our weekly magazine. I want each one of you to say something about her based on the pictures you have." Does any one of you want to start? I will go first.

Researcher: In my pictures I see a piano and a guitar. Linda is pointing to her piano. I think Linda can play a piano, but she cannot play a guitar. Or may be she can teach piano. Or she can play both of them. Oh, no I can see (X) signs so she cannot play a guitar but can play a piano. Now each one of you will take a turn to describe the pictures.

Note, students were expected to say the following:

Student 1: Linda can drive a car but she cannot ride a horse. Or she can do both.

Student 2: Linda can make cakes but she cannot cook. Or she can sell cakes.

Package 1

XXX

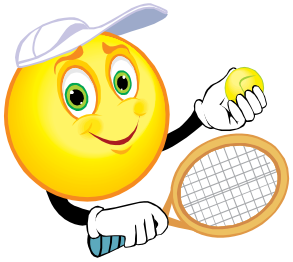
Package 2

Package 3

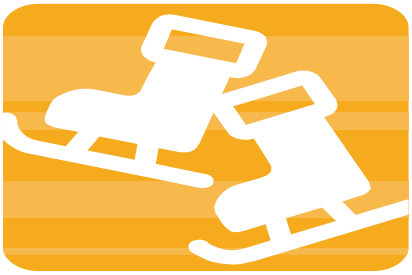
Package 4

Package 5

XXXX



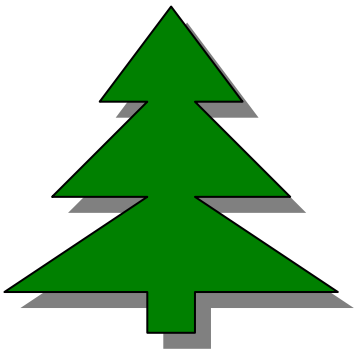
Package 6



Package 7

XXX

Package 8



XXX

Activity (4)

In this activity students will be asked to give suggestions to elicit text containing English modals.

1) Researcher: In this slip of paper, you will have some suggestions or tips for parents to help their children with their homework. Please read the tip in the slip, and then turn the paper to the other side.

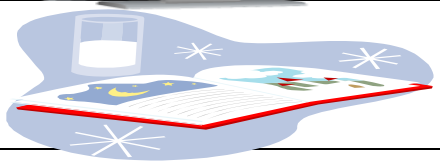
Parents can facilitate successful homework practices without becoming directly involved in the completion of assignments by providing a quiet, study space and ensuring that all required materials (books, paper, pencils, etc.) are available.
Parents can help with time management to ensure that children set aside time for homework and that the work is not put off until the last minute.
Parents can also help with workload management by encouraging their children to start with more difficult homework tasks, leaving easier tasks for the end of homework sessions when children are more tired.
Parents can also model attitudes and behaviours by expressing positive attitudes toward homework and doing "homework" at the same time as their children (e.g., reading, paying bills, doing other paperwork).
Too much interference from parents can reduce the beneficial effects of homework: learning how to work independently is an important lifelong learning skill that all children need to develop.
When children ask for help, parents can be most effective by helping children find answers for themselves rather than actually providing the answers.
When parents notice their children experiencing too much difficulty, parents should communicate with the child's teacher: teachers can provide the best advice on how parents can help and on what other sources of help (e.g., tutoring) are available.

2) Researcher: Now you need to give your suggestions to Ahmad who is a father for an 11 year old son, who needs our help in reducing the amount of time his son spent watching TV and using the internet. Each one of you could tell him some suggestions that can help in solving his problem similar to the one you have just read.

Activity (5)

Researcher: "Fatema is at home alone. She is bored. She does not know what to do to pass the time. You call her and she is telling you her problem."

A group of two will use the given pictures (in an envelope) to make a suggestion or suggestions to Fatema. Each student will be given a turn to say what suggestions he/she came up with.

PICTURE 1	
PICTURE 2	
PICTURE 3	
PICTURE 4	
PICTURE 5	
PICTURE 6	
PICTURE 7	
PICTURE 8	

B.2 Second Session

Second Day Session Includes Five Activities:

Activity (1)

Students will answer the following questions (Question and Answer format lead the pupils step by step to the outcome, Samuda & Bygate 2008.)

Researcher: What sports do you like?
Researcher: Are there certain rules that need to be followed when you are practicing that game?

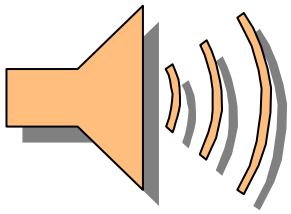
Researcher: What do you see in this picture?
Researcher: 'Now I want you to listen carefully to the rules that football players must follow. Then from your own experience or previous knowledge, I want you to tell me the rules that must be followed in playing tennis.' Each student has to mention at least one rule.

The following is the recorded material:

1- The players must not touch the football with their hands or arms.
2- The teams must not have more than 11 players playing on the field during the match.
3- Players must not wear anything which might injure another player.
4- A substituted player must not return to the game.
5- Players must not leave the game without the referee's permission.
6- Players must not hold an opponent.
7- If a player is sent off during the game, he must not be replaced.

Activity (2)

Researcher: I think all of you are living with a host family, aren't you?
Researcher: This means you are familiar with English costumes and ways of life.
Researcher: The Saudi Cultural Bureau is trying to issue a pamphlet including some of the rules that must be followed by students who live with host families.
Researcher: We want to think of some of those rules that could be included in this booklet.
Researcher: Each one of you will have some time to look at the pictures (a package of two pictures will be given to each student) and create rules that could be helpful.

Early 	**X** **Internet access** 	**X** **Loud music** 
X **mess** 	**TV** 	**X friends in the house** 

No Alcohol

Bathtub

Lock door

Tidy bed room

Late

Activity (3)

An application form will be brought to class and presented to the students. I will read a couple of the instructions after asking the following warming up questions.

- Before you came here, did you fill in any application form?
- Did you follow the instructions given in the application form?

Each student will be given one instruction to read silently, and then he /she will share what he has read with the rest of the class without looking at the paper.

An application must include the following elements before the company will accept it:

- the name of the applicant;
- a name and address for correspondence;
- a listing of the goods or services required; and
- the submission fee for at least one class of goods or services.

If your application does not meet these requirements, the company will return the application papers and refund any fees submitted.

If you submit a paper application, the company will assign a serial number and send a submission receipt. You should review this receipt for accuracy.

An electronically submitted application must include the same information to receive a submission date.

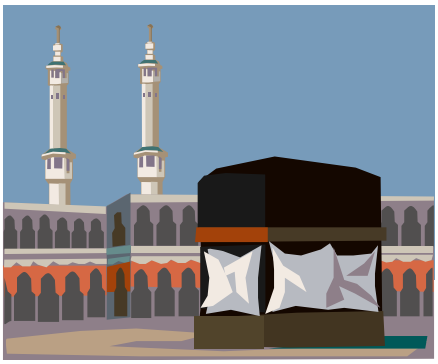
If through later review the company determines that the application did not include the required information, they will cancel the serial number, return the application, and refund the submission fee.

You must keep your mailing address up-to-date with the company.

Every application must include a clear representation of what you want to register.

You must list the specific goods/services for which registration is sought, regardless of the basis for the application.

Activity (4)



Ahmad is a new Muslim who has just converted into Islam. He wants to get into the Holy Mosque.

- What instructions would you give him?
- Each student will have one picture to give a rule or rules to Ahmad

1 	**2** 	**3** 
4 	**5** 	**6** 
7 		

B.3 Third Session

Third Day Session Includes Three Activities

Activity (1)

Warm-up questions

- Have you been to a restaurant?
- What kind of food do you like?

Researcher: Now, I want you please to listen carefully to the following conversation that took place in a restaurant between a customer and a waiter.

Researcher: A copy of a restaurant menu will be given to each group (a group of two students) and ask them to act out the conversation based on the items of the menu they have.

Recorded Conversation:

Waiter: Hi! How are you doing this afternoon?
Customer: Fine, thank you.
Waiter: Here is your menu.
Customer: What are your specials today?
Waiter: Our special today is Grilled Chicken Breast. It comes with a baked potato and a side salad.
Customer: Sounds good, but I'm going to take a look at the menu.
Waiter: OK, can I get you anything to drink?
Customer: Yes, I'd like a Diet Coke.
Waiter: Sure, I'll bring that right out. (Coming with the drink)
Waiter: Here you go. Have you had time to decide?
Customer: Yes. I think I'll have the New York steak.

Waiter: Oh, good choice. How would you like that cooked?

Customer: I'd like that well done.
Waiter: OK and you have a choice of potatoes with that.
Customer: I'll have fries, please.
Waiter: Your meal also includes a choice of soup or salad.
Customer: I think I'll have the salad.
Waiter: and what kind of dressing would you like with your salad?
Customer: I will have the salad with ranch dressing.
Waiter: I'll be right back with your salad**.**

Activity (2)

Researcher: I think most of you have just graduated from high school, have not you?

Researcher: What were your parents' concerns about going to college?

Researcher: What were your concerns?

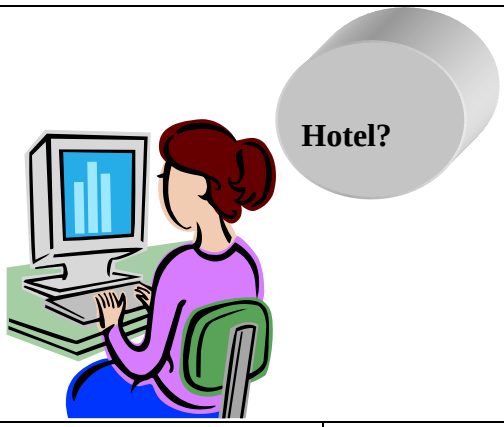
Researcher: Each one of you will be given some concerns and then you will share your concerns with the rest of the class.

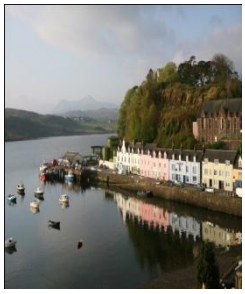
Parents can begin working on helping their child even before a child graduates from high school.
You can begin by collecting documentation you will need to complete the application. You will want to have the application completed and submitted.
You will also want to visit with your school counsellor. You will want to visit the colleges that have invited you to enrol.
College is an exciting time and as a fresher you may be wondering what you will need for your room.
Find out what is allowed and not allowed in the rooms as well. Some schools do not allow microwaves, hot plates, candles and more.
Moving in can be stressful and annoying so by talking to your roommate ahead of time you can find out what kind of person they are, what you have in common and what you each can bring or leave at home.
You will want to live frugally but there will be an occasion when you need your own money for those extra things.
By addressing all of your concerns before you arrive on your first day of college, you won't find this large life transition so difficult.

Activity (3)

Exercise (1)

Researcher: Enas and her family are going to the beach next weekend. She is thinking of ways to make the journey interesting and enjoyable. We will help her in planning the trip. Each one has to tell her what she will need to do according to the pictures you have in your package.





Canoe or Boat

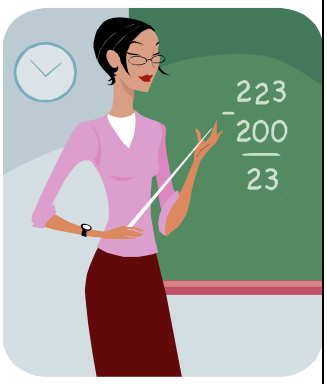
Friends and Games

Clothes, swimming suits

Food and Drinks

Sun Cream

Exercise (2)
Researcher: Do you usually get into class on time?
Researcher: Have you been late to class?
Researcher: What did your teacher do?
Researcher: Teachers like their students to come to class on time. I want each one of you to look at the pictures and tell me what your teacher will do if you come late to class.

<table>
<tr>
<td>1
X explain the lesson
 
</td>
<td>2 X participation
  </td>
<td>3
The head teacher’s office
  </td>
</tr>
<tr>
<td>4 X
 
</td>
<td>5
 
</td>
<td>6
X
 
</td>
</tr>
</table>

B.4 Fourth Session

Fourth Day Session Includes Four Activities

Activity (1)

Participants will be divided into two groups. Each group will come up with different ideas to share with the rest of the class. Then, they will be asked to put their instructions in proper order starting by the very important.

Researcher: You are a father or a mother for a 7 year old child, before you go out you will give your child some safety instructions. What will you say? Include the following words in your instructions.

Oven, (No)
Iron, (No)
Answer the phone, (No)
Door for strangers, (No)
DS game,
Computer

Activity (2)

Learners will be acting out the following:

Researcher: Pretend that you are a nutritionist. Your friend is an overweight who suffers from being fat. Give him/her some advice in order to lose weight and stay healthy by using the following.

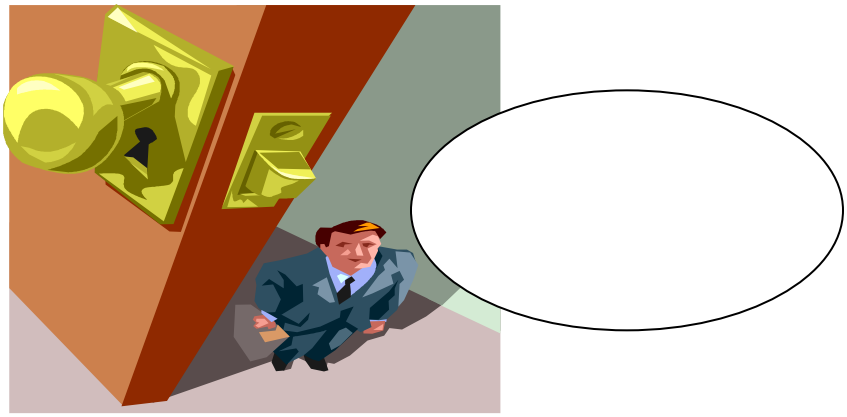
Water
Fresh air
Protein
Exercise
Sleep
Eat more greens
Junk food
Cake

Activity (3)

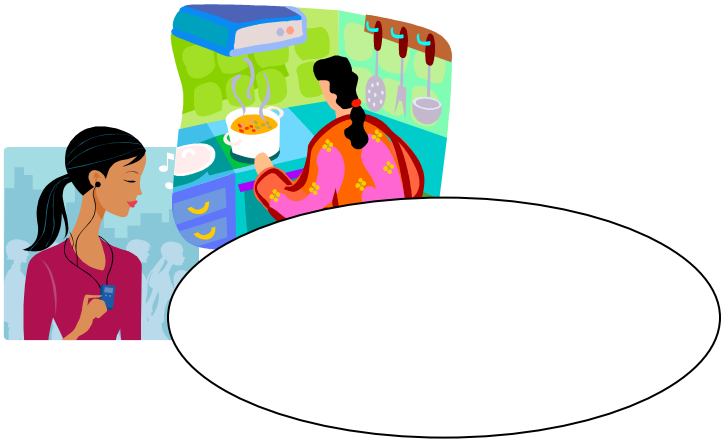
Each group will be given 18 cards. Three faced down sentences were assigned for each scenario. Learners will be asked to create a scenario by matching the picture with the very appropriate sentence. Then they will share their outcome with the rest of the class to discuss and find out the best selection for each scenario.

A:

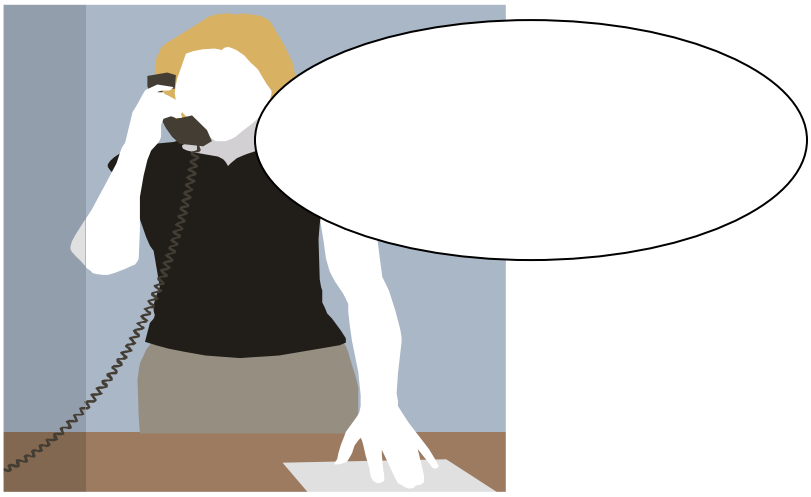
Mr John is reaching Mr Ali's house but the bell was broken so he starts knocking on the door but no answer. Finally, someone comes to the window. What do you think Mr. John says?



B:
Lora is a good girl. She always helps her mother in the kitchen. Now, she is offering her mother some help. What do you think she says?



C:
Mary is talking to her friend on the phone to apologize for not going to the party tonight. What do you think Mary is saying?

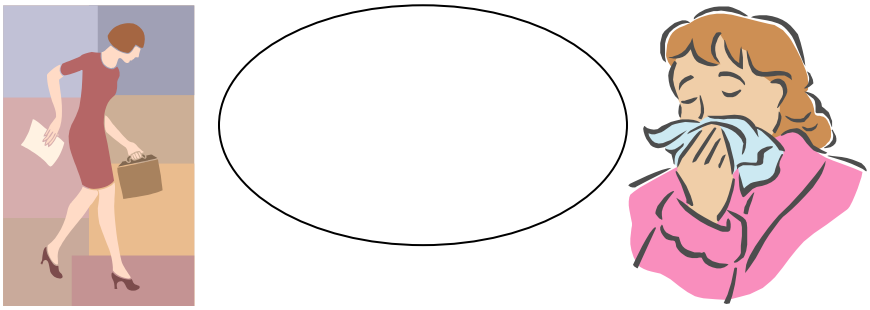


D:

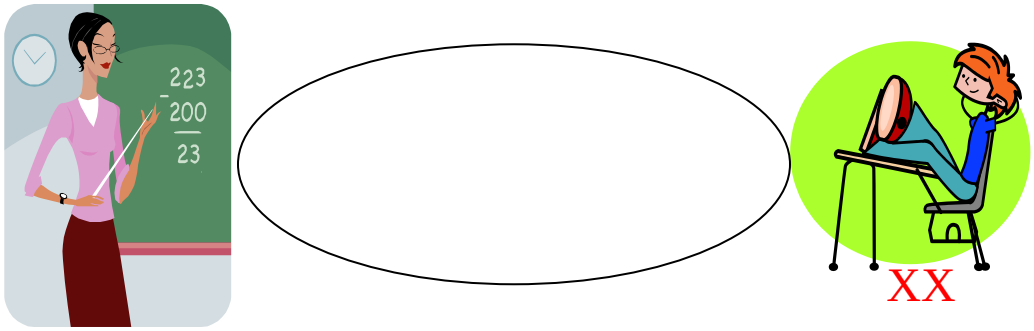
Mrs Smith is at the checkpoint counter in the airport. The clerk is asking her to present her ticket. What do you think he says?

E:

Linda is on the way home. She meets one of her closest friends 'Anna'.
Anna looks sick so Linda is advising her to see the doctor. What do you think Linda is saying**?**



F:
Henry is a horrid boy in his class. He puts his foot on the top of the desk.
His teacher reminds him about the classroom's rules. What do you think she tells Henry?



Sentences to be used with the scenarios above
Come open the door.
You should see a doctor.
You can see a doctor.
You must put your feet down.
Where is your ticket?
Show me your ticket.
Can I see your ticket, please?
I cannot come tonight.
I am so sorry to say that I cannot come to the party tonight.
Sorry, I will not come to the party tonight.
You should be in bed.
You will put your feet down.
You can put your feet down.
Do you want me to help you?
Mum, can I help you?
Can you open the door for me, please?
Will you open the door?
Do you need help?

Activity 4

In order to elicit more texts containing modals, each student, in this activity, will pick up one of the following questions. The main goal of this exercise is to give each student the chance to have a free production that targeted modals.

How will you spend the rest of the day today?
What will you do in Easter holiday, Christmas holidays' or whatever holiday is coming up? And where will you go?
What must you do in Saudi Arabia to obtain a driver's license?
Name three things that you couldn't do last year that you can do this year.
Can you do anything creative? For example, can you fix a computer, play a musical instrument, paint, draw, or write poetry?
What will you do when you graduate?
Can you mention some of the examination rules?
What can you do to improve your English?

Appendix C: Elicited Imitation Test

C-1 Questionnaire for Natives

- Have you figured out the target structure?
- Do you think the incorrect sentences are easy or difficult to notice?
- Do you think the sentences that you have heard comprehensible?
- Which sentences were difficult to judge?
- Do you the think the length of sentences is reasonable?
- Do you think one minute break is sufficient?
- Do you think 8 sentences for training purposes is sufficient?
- Do you think 50 sentences are too many for non native speakers?
- Was the voice clear?

C-2 Questionnaire for Non-Natives

- Have you figured out the target structure?
- Do you think you need to focus on meaning or forms or both, in your answers?
- Were there any difficult words? What are they?
- Do you prefer having a written version of this test rather than listening only? Or having both?
- Was repeating the instruction for the first 8 sentences helpful?
- Do you think the test needs to be changed? What kind of changes?
- Was the voice clear?

C-3 Length of Syllables and the Different Feelings

	TRAINING SENTENCES	Syllables	Feeling
1	*People have been using computers since 50 years ago.	13	NS
2	*In England, every child have to go to school.	10	T
3	*Nobody likes to watch terrified movie.	10	F
4	*It is harder to learn English from learning Japanese.	14	F
5	Swine flu causes death all over the world.	9	NS
6	A flood is more dangerous than an electrical shock.	14	T
7	The level of the Red sea is not rising.	9	F
8	Using the internet is the easiest way to get information.	15	T
9	*Girls plays football better than boys.	8	F
10	Children are not supposed to stay up late.	11	T
11	*Grandparents loving to play with their grandchildren.	12	NS
12	It is easy for disabled people to move around.	13	F

GRAMMATICAL SENTENCES TARGETING MODALS			
	Can/ Cannot	Syllables	Feeling
1	You cannot drive fast in the express lane on the highway.	12	F
2	The Iraqi People can move from one place to another very easily.	17	NS
3	Tourists can see the Pyramids if they go to Egypt.	10	T
4	I can buy a computer at a hair dresser's shop.	13	F
5	An overweight person cannot lose weight easily.	12	NS
6	Muslims cannot celebrate Christmas.	9	T

Must/ Must not			
1	Everyone must breath oxygen to live.	9	T
2	Women must pray 5 times a day.	8	NS
3	Women must follow men when walking in the city.	11	NS
4	Students must pass an entrance exam to get an American visa.	17	NS
5	You must not drive on a slippery road.	9	T
6	Muslim women must not be alone with a strange man.	12	T

Will/ Will not			
1	King AbdulAllah will go to London next week.	12	NS
2	The HSBC bank will give one million pounds to 100 costumers.	11	F
3	Old Saudi people and children will not travel abroad this summer.	17	NS
4	Barack Obama will not support poor countries.	12	F
5	Doctors will try their best to find medication for patients with swine flu.	16	T
6	Drivers will get fined for breaking the speed limit on country roads.	15	T

	UNGRAMMATICAL SENTENCES TARGETING MODALS	Syllables	Feeling
	Can/ Cannot		
1	*Chinese people can be eat rice with chopsticks very well.	12	T
2	*Women cannot driving in Saudi Arabia.	11	T
3	*The president of the United States can talks Arabic.	14	F
4	*School teachers cannot designed their own lessons.	12	NS
5	*Students cannot be mark their tests.	7	NS

	Will / Will not		
1	*The ministry of health will produced more vaccine for Swine Flu.	13	T
2	*Saudi Government does not will build new primary schools.	13	F
3	*Marks and Spencer will can have a big sale at the end of this year.	15	NS
4	*Microsoft wills provide a new program for deaf people.	13	NS
5	*The weather will remains hot forever.	9	F

	Must / Must not		
1	*To stay healthy, you must not drinking water every day.	13	F
2	*A driver must fastens a seatbelt when driving a car.	14	T
3	*Parents must don't leave their kids alone at home.	11	T
4	*To get a better job, you must to work very hard.	12	NS
5	*Muslims don't must serve Alcohols.	8	T

	DISTRACTOR SENTENCES	Syllables	Feeling
1	*Every child needs to have father.	8	T
2	* United States is big than Europe.	10	T
3	*Doctors does not like to save people's life.	10	F
4	*Scientists agrees that global temperature is not increasing.	16	NS
5	*There are few than 100 members of the United Nations.	12	F
6	*Rivers and stream formed from rain.	8	T
7	English is not an international language.	12	F
8	Pasta is a famous Italian dish.	10	T
9	Arabic is a difficult language to learn.	12	NS
10	The weather in Canada is always warm.	10	F
11	When a baby is born, he needs to be given a name.	11	T
12	Artists need a nice atmosphere to get inspired.	13	NS
13	Poets are always using their own experience to write poems.	14	NS

14	The sun rise in the west and sets in the east.	8	F
15	There are 1 billion planets in the galaxy.	9	NS

T: True

F: False

NS: Not Sure

*: Ungrammatical Sentences

C-4 Application Guidelines for Elicited Imitation

- To record, please un hold the recorder and press record and make sure that the recorder is functioning.
- First, ask the participant to identify herself, and to clearly write out her name level, and the date of the test on the top of the answer sheet.
- Tell the student that she needs to listen carefully to the instructions.
- You can summaries the instructions in three words (listen, tick and repeat). Please make sure that the student does not repeat the sentence before judging it.
- You can repeat the sentence back to her, for the first 12 training sentences but not the rest.
- For the first 12 practice sentences, you need to remind the student with part of the instructions- "Remember, you must use the best, most correct English you can. Please do not change the meaning of the sentence".
- If a student changed the meaning, for the first 12 sentences, tell her that she does not need to do that and you can model it for her.
- Tell the student that she can skip the sentence to the next one if she cannot say it back to you.
- In the answer sheet, make sure that the student is in the right order of the sentences and she is following the recorder.
- If a student needs to know the meaning of a sentence in Arabic, you can translate it to her.
- You need to end up your meeting with an exit questionnaire.

C-5 Test Format

NAME: **LEVEL:**

DATE:

INFORMATION

1- **You don't need to worry about grades in these tests.**
2- **Your participation is appreciated and of a great value to understanding how we teach and learn English.**
3- **We need to record what you say so we can listen to it again for the purpose of the research.**

INSTRUCTIONS

1- **Listen to each sentence carefully.**

2- **Decide whether the sentence is True, Not True, or Not Sure. Note your decision by ticking in the space provided on the sheet.**

3- **Then, say the sentence you have just heard to the researcher in correct English. <u>Remember, you must use the best, most correct English you can.</u>**

4- **Please <u>do NOT change the meaning</u> of the sentence when you say it back, even if you think the sentence is not true.**

5- **So, repeat the meaning of the sentence using good English!**

The first 12 sentences are practice sentences, so you understand what you have to do. After you have said each practice sentence back to the researcher, the researcher will repeat the instructions.

After these 12 training sentences, the researcher will not comment on your response!

You will get a break of 1 minute after 30 sentences.

Thank You for Your Participation.

Note to the reader:

The test consists of 60 sentences as follows:

	Type of sentences	Sentence numbers
1	(12) Training sentences	1 to 12
2	(33) Targeted modals (can, will and must), (18) correct and (15) incorrect sentences	14, 16, 17, 18, 19, 20, 22, 24, 26, 28, 29, 31, 32, 33, 34, 35, 37, 38, 39, 41, 42, 44, 46, 47, 48, 50, 51, 52, 54, 55, 57, 59, 60
3	(15) Distractor sentences targeted different structure.	13, 15, 21, 23, 25, 27, 30, 36, 40, 43, 45, 49, 53, 56, 58
	* = incorrect	

1. *****In England, every child have to go to school.
2. Children are not supposed to stay up late.
3. *****People have been using computers since 50 years ago.
4. Swine flu causes death all over the world.
5. *****Girls plays football better than boys.
6. It is easy for disabled people to move around.
7. *****Grandparents loving to play with their grandchildren.
8. The level of the Red Sea is not rising.

9. *****Nobody likes to watch terrified movie.

10. A flood is more dangerous than an electrical shock.

11. *****It is harder to learn English from learning Japanese.

12. Using the internet is the easiest way to get information.

13. United States is big than Europe.

14. Tourists can see the Pyramids if they go to Egypt.

15. Artists need a nice atmosphere to get inspired.

16. Students must pass an entrance exam to get an American visa.

17. *****School teachers cannot designed their own lessons.

18. The HSBC bank will give one million pounds to 100 customers.

19. *****The ministry of health will produced more vaccine for Swine Flu.

20. Women must pray 5 times a day.

21. English is not an international language.

22. *Muslims don't must serve Alcohol.

23. There are 1 billion planets in the galaxy.

24. An overweight person cannot lose weight easily.

25. Pasta is a famous Italian dish.

26. *****Marks and Spencer will can have a big sale at the end of this year.

27. *Rivers and stream formed from rain.

28. *****To stay healthy, you must not drinking water every day.

29. *****Saudi Government does not will build new primary schools.

30. The weather in Canada is always warm.

31. Muslim women must not be alone with a strange man.

32. *****Chinese people can be eat rice with chopsticks very well.

33. Old Saudi people and children will not travel abroad this summer.

34. *****Parents must don't leave their kids alone at home.

35. Muslims cannot celebrate Christmas.

36. *Scientists agrees that global temperature is not increasing.

37. Women must follow men when walking in the city.

38. *****Microsoft wills provide a new program for deaf people.

39. *****Women cannot driving in Saudi Arabia.

40. Arabic is a difficult language to learn.

41. Drivers will get fined for breaking the speed limit on country roads.

42. I can buy a computer at a hair dresser's shop.

43. When a baby is born, he needs to be given a name.

44. *****A driver must fastens a seatbelt when driving a car.

45. Doctors does not like to save people's life.

46. You cannot drive fast in the express lane on the highway.

47. *****To get a better job, you must to work very hard.

48. Barack Obama will not support poor countries.

49. The sun rise in the west and sets in the east.

50. Everyone must breathe oxygen to live.

51. *****The president of the United States can talks Arabic.

52. King AbdulAllah will go to London next week.

53. *There are few than 100 members of the United Nations.

54. You must not drive on a slippery road.

55. Doctors will try their best to find medication for patients with swine flu.

56. *Every child needs to have father.

57. *****Students cannot be mark their tests.

58. Poets are always using their own experience to write poems.

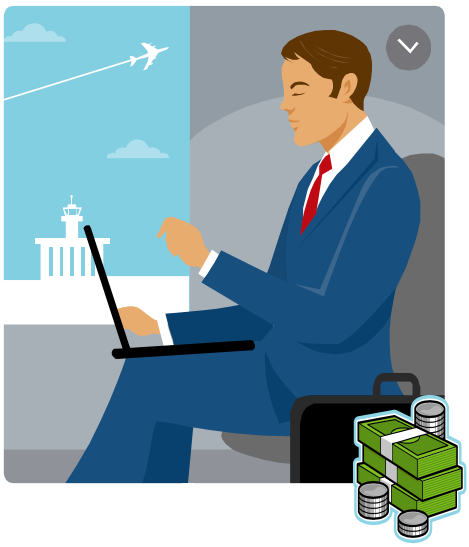
59. *****The weather will remains hot forever.

60. The Iraqi People can move from one place to another very easily.

Appendix D: Free Oral Picture Description

What are the differences in lifestyle between Mr Bell and Jack?

Mr Bell is a rich business man



Jack is a poor window cleaner

Nice car

Big house

Travel

Nice hotels and restaurants

Shopping

Nice clothes

Laptop

What are the things that Lana is able to do but Lolo is not? Use the following words to describe the pictures.

Lolo is a baby

Lana is 8 years old

Read

Watch TV

Play with snow and friends

Write

Go to school

Eat chips and chocolate

Every person has wishes in his/her life. Describe your life in ten years time from now.

Full of adventure

A millionaire

A doctor or an engineer

A teacher

A mother or a father

A designer

A traveller

The police are investigating a murder that has just happened. The officer is giving orders to people around. Tell me the instructions he gives using the following words.

Handcuffs

Keep a way

Touch the gun

Evacuate the road

Call an ambulance

Move the body to the hospital

Find the killer

Describe these two ladies by using the words provided.

Lora is a disabled lady

Anna is a secretary at the Department of English

Do any kind of physical activity

Walk

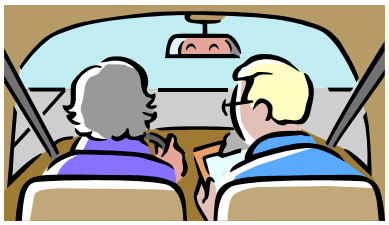
Play football

Sit in any chair

Climb Mountains

Swim

To drive a motorcycle or a car there are certain rules to be followed. What are they?



Helmet

Fasten seat belt

Cell phone

Car seat for children

Speed limit

Car insurance

Traffic lights

You are planning to go to Disney Land. What are the preparations needed before you go?

Call a travel agent

Reserve a hotel

Transportation

Clothes

Food

Maps and books

Camera

Appendix E: Gap Fill

Assessment Test (1)

Name:

Nationality:

Age:

Level:

Date of Exam:

Time of Exam:

Grade:

Instructions to the Participants:

1- **You do not need to worry about grades in these tests.**

2- **Your participation is appreciated and of a great value to understanding how we teach and learn English.**

3- **You need to read the instructions carefully before you start answering the questions.**

4- **You need to answer all questions.**

5- **If you have difficulty understanding some words, don't hesitate to ask the instructor.**

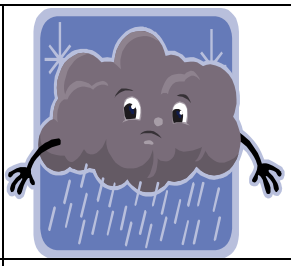
6- **Concentrate on your paper. You do not need to look around, and do not turn back to the previous pages.**

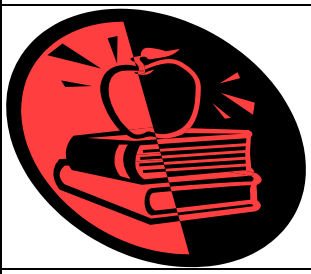
7- **Distributed pens are allowed only. No liquid, no eraser or anything else.**

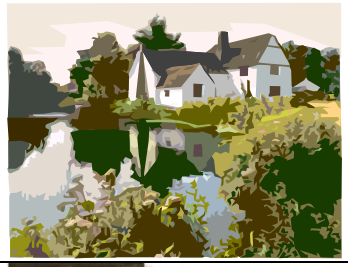
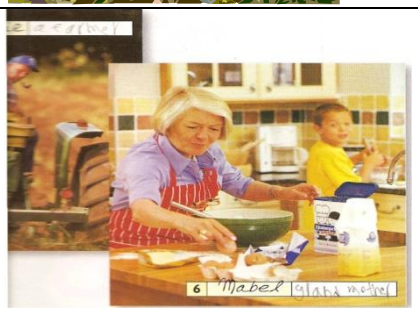
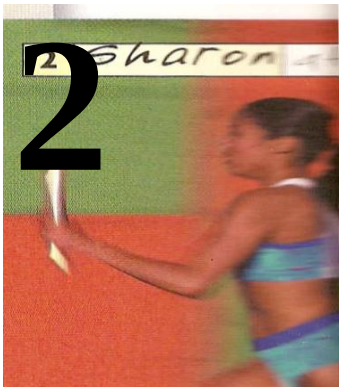
Thank You for Your Participation

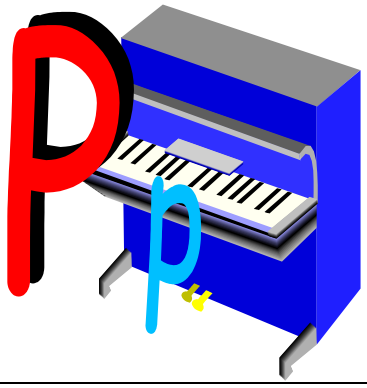
Note to the reader:
The test consists of 22 sentences (14) targeted Modals 1,3,4,6,8,9,11,13,15,16,18,20,21,2 and (8) distracter sentences 2,5,7,10,12,14,17,19

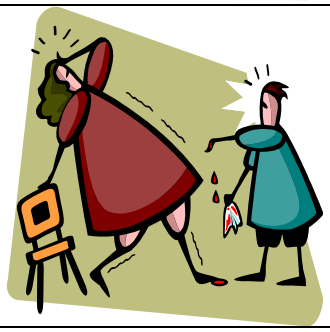
Based on the pictures, fill in the following blanks with the appropriate words. Some gaps need to be filled with one word, some with two or more, and some blanks do not need to be filled.

1		It has been raining for three days, I am not sure if it _____ tomorrow.
2		How old _____ you?
3		Wow I _____ the moon!!
4		This picture is very expensive, you _____ it.
5		Maths _____ very _____ for most students.

6		The receptionist: _____ I_____ you?" The customer: "Yes, indeed, I need someone to help me with my luggage."
7	 **English**	He has already _____ that course.
8		Shhh! The baby is sleeping. You _____ quiet.
9		I am really good at climbing: I _____ very high.
10		This is good! I was so _____

11		"Speak up please, I _____ you."
12		They have lived in this house since 1998_____.
13		Ahmad is talking to his friend: "I am going to the Lake District next weekend, _____ me?"
14		Grandma _____ cake and grandpa _____ the garden.
15		I am sure Sharon_____ first.

16		Exit 52 is temporarily closed. To reach the central part of the city, you _____ exit 56.
17		How many brothers do you have_____?
18		One of the examination rules is you _____ around.
19		In this picture, the letter **P** _____ for piano.

20		This box is very heavy for me, but Michael is very strong, he _____ up to 40KG.
21		Before you get into your head's office you _____ on the door.
22		Do you think the lady_____ him to the hospital?

Appendix F: Grammaticality/Ungrammaticality Judgment Test

Grammaticality Judgment (Test 1)

Name:

Nationality: Saudi

Age:

Level:

Date of Exam:

Time of Exam:

Grade:

Note to the reader:
The test consists of (9) correct sentences targeted English modals (can, will and must) 6, 10, 24, 26, 28, 32, 34, 36, 38, and (9) incorrect sentences targeted English modals 4, 8, 9, 12, 14, 17, 19, 21, 39.
There are 21 distractor sentences targeted different structures 1, 2, 3, 5, 7, 11, 13, 15, 16, 18, 20, 22, 23, 25, 27, 29, 30, 31, 33, 35, 37

Instructions to the Participants:

1- You don't need to worry about grades in these tests.

2- Your participation is appreciated and of a great value to understanding how we teach and learn English.

3- You need to read the questions carefully before you start answering them.

4- You need to answer all questions.

5- Concentrate on your paper. You do not need to look around, and do not turn back to the previous pages.

6- Please write out any corrections you think are necessary. Just write out the CORRECTION ONLY. DO NOT write out the whole sentence again.

7- Please provide ANY RULES to any errors you think you have found.

8- If you have difficulty understanding some words, please ask the instructor.

Thank You for Your Participation

In this activity, you are asked to judge whether the following sentences are right, wrong or not sure. <u>Underline</u> any part or parts you think are wrong. <u>Write out any corrections</u> you think are needed and <u>provide any rules</u> to any errors you think you have found as illustrated in the example below:

Example already completed by a student:

	Sentences	**Right**	**Wrong**	**Not sure**
1	**Ali <u>cleans</u> his room yesterday.**		**X**	
	Correction? *cleaned*			
	Rule/s for error/s?	**The verb must be in simple past tense.**		

	Sentences	Right	Wrong	Not sure
2	**Does Hanna speak English fluently?**			
	Correction?	**X**		
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
1-	My house is at the end of this street.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
2-	My brother is a soldier. He is in army.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
3-	Lisa is a youngest student in her class.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
4-	The book isn't in right now but we order can one for you.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
5-	My brother like to play with his friends.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
6-	I will not be doing my homework tonight.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
7-	California is a nice place. Many people go there for a holiday.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
8-	Must people to put on a helmet when they ride a bicycle?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
9-	Where will can you go this summer holiday?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
10-	Why must students follow the school's rules?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
11-	I eat my food very slow.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
12-	If Josh goes to bed early, he wills feel better in the morning.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
13-	We always go to the same place. Let's go somewhere else.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
14-	Can Tom has a piece of chocolate cake?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
15-	George jacket is black but his T-shirt is white.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
16-	Last year, my dad go to Paris.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
17-	He must goes to the post office. He needs some stamps.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
18-	Where are you going next weekend?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
19-	The players will not playing tennis next week.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
20-	English schools start at 9.00 am and end at 3.15 pm.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
21-	It’s a terrifying film. Your son is must not see it.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
22-	The telephone was invented by Alexander Bell in 1876.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
23-	She didn’t eat anything, but she drank a little water.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
24-	Can we stay with your brother when we are in Paris?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
25-	I know Mrs Watson but I don’t know his husband.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
26-	When you are driving, one of the rules is "You must fasten your seatbelt."			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
27-	Sara want to go with her mother to the party.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
28-	Jane and Allen cannot come to the party.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
29-	What sort of music he like?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
30-	Did you call your mother last night?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
31-	There is some pictures in this classroom.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
32-	I like this hotel room. You can see the mountains from the windows.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
33-	Don’t cook that meat. It doesn’t smell good.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
34-	Do you think there will be a lot of people at the party on Saturday?			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
35-	I'm tired this morning. I did not sleep well last night.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
36-	We must not forget to turn off the lights before we leave the house.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
37-	When I was having breakfast, the phone suddenly rings.			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
38-	‘Oh! My bag is very heavy.’ ‘I will carry it for you.’			
	Correction?			
	Rule/s for error/s?			

Item	Sentence	Right	Wrong	Not Sure
39-	I do not can speak English.			
	Correction?			
	Rule/s for error/s?			

Appendix G: Questionnaires

G.1 Exit Questionnaire in the UK

Name: _____ Date: _____

Please answer the following question briefly.

Now that you have finished the test, what do you think this test was about? Why have you written that?

Thank You for Your Participation.

G.2 Exit Questionnaire in SA

Name: _____ Date: / /

Please answer the following questions briefly.

1- What do you think I was testing in that test? What was the test about?
2- Did you think about any rules during the test; what was the rule; could you give me the rule now that you were thinking about during the test?
3- Can you give me any examples?

Thank You for Your Participation.

G.3 Attitudinal Questionnaire

Attitudinal Questionnaire

Name: _____

Date of Birth: _____

Gender: ___ Male _____ Female

Homeland: _____

Institution: _____

Level: _____

How many years have you been studying English?

Where did you study English before coming here? (A school? language institute? a private tutor? or a combination of these?)

_____
_____
_____

Did you have any contact with any English-speaking people after school? When? Before or now?

□ Yes □ No

If yes, when? _____
Who was it? _____

How did you communicate with him / her?

□ Face to face □ Skype

□ MSN □ others

Have you been to any English –speaking countries before (e.g., UK, USA, Canada, etc)?

□Yes □ No

If yes,

Where? _____

When? _____

How long did you stay there?

_____

How long have you been studying English in here?

_____

How many classes do you, now, have per week (once a week, twice, etc)?

Now, do you have any contact with native English speakers after school?

□Yes □No

How do you communicate with him / her?

□ Face to face □ Skype

□ MSN □ other

How often do you talk / write to him /her in English?

_____

Tell me your feelings about what we have been doing in these sessions. Please honestly circle one (ONLY) of the numbers that best describes your feelings. Thank you for your time.

#	Sentences	1	2	3	4	5
		Strongly disagree	disagree	uncertain	agree	Strongly agree
1	The activities are interesting.					
2	The activities are not up to my level.					
3	The activities are easy.					
4	The activities are short.					
5	I feel it is my teacher's duty to correct my errors all the time.					
6	I feel frustrated when you correct me.					
7	I feel better when you give me the rules.					
8	I feel discouraged when I repeat the same errors.					
9	I feel nervous about speaking after you have corrected my errors.					
10	I feel it is better for me to know the corrections of my errors.					
11	I feel that I am not used to being corrected when I do grammatical mistakes.					
12	I feel that this way of correction is new for me.					
13	I am benefitting from your corrections.					
14	Having my errors corrected is the best way to learn English.					
15	I feel most comfortable with your direct corrections.					
16	The corrections you have been providing are not important.					
17	I prefer providing me with rules and information.					

18	I think the most helpful way is correcting my errors directly.					
19	I need a lot of time to think about my mistakes.					
20	I need to finish the activities fast so I can attend my other classes.					
21	What you are doing does not improve my English.					

Appendix H: Students' Production

H-1 Recast Group

Students non-native like production	Teacher's correction	Students subsequent utterance
*S1: my friend cannot song.	T: he cannot sing a song	S1: he cannot sing a song.
		S1: he cannot swim (Continuation)
		S1: he can ride a bicycle (Continuation)
		S1: cannot flip backwards (Continuation)
*S2: you must to ask him	T: you must ask him	S2: you must ask him
		*S2: he must to give him
	T: you must give him	S2: You must give him special program.
*S3: I will learning English	T: so you will learn and study English	S3: My parents concern first comes here
*S4: he will has	T: he will have	S4: he will have a big car
*S5: he will shopping for food	T: he will buy some grocery	S5: yeah, he must buy some chicken

H-2 Metalinguistic Group

Students non-native like production	Teacher's correction	Students subsequent utterance
*S1: He can play flip.	T: Okay after can, after modals, we have to have one verb and flip here is a verb.	S1: Flip! He can, he can flip backwards.
*S2: He can't driver tractor.	T: We have to have a verb after modals.	*S2: He can't driving. He can't drive.
	T: we need to have a simple form of the verb.	S2: He can't drive. * S2: Uh-huh. She can driver her and she S2: verb? *S2: She can't. she can't rode a horse.
	T: A simple form of a verb after modal.	S2: Please example.
	T: the simple form of the verb is an inflected verb, no change, without endings, without anything (e.g., ing, s or ed).	S2: She can't ride a horse.
*S3: Ahmed the father stop his son because he watch TV too much and if he watching too much and use computer too much again, this maybe *can wearing sunglasses.	T: We have to have a simple form of the verb after can	S3: wear glasses, this I think problem for child. Put him in the room.
*S4: sometimes they do that. I think there is one rule when you go to bath room you must to wear slippers or something like that.	T: With must we do not use *to*. Ok? 'Must' comes with no *to*. We use the main verb only.	S4: you must wear slipper.
*S5: you must when you listen to music you must do not make it loud.	T: with must we do not use do auxiliary. We use one main verb only	S5: Hum. You must make the music low and you must smoke outside the house and you must keep the bed tide every morning and the room keep it clean.
*S6: You must took off	T: with must we use simple form of the verb not the past	S6: take off? you must take off your shoes for enter the holy mosque and you must turn off mobile

		phone
*S7: he will not have a big dinner or big meal just snack or something light, and when he got to the hotel, he *will made dinner.	T: You need to have a simple form of the verb after modals	S8: he will make
*S7: I think he will not participation with	T: A gain we need to have a verb, abase form of the verb after will T: participation is a noun, ok? T: yes.	*S9: he will not participation *S7: he will not be participation
S7: with do?		S7: do? S7: what part of speech participation? Noun?
S10: participation, participate?	T: We need to have a verb.	S10: he will not participate with the teacher and he will stand like this, and will not understand the lesson all the time.

H-3 Task Only

Students non-native like production
S1: the most important thing. They must manage the house well. The parents do not have strong personality.
S2: you must pay for Ahmad any play.
S3: if they can register them in any sport club.
*S4: I think I will said to Ahmad he should to be near to his son. The first day do the homework with him and the second day far from him. When he need help he will help him about the TV because make it problem with his son. And about the internet he said *you must used the internet in the weekend or when you have free time.
T: so he has to set a time for him. Yeah?
S4: yeah and he must do timetable for his son
*S3: Ahmad is should do a timetable daily for do anything everyday for do his homework and talking with his parents and help his mother or his father and do any sport and will he. He will be busy at every day.
T: what do you think?

*S5: You must should make a lot of friends.
T: Do you think this might help?
S3: He will know a lot of friends.
S3: He can draw.
*S4: He should phones his parents.
S5: If he has he can go to his friends. I suggest him stay home, play station or watch TV.
S6: He can work in a restaurant. He will know lots of people
S4: Where is his parents? What about when he works? He will know a lot of people he will spend his time, and when he come back to home, he will be tired.
*S6: He should makes a lot of friends and go to the library, *he should phones his friends, and in his free time *he can phones his friends.
*: incorrect items

Appendix I: Scoring Guidelines

I-1

Tests	0 Mark	0.5 Mark	1 Mark
Grammaticality Judgment (GJT)	Ticking right for incorrect item	underling and ticking wrong only	underlining and writ out the correction
	No answer indication	writing the correction only	ticking wrong and writing out the correction
	Ticking not sure	ticking wrong only	underlining, ticking and writing out the correction
Gap Fill (GF)	no obligatory context, or no verb after modal, e.g., *S: will first winner *S: can 100 kg. *S: must exit 216.	Modal plus an inflected verb (s, ing or ed), insertion of be or do or addition of to, or semantically incorrect modals or incorrect verb. *S: It will be rain tomorrow. *S: You must don't look around. S: You will be quite. S: You cannot pay a car.	grammatically and semantically correct modal and verb S: It will rain tomorrow. S: You must be quite.

I-2

Tests	0 Mark	0.5 Mark	1 Mark
Elicited Imitation	create an obligatory occasion but no main verb, e.g., *S: Mothers cannot in charismas. *S: Parents must children home. *S: Parents is must don't the children in the home. *S: Saudi government will not children to school.	Obligatory context + inflected verb, insert be or don't and or adding to. For example: *S: Students must to pass an entrance exam. *S: Marks and Spencer can be have a big sale. *S: Mothers cannot be play. *S: The driver have must pass it.	Obligatory occasion + verb correctly produced regardless of other lexical items for example, S: Muslim women mustn't stay with strange man. S: All Saudi people will not go to ground S: Parents must not let their children
Free Oral Picture Description (PD)	create an obligatory occasion but no main verb, e.g., *S: Must a seatbelt. *S: can a house. *S: Must a traffic light.	creating an obligatory context plus an inflected verb, inserting be, do or adding to, e.g., *S: Mr. Jack cannot shopping. *S: He can has a big house *S: You must don't use cell phone.	creating an obligatory context with modal plus correct verb, e.g., S: Anna can read, S: she can write, eat

I-3

Test	0 mark	.05 mark	1 mark
Metalinguistic questions (MQ) Students were required to state rules for any incorrect sentences.	Rule: S: It's a sentence not a quiz S: grammar S: it's not a perfect tense	Rule : S: it's a rule *S: pecause (ing)	Rule: *S: models come before the verb *S: No 2 models S: Can with ability S: we cannot use (can ,do) together *S: We must put the subject before the model and the verb S: Use should for give advice
Note, the students' production in the MQ test is exactly the same as it was written in the participants' test paper			

Appendix J: Normality Tests and Baselines

Table J.1 Normality, Baseline, Tests Required, and Research Questions for GF and GJT in SA

Saudi Arabia									
No	Test		Shapiro	ANOVA	K. W	Base-line	Normality	Test Required on Data	Research Questions
1	GF	Pre	P<0.01		0.64	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	RQ1 Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	P<0.01				NN		
		Delayed post	P<0.01				NN		
2	GJT	Pre	0.63	0.78		same	N	Raw scores RM-ANOVA+ Planned Contrasts for interaction time*group	
		Post	0.21				N		
		Delayed Post	0.77				N		

Table J.2 Normality, Baseline, Tests Required, and Research Questions for GJT-Cor and GJT-Inc in SA

Saudi Arabia									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
3	GJT-Cor.	Pre	P<0.01		0.63	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	RQ1 Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	P<0.01				NN		
		Delayed Post	P<0.01				NN		
4	GJT-Inc	Pre	P<0.01		0.48	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	
		Post	0.13				N		
		Delayed Post	P<0.01				NN		

Table J.3 Normality, Baseline, Tests Required, and Research Questions for PD and EI in SA

Saudi Arabia									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
5	PD	Pre	P<0.01		0.29	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	RQ1 Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	P<0.01				NN		
		Delayed Post	P<0.01				NN		
6	EI	Pre	P<0.01		0.42	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	
		Post	P<0.01				NN		
		Delayed Post	P<0.01				NN		

Table J.4 Normality, Baseline, Tests Required, and Research Questions for EI-Cor and EI-Inc in SA

Saudi Arabia									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
7	EI-Cor	Pre	P<0.01		0.57	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	RQ1 Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	P<0.01				NN		
		Delayed Post	P<0.01				NN		
8	EI-Inc	Pre	P<0.01		0.10	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	
		Post	P<0.01				NN		
		Delayed Post	P<0.01				NN		

Table J.5 Normality, Baseline, Tests Required, and Research Questions for MQ in SA

Saudi Arabia									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
9	MQ	Pre	P<0.01		0.24	Same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	RQ1) Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	P<0.01				NN		
		Delayed Post	P<0.01				NN		

Table J.6 Normality, Baseline, Tests Required, and Research Questions for PD and GF in UK

United Kingdom									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
10	PD	Pre	P<0.01		0.26	Same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	RQ1) Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	P<0.01				NN		
		Delayed Post	P<0.01				NN		
11	GF	Pre	0.33		0.45	same	N	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	
		Post	0.03				NN		
		Delayed Post	0.03				NN		

Table J.7 Normality, Baseline, Tests Required, and Research Questions for GJT and GJT-Cor in UK

United Kingdom									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
12	GJT	Pre	0.38	0.31		same	N	Raw scores RM-ANOVA+ Planned Contrasts for interaction time*group	RQ1) Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	0.37				N		
		Delayed Post	0.13				N		
13	GJTT-Cor	Pre	0.02		0.02	Not the same	NN	Gain scores 1) Compare pre-post gains across groups. 2) Compare pre- delayed post gains across groups. 3) compare post delayed post gains across groups Using KW+MW	
		Post	P<0.01				NN		
		Delayed Post	0.02				NN		

Table J.8 Normality, Baseline, Tests Required, and Research Questions for GJT-Inc in UK

United Kingdom									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
14	GJT-Inc	Pre	0.04		0.48	same	NN	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and delayed post scores across groups Using KW+MW	RQ1) Do recasts, metalinguistic information, and 'oral task alone' help the development of English modals amongst speakers of Arabic? 1a) What is the effectiveness of these three intervention types relative to each other? 1b) Are any gains maintained after a delay of about 7 weeks? 1c) Are gains observed differentially on different outcome measures? RQ3 What are learners' opinions about the different feedback techniques?
		Post	0.01				NN		
		Delayed Post	0.14				N		

Table J.9 Normality, Baseline, Tests, and Research Questions on PD for SA and UK (MI) Groups

Metalinguistic Groups in UK and SA									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
1	PD	Pre	P<0.01		0.82	Same	NN	Raw scores 1 - split data , do Non-Parametric Friedman + Wilcoxon, 2) compare pre-delayed post gains across groups	RQ2 Are results observed differentially in different contexts: EFL in SA, and ESL in the UK? RQ3a) Do opinions differ according to the context in which the study was done?
		Post	P<0.01				NN		
		D-Post	P<0.01				NN	2 - Compare post-test scores across groups and delayed post scores across groups. Using KW+MW	

Table J.10 Normality, Baseline, Tests, and Research Questions on GF for SA &UK (MI) Groups

Metalinguistic Groups in UK and SA									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
1	GF	Pre	0.26	P<0.01		Not Same	N	Raw score 1) Repeated measure, ANCOVA 2) planned contrasts for interaction time *group comparing post-test	RQ2 Are results observed differentially in different contexts: EFL in SA, and ESL in the UK? RQ3a) Do opinions differ according to the context in which the study was done?
		Post	0.15				N		
		D-Post	0.06				N		

Table J.11 Normality, Baseline, Tests, and Research Questions on PD for SA and UK (R) Groups

Recast Groups in UK and SA									
No	Test		Shapiro	ANOVA	K.W	Base-line	Normality	Test Required on Data	Research Questions
2	PD	Pre	P<0.01		0.41	Same	NN	Raw scores 1) Non-Parametric Friedman + Wilcoxon, 2) compare pre-delayed post gains across groups	RQ2 Are results observed differentially in different contexts: EFL in SA, and ESL in the UK? RQ3a) Do opinions differ according to the context in which the study was done?
		Post	P<0.01				NN		
		Delayed-Post	P<0.01				NN	3) Compare post-test scores across groups and delayed post scores across groups. Using KW+MW	

Table J.12 Normality, Baseline, Tests, and Research Questions on GF for SA &UK (R) Groups

Recast Groups in UK and SA									
No	Test		Shapiro	ANOVA	K W	Base-line	Normality	Test Required on Data	Research Questions
2	GF	Pre	0.01		P<0.01	Not Same	NN	Gain scores 1) compare gains across groups Using KW and MW	RQ2 Are results observed differentially in different contexts: EFL in SA, and ESL in the UK? RQ3a) Do opinions differ according to the context in which the study was done?
		Post	P<0.01				NN		
		Delayed-Post	0.05				NN		

Table J.13 Normality, Baseline, Tests, and Research Questions on PD for SA and UK (TO) Groups

Task Only Groups in UK and SA									
No	Test		Shapiro	ANOVA	K W	Base-line	Normality	Test Required on Data	Research Questions
3	PD	Pre	P<0.01		0.76	Same	NN	Raw scores 1 - split data , do Non-Parametric Friedman + Wilcoxon, 2) compare pre-delayed post gains across groups	RQ2 Are results observed differentially in different contexts: EFL in SA, and ESL in the UK? RQ3a) Do opinions differ according to the context in which the study was done?
		Post	P<0.01				NN		
		Delayed-Post	P<0.01				NN	2 - Compare post-test scores across groups and delayed post scores across groups. Using KW+MW	

Table J.14 Normality, Baseline, Tests, and Research Questions on GF for SA and UK (TO) Groups

Task Only Groups in UK and SA									
No	Test		Shapiro	ANOVA	K W	Base-line	Normality	Test Required on Data	Research Questions
3	GF	Pre	0.02		0.01	Not Same	NN	Gain scores 1) compare gains across groups Using KW and MW	RQ2 Are results observed differentially in different contexts: EFL in SA, and ESL in the UK? RQ3a) Do opinions differ according to the context in which the study was done?
		Post	0.05				NN		
		Delayed-Post	P<0.01				NN		

Appendix K: Tests Administered

Table K.1 Free Oral Picture Description Test Administered in United Kingdom

UK	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test	Results	Test	Group	Results
1	PD (k=7)	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and D post scores across groups Using KW&MW	Meta (n=13)	χ^2 (2)=10.92 P <0.01	Post-pre	Z = -2.76, p <0.01	Post	H(2) =3.18 P = 0.20	Post	MI vs. TO	Z=0.53,p=0.61
					D post-pre	Z=-2.82, p<0.01				R vs. TO	Z=-1.18,p=0.26
					D post-post	Z=-0.94, p=0.38				MI vs. R	Z=-1.70,p=0.09
			Recast (n=13)	χ^2 (2)=12.67 P<0.01	Post-pre	Z=-1.49, p=0.08					
					D post-pre	Z = -2.67, p<0.01	D post	H(2) = 0.27 P = 0.87	D post	MI vs. TO	Z=-0.06,p=0.48
					D post-post	Z=-3.06, p<0.01				R vs. TO	Z=-0.31,p=0.39
			Task only (n=10)	χ^2 (2)=4.20 P = 0.12	Post-pre	Z=1.38, p=0.19				MI vs. R	Z=-0.54,p=0.30
					D post-pre	Z=-1.38, p=0.19					
					D post-post	Z=-0.87, p=0.22					

Meta : Metalinguistic information group

Table K.2 Gap Fill Test Administered in United Kingdom

UK	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test	Results	Test	Group	Results
2	**GF (k=14)**	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and D post scores across groups Using KW&MW	Meta (n=13)	χ^2 (2)=9.18, P =0.01	Post-pre	Z = -1.68, p = 0.09	Post	H(2) =3.22 P =0.20	Post	MI vs. TO	Z=-0.41, p=0.69
					D post-pre	Z=-2.67, p=0.01				R vs. TO	Z=-0.78, p=0.43
					D post-post	Z=-1.34, p=0.18				MI vs. R	Z=-1.97, p=0.05
			Recast (n=13)	χ^2 (2)= 7.68, P=0.02	Post-pre	Z=-2.47, p=0.01					
					D post-pre	Z =-2.49, p= 0.01	D post	H(2)= 0.47 P = 0.79	D post	MI vs. TO	Z=-0.41, p=0.68
					D post-post	Z=-0.63, p=0.53				R vs. TO	Z=-0.60, p=0.55
			Task only (n=10)	χ^2 (2)= 1.88, P =0.39	Post-pre	Z=-1.91, p=0.06				MI vs. R	Z=-0.44, p=0.66
					D post-pre	Z=-1.68, p=0.09					
					D post-post	Z=-0.54, p=0.59					

Meta : Metalinguistic information group

Table K.3 Timed Grammaticality Judgment Test Administered in United Kingdom

UK	Overall results for Grammaticality judgment test (k=18) in UK Tests of Within-Subjects Effects					
3	Source	Type ||| sum of Squares	Df	Mean square	F	Sig
	Tests	2220.51	2	1110.26	9.70	0.00
	tests * Treatment group	554.79	4	138.70	1.21	0.32
	Error (tests)	7560.78	66	114.56		

Table K.3 continued

UK	Overall results for Grammaticality judgment test (k=18) in UK Tests of Within-Subjects Contrasts				
3		tests	Df	F	Sig
	Tests	Pre-test vs. Post-test	1	8.39	0.00
		Pre-test vs. D Post-test	1	16.01	0.00
		Post-test vs. D Post-test	1	2.98	0.09
	Error (tests)		33		

Table K.3 continued

UK	Grammaticality Judgment Test (k=18) in UK Tests of Between-Subjects Effects					
3	Source	Type ||| sum of Squares	Df	Mean square	F	sig
	Treatment group	479.91	2	239.96	1.32	0.28
	Error (tests)	5990.42	33	181.53		

Table K.3 continued

UK	Overall results for Grammaticality judgment test (k=18) in UK Tests of Within-Subjects Contrasts				
3		Tests* treatment groups	Df	F	Sig
	Tests* treatment group	Pre-test vs. Post-test	2	1.27	0.29
		Pre-test vs. D Post-test	2	0.30	0.74
		Post-test vs. D Post-test	2	2.29	0.12
	Error (tests)		33		

Table K.4 Correct Items in Timed Grammaticality Judgment Test Administered in United Kingdom

UK	Test	Test Required on Data	Comparison Across Groups (Gain Scores)			
			Gain	KW	MW	
					Group	Result
4	GJT-Cor (k=9)	Gain scores 1)Compare pre-post gains across groups, 2) compare pre-d post gains across groups, 3) compare post d post gains across groups Using KW&MW	Pre-post gains	H(2) =1.87, P = 0.39	MI vs. TO	Z= -.789,p=0.43
					R vs. TO	Z=-1.26,p=0.23
					MI vs. R	Z=-0.78,p=0.45
			Pre-D post gains	H(2) = 7.66, P = 0.02	MI vs. TO	Z=-2.73,p=0 .01
					R vs. TO	Z=-1.78,p=0.08
					MI vs. R	Z=-1.02,p=0.28
			Post-D post gains	H(2) =5.70, P = 0.06	MI vs. TO	Z= -2.22, p=0.03
					R vs. TO	Z=-0.69,p=0.52
					MI vs. R	Z=-1.76,p=0.04

Table K.5 Incorrect Items in Timed Grammaticality Judgment Test Administered in United Kingdom

UK	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test	Results	Test	Group	Results
5	**GJT-Incor (k=9)**	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and D post scores across groups Using KW&MW	Meta (n=13)	$\chi^2(2)=11.6$, P <0.01	Post-pre	Z =-2.56, p =0.01	Post	H(2)=1.05, P =0.59	Post	MI vs. TO	Z=-0.16, p=0.88
					D post-pre	Z=-2.92, p<0.01				R vs. TO	Z=-0.66, p=0.51
					D post-post	Z=-1.47, p=0.14				MI vs. R	Z=-1.023, p=0.31
			Recast (n=13)	$\chi^2(2)=9.17$, P=0.01	Post-pre	Z=-1.48, p=0.14					
					D post-pre	Z =-2.83, p= 0.01	D post	H(2)=1.62, P =0.45	D post	MI vs. TO	Z=-1.34, p=0.18
					D post-post	Z=-1.72, p=0.09				R vs. TO	Z=-0.81, p=0.42
			Task only (n=10)	χ^2 (2)=0 .79, P =0.67	Post-pre	Z=-1.48, p=0.14				MI vs. R	Z=-0.29, p=0.77
					D post-pre	Z=-0.42, p=0.67					
					D post-post	Z=-0.36, p=0.72					

Meta : Metalinguistic information group

Table K.6 Elicited Imitation Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test		Test	Group	Results
6	EI (k=33)	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW&MW	Meta (n=19)	$\chi^2(2)=14.14$ P <0.01	Post-pre	Z=-3.46,p=00	Post	H(2)= 3.16, P = 0.21	Post	MI vs. TO	Z=-1.74, p=0.08
					D post-pre	Z=-2.37,p=0.01				R vs. TO	Z=-0.43, p=0.67
					D post-post	Z=-0.82,p=0.21				MI vs. R	Z=-1.29, p=0.20
			Recast (n=25)	χ^2 (2)=27.75 P<0.01	Post-pre	Z=-4.16,p=00					
					D post-pre	Z=-3.44,p=00	D post	H(2)= 0.77, P = 0.68	D post	MI vs. TO	Z=-0.21, p=0.42
					D post-post	Z=-0.21,p=0.42				R vs. TO	Z=-0.88, p=0.19
			Task only (n=20)	χ^2 (2)=13.00 P <0.01	Post-pre	Z=-2.76,p<0.01				MI vs. R	Z=-0.53, p=0.30
					D post-pre	Z=-2.77,p<0.01					
					D post-post	Z=-2.09,p=0.02					

Meta : Metalinguistic information group

Table K.7 Correct Items in Oral Elicited Imitation Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test	Results	Test	Group	Results
7	EI-Cor (k=18)	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW&MW	Meta (n=19)	χ^2 (2)=15.13 P <0.01	Post-pre	Z=-3.21,p<0.01	Post	H(2) =3.42, P = 0.18	Post	MI vs. TO	Z=-1.25, p=0.21
					D post-pre	Z=-2.77,p<0.01				R vs. TO	Z=-0.71, p=0.48
					D post-post	Z=-0.27,p=0.41				MI vs. R	Z=-1.77, p=0.08
			Recast (n=25)	χ^2 (2)=9.29 P=0.01	Post-pre	Z=-1.87,p=0.06					
					D post-pre	Z=-2.47,p=0.01	D post	H(2) =0.26, P = 0.88	D post	MI vs. TO	Z=-0.41, p=0.45
					D post-post	Z=-1.59,p=0.06				R vs. TO	Z=-0.15, p=0.45
			Task only (n=20)	χ^2 (2)=10.83 P <0.01	Post-pre	Z=-2.30,p=0.02				MI vs. R	Z=-0.47, p=0.33
					D post-pre	Z=-2.42,p=0.01					
					D post-post	Z=-1.26,p=0.11					

Meta : Metalinguistic information group

Table K.8 Incorrect Items in Oral Elicited Imitation Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test	Results	Test	Group	Results
8	**EI-Incor (k=15)**	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW&MW	Meta (n=19)	χ^2 (2)=9.12 P = 0.01	Post-pre	Z=-3.44, p<0.01	Post	H(2) = 3.69, P = 0.16	Post	MI vs. TO	Z=-1.82, p=0.07
					D post-pre	Z=-1.78, p=0.08				R vs. TO	Z=-1.27, p=0.21
					D post-post	Z=-1.25, p=0.12				MI vs. R	Z=-0.87, p=0.39
			Recast (n=25)	χ^2 (2)=29.46 P=0.01	Post-pre	Z=-4.19, p<0.01					
					D post-pre	Z=-3.86, p<0.01	D post	H(2) = 0.67, P = 0.72	D post	MI vs. TO	Z=-0.08, p=0.47
					D post-post	Z=-1.42, p=0.08				R vs. TO	Z=-0.85, p=0.20
			Task only (n=20)	χ^2 (2)=6.63 P = 0.04	Post-pre	Z=-1.55, p=0.13				MI vs. R	Z=-0.50, p=0.31
					D post-pre	Z=-1.94, p=0.05					
					D post-post	Z=-1.33, p=0.10					

Meta : Metalinguistic information group

Table K.9 Free Oral Picture Description Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split data				Comparison across groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test		Test	Group	Results
9	**PD (k=7)**	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW&MW	Meta (n=19)	χ^2 (2)=4.62 P = 0.10	Post-pre	Z=-2.37,p=0.02	Post	H(2) = 1.30, P = 0.52	Post	MI vs. TO	Z=-0.30, p=0.77
					D post-pre	Z=-1.37,p=0.18				R vs. TO	Z=-0.72, p=0.47
					D post-post	Z=-0.52,p=0.31				MI vs. R	Z=-1.13, p=0.26
			Recast (n=25)	χ^2 (2)=4.85 P=0.09	Post-pre	Z=-2.21,p=0.03					
					D post-pre	Z=-2.47,p=0.01	D post	H(2) = 1.45, P = 0.49	D post	MI vs. TO	Z=-1.01, p=0.32
					D post-post	Z=-0.37,p=0.37				R vs. TO	Z=-1.06, p=0.14
			Task only (n=20)	χ^2 (2)=2.69 P = 0.26	Post-pre	Z=-0.50,p=0.63				MI vs. R	Z=-0.22, p=0.42
					D post-pre	Z=-1.14,p=0.27					
					D post-post	z=-1.98,p=0.02					

Meta : Metalinguistic information group

Table K.10 Gap Fill Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test	Results	Test	Group	Results
10	**GF (k=14)**	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW&MW	Meta (n=19)	χ^2 (2)=16.13 P <0.01	Post-pre	Z= -3.34, p<0.01	Post	H(2) =1.09, P = 0.58	Post	MI vs. TO	Z=-0.47, p=0.64
					D post-pre	Z=-3.21, p<0.01				R vs. TO	Z=-1.02, p=0.31
					D post-post	Z=-0.13, p=0.46				MI vs. R	Z=-0.56, p=0.58
			Recast (n=25)	χ^2 (2)=14.38 P<0.01	Post-pre	Z=-3.62, p<0.01					
					D post-pre	Z= -2.60, p<0.01	D post	H(2) = 0.72, P = 0.70	D post	MI vs. TO	Z=-0.59, p=0.28
					D post-post	Z=-0.60, p=0.29				R vs. TO	Z=-0.40, p=0.35
			Task only (n=20)	χ^2 (2)=12.57 P <0.01	Post-pre	Z=-2.95, p<0.01				MI vs. R	Z=-0.77, p=0.23
					D post-pre	Z=-2.05, p=0.04					
					D post-post	Z=-1.61, p=0.06					

Meta : Metalinguistic information group

Table K.11 Untimed Grammaticality Judgment Test Administered in Saudi Arabia

AS	Untimed Grammaticality Judgment Test in SA Tests of Within-Subjects Effects (k=18)					
11	Source	Type ||| sum of Squares	Df	Mean square	F	Sig
	Tests	4037.37	2	2018.69	19.00	0.00
	tests * Treatment group	1448.50	4	362.12	3.41	0.01
	Error (tests)	12746.75	120	106.22		

Table K.11 Continued.

SA	Untimed Grammaticality Judgment Test in SA Tests of Within-Subjects Contrasts (k=18)				
11		Tests	Df	F	Sig
	Tests	Pre-test vs. post-test	1	38.49	0.00
		Pre-test vs. delayed post-test	1	15.86	0.00
		Post-test vs. delayed post-test	1	4.50	0.04
	Error (tests)		60		

Table K.11 Continued.

SA	Untimed Grammaticality Judgment Test in SA Tests of Within-Subjects Contrasts (k=18)				
11		Tests* treatment groups	Df	F	Sig
	Tests* treatment group	Pre-test vs. Post-test	2	1.39	0.26
		Pre-test vs. D Post-test	2	2.22	0.12
		Post-test vs. D Post-test	2	6.23	0.00
	Error (tests)		60		

Table K.11 Continued.

SA	Untimed Grammaticality Judgment Test in SA Tests of Between-Subjects Effects (k=18)					
11	Source	Type ||| sum of Squares	Df	Mean square	F	sig
	Treatment group	123.52	2	61.76	0.24	0.79
	Error (tests)	15760.30	60	262.67		

Table K.12 Correct Items in Untimed Grammaticality Judgment Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test		Test	Group	Results
12	**GJT-Cor. (k=9)**	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW& MW	Meta (n=19)	χ^2 (2)=2.84 P = 0.24	Post-pre	Z=-0.86, p=0.40	Post	H(2) =0.35, P = 0.84	Post	M I vs. TO	Z=-0.43, p=0.67
					D post-pre	Z=-0.51, p=0.61				R vs. TO	Z=-0.56, p=0.58
					D post-post	Z=-1.88, p=0.04				MI vs. R	Z=-0.13, p=0.89
			Recast (n=25)	χ^2 (2)=11.45 P<0.01	Post-pre	Z=-2.18, p=0.03					
					D post-pre	Z=-2.47, p=0.01	D post	H(2)= 7.14, P = 0.03	D post	MI vs. TO	Z=-2.44, p<0.01
					D post-post	Z=-0.48, p=0.33				R vs. TO	Z=-0.82, p=0.21
			Task only (n=20)	χ^2 (2)=8.22 P = 0.02	Post-pre	Z=-1.59, p=0.12				MI vs. R	Z=-2.11, p=0.02
					D post-pre	Z=-2.63, p<0.01					
					D post-post	Z=-1.69, p=0.05					

Meta : Metalinguistic information group

Table K.13 Incorrect Items in Untimed Grammaticality Judgment Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test		Test	Group	
13	**GJT-Incor (k=9)**	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW+MW	Meta (n=19)	χ^2 (2)=15.60 P <0.01	Post-pre	Z=-3.63, p<0.01	Post	H(2) =5.56, P = 0.06	Post	MI vs. TO	Z=-1.45, p=0.15
					D post-pre	Z=-1.07, p=0.31				R vs. TO	Z=-0.57, p=0.57
					D post-post	Z=-2.64, p<0.01				MI vs. R	Z=-2.42, p=0.02
			Recast (n=25)	χ^2 (2)=6.27 P=0.04	Post-pre	Z=-2.11, p=0.04					
					D post-pre	Z=-1.30, p=0.19	D post	H(2) = 0.75, P = 0.69	D post	MI vs. TO	Z=-0.78, p=0.22
					D post-post	Z=-0.50, p=0.32				R vs. TO	Z=-0.50, p=0.31
			Task only (n=20)	χ^2 (2)=3.03 P = 0.22	Post-pre	Z=-2.26, p=0.02				MI vs. R	Z=-0.56, p=0.29
					D post-pre	Z=-1.13, p=0.27					
					D post-post	Z=-0.49, p=0.33					

Meta : Metalinguistic information group

Table K.14 Metalinguistic Questions Test Administered in Saudi Arabia

SA	Test	Test Required on Data	Split Data				Comparison Across Groups				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test	Results	Test	Group	Results
14	**MQ (k=9)**	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW+MW	Meta (n=19)	χ^2 (2)=7.14 P = 0.03	Post-pre	Z=-2.67, p<0.01	Post	H(2) =0.35, P = 0.84	Post	MI vs. TO	Z=-0.09, p=0.93
					D post-pre	Z=-1.65, p=0.11				R vs. TO	Z=-0.64, p=0.53
					D post-post	Z=-1.03, p=0.17				MI vs. R	Z=-0.32, p=0.76
			Recast (n=25)	χ^2 (2)=11.04 P<0.01	Post-pre	Z=-3.52, p<0.01					
					D post-pre	Z=-1.59, p=0.12	D post	H(2) = 0.63, P = 0.73	D post	MI vs. TO	Z=-0.40, p=0.35
					D post-post	Z=-1.55, p=0.07				R vs. TO	Z=-0.55, p=0.29
			Task only (n=20)	χ^2 (2)=7.17 P = 0.03	Post-pre	Z=-2.69, p<0.01				MI vs. R	Z=-0.69, p=0.25
					D post-pre	Z=-0.84, p=0.44					
					D post-post	Z=-1.54, p=0.07					

Meta : Metalinguistic information group

Table K.15 Free Oral Picture Description Test Administered across Metalinguistic Groups in SA and UK

No	Test	Test Required on Data	Split Data				Comparison Across Countries				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test		Test	Group	Results
15	**PD (k=7)**	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2-compare post-tests cores across groups and d post scores across groups Using KW+MW	Meta.SA (n=19)	$\chi^2(2)=4.62$, p=0.09	Post-pre	Z=-2.37,p=0.02	Post	H(1)= 0.31, P = 0.58	Post	M(SA) vs. M(UK)	Z=-0.56, p=0.58
					D post-pre	Z=-1.37,p=0.17					
					D post-post	Z=-0.52,p=0.60					
			Meta UK (n=13)	$\chi^2(2)=10.92$, p<0.01	Post-pre	Z=-2.76,p=00					
					D post-pre	Z=-2.82,p=00	D post	H(1)= 0.54, P = 0.46	D post	M(SA) vs. M(UK)	Z=-0.73, p=0.47
					D post-post	Z=-0.94,p=0.35					

Meta : Metalinguistic information group

Table K.16 Free Oral Picture Description Test Administered across Recast Groups in SA and UK

No	Test	Test Required on Data	Split Data				Comparison Across Countries				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test		Test	Group	Results
16	**PD (k=7)**	Raw scores 1 - split data Non-Parametric Friedman &Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW+MW	Recast-SA (n=25)	$\chi^2(2)=4.85$, p=0.09	Post-pre	Z=-2.21,p=0.03	Post	H(1)= 0.06, P = 0.81	Post	R(SA) vs. R(UK)	Z=-0.25, p=0.81
					D post-pre	Z=-2.47,p=0.01					
					D post-post	Z=-0.37,p=0.72					
			Recast - UK (n=13)	$\chi^2(2)=12.67$, p<0.01	Post-pre	Z=-1.49,p=0.14					
					D post-pre	Z=-2.67,p=0.01	D post	H(1)= 1.10, P = 0.22	D post	R(SA) vs. R(UK)	Z=-1.22, p=0.23
					D post-post	Z=-3.06,p<0.01					

Table K.17 Free Oral Picture Description Test Administered across Task Only Groups in SA and UK

No	Test	Test Required on Data	Split Data				Comparison Across Countries				
			Group	Friedman	Wilcoxon		KW		MW		
					Test	Results	Test		Test	Group	Results
17	**PD (k=7)**	Raw scores 1 - split data Non-Parametric Friedman & Wilcoxon 2 - compare post-test scores across groups and d post scores across groups Using KW+MW	Task only SA (n=20)	$\chi^2(2)=2.69$, p=0.26	Post-pre	Z=-0.50,p=0.62	Post	H(1)= 0.33, P = 0.57	Post	TO (SA) vs. TO (UK)	Z=-0.57,p=0.57
					D post-pre	Z=-1.14,p=0.26					
					D post-post	Z=-1.98,p=0.05					
			Task only UK (n=10)	$\chi^2(2)=4.20$, p=0.12	Post-pre	Z=-1.38,p=0.17					
					D post-pre	Z=-1.38,p=0.17	D post	H(1)= 0.02, P = 0.89	D post	TO(SA) vs. TO(UK)	Z=-0.13,p=0.91
					D post-post	Z=-0.87,p=0.39					

Table K.18 Gap Fill Test Administered across SA and UK for the Metalinguistic Groups

ANCOVA Within the group	Gap Fill Test (k=11) (Dependent V: Post-test) Tests of Within-Subjects Effects (covariate: pre-test)					
	Source	Type ||| sum of Squares	Df	Mean square	F	Sig
	Tests (Pre-test GF)	11235.915	1	11235.915	29.467	0.00
	Error (tests)	11439.060	30	381.302		

Table K.18 Continued

ANCOVA Within the group	Gap Fill Test (k=11) (Dependent Variable: Delayed post-test) Tests of Within-Subjects Effects (covariate: pre-test)					
	Source	Type ||| sum of Squares	Df	Mean square	F	Sig
	Tests (pre-test GF)	17731.201	1	17731.201	38.211	0.00
	Error (tests)	13921.046	30	464.035		

Table K.18 Continued

ANCOVA Within the group	Gap Fill Test (k=11) (Dependent Variable: Delayed post-test) Tests of Within-Subjects Effects (covariate: post-test)					
	Source	Type ||| sum of Squares	Df	Mean square	F	Sig
	Tests (pre-test GF)	23351.433	1	23351.433	84.394	0.00
	Error (tests)	8300.814	30	276.694		

Table K.18 Continued

ANCOVA across Country						
Gap Fill Test (k=11) **Tests of Within-Subjects Effects (covariate: pre / post-test)**						
Context	**Time**	**Type ||| sum of Squares**	**df**	**Mean square**	**F**	**sig**
SA	Pre-post	6810.939	1	6810.939	21.852	0.00
	Error(tests)	5298.674	17			
	Pre-delayed	9512.053	1	9512.053	21.963	0.00
	Error(tests)	7362.676	17			
	Post-delayed	12697.261	1	12697.261	51.671	0.00
	Error(tests)	4177.467	17			
UK	Pre-post	380.955	1	380.955	0.972	0.35
	Error(tests)	4310.717	11			
	Pre-delayed	1074.380	1	1074.380	2.531	0.14
	Error(tests)	4669.421	11			
	Post-delayed	2358.486	1	2358.486	7.663	0.02
	Error(tests)	3385.315	11			

Table K.19 Gap Fill Test Administered across SA and UK for the Recast Groups

No	Test	Test Required on Data	Comparison Across Country			
			Gain Scores	KW	MW	
					Recast Group	Result
19	**GF (k=11)**	Gain scores 1)compare pre-post gains across groups, 2)compare pre-d post gains across groups, 3) compare post-d post gains across groups using KW+MW	Pre-post gains	H(1) =1.55, P = 0.21	SA vs. UK	Z=-1.24, p=0.21
			Pre-d post gains	H(1) = 0.62, P = 0.43	SA vs. UK	Z= -0.79, p=0.43
			Post-d post gains	H(1) =0.01, P = 0.94	SA vs. UK	Z=-0.08, p=0.94

Table K.20 Gap Fill Test Administered across SA and UK for the Task Only Groups

No	Test	Test Required on Data	Comparison Across Country			
			Gain Scores	KW	MW	
					Task only Group	Result
20	**GF (k=11)**	Gain scores 1)Compare pre-post gains across groups, 2) compare pre-d post gains across groups, 3) compare post-d post gains across groups using KW and MW	Pre-post gains	H(1)=1.58, p = 0.208	SA vs. UK	Z= -1.26, p=0.21
			Pre-d post gains	H(1) = 0.33 p = 0.56	SA vs. UK	Z= -0.58, p=0.56
			Post-d post gains	H(1) =2.29, p= 0.13	SA vs. UK	Z= -1.51, p=0.13

Table K.21 Free Oral Picture Description Test Administered in UK and SA for All Groups

No	Test	Test Required on Data	Split Data				Comparison Across Groups in UK and SA				
			Group	Friedman	Wilcoxon		KW		MW		
					Period	Result	period	Result	Test	Group	P
21	**PD**	Raw scores 1 - split data Non-Parametric Friedman + Wilcoxon 2 - compare post-test scores across groups and D post scores across groups using KW+MW	Meta(32)	$\chi^2(2)= 14.13$, P <0.01	Post-pre	(Z= -3.63, p =0.00)	Post	H (2) = 3.75, p = 0.15	Post	MI vs. TO	*z* = -0.57, *p*= 0. 57.
					D post-pre	(Z= -2.77, p =0.01)				R vs. TO	*z* = -1.316, *p*= 0.19.
					D post-post	(Z= -0.88, p= 0.38)				MI vs. R	*z* = --1.84, *p*= 0. 07.
			Recast(38)	$\chi^2(2)= 13.92$, P <0.01	Post-pre	(Z = -2.63, p=0. 01)					
					D post-pre	(Z=-3.36, p<0.01)	D post	H (2) = 0.84, p = 0. 66	D post	MI vs. TO	*z* = -0. 84, *p*= 0. 40
					D post-post	(Z= -1.90, p = 0.06)				R vs. TO	*z* = -0.75, *p*= 0.46.
			Task only(30)	$\chi^2(2)= 3.45$, P = 0.18	Post-pre	(Z= -0.20, p = 0. .85)				MI vs. R	*z* = -0.14, *p*= 0. 89.
					D post-pre	(Z= -1.64, p = 0. 10)					
					D post-post	(Z= -2.16, p = 0. .03)					

Meta : Metalinguistic information group

Appendix L: Studies on the Effects of Corrective Feedback in L2 Learning

Study	Interactional Focus	Linguistic Focus	Participant	Tests and Findings
Adams (2007)	Incidental feedback on lexical items	Questions, past tense, locative prepositions	25 ESL adults	Tailor-made post-test. Evidence of learning occurred most on past tense items but less frequent on the other three item types. The feedback episodes were more helpful for some students than for others.
Ayoun (2001)	Recasts, models (written/ computerized)	Verb tense	145 French FL adults	Grammaticality judgment task/ correction task and free composition (pre-test and post-test) written recasts were more effective than modelling and traditional grammar instruction
Carpenter, Jeon, MacGregor, & Mackey (2006)	Recasts, repetitions, learners' perceptions	Morphosyntax, lexis, phonology	34 ESL adults	Videotape stimulus throughout a communicative task based activity in dyads with one of 2 NSs.
Carroll (2001)	Explicit and implicit negative feedback	Elicited verb-noun conversations in a sentence format.	100 adult low-intermediate ESL learners	Elicited verb-noun conversions in a sentence format. All types of feedback helped learners to learn the targeted items. Explicit metalinguistic information and indirect prompting enabled learners to form a generalization. Recasts did not facilitate generalization.
Carroll & Swain (1993)	Explicit and implicit negative feedback	Dative Verbs	100 Spanish ESL adults	Recall production following each feedback session. All of the treatment groups performed better than the control group on

				both recalls tasks. Direct metalinguistic group outperformed the other groups.
DeKeyser (1993)	The effect of error correction on L2 grammar knowledge and oral proficiency	Variety of features, predominantly morphosyntactic	25 high school seniors of L2 French	Three oral communications (interview, picture description, storytelling), fill in the blank (pre/post-tests). No statistically significant differences evident between groups A and B (explicit and implicit feedback) Learners with high previous achievement, high language aptitude, high extrinsic motivation and low anxiety benefited the most from error correction.
Doughty & Varela (1998)	Corrective recasts	Past tense	34 ESL learners	Written and oral science reports (pre-test, post-test and delayed post-test). Positive effect on post tests and delayed post tests
E la Fuente (2002)	Negotiation of meaning output	nouns	32 adult learners	Self-report scale on receptive and productive target vocabulary knowledge (pre-test, post-test and 2deakyed post- tests). positive effect on comprehension, receptive and productive vocabulary knowledge in immediate and delayed post tests
Egi (2007)	Recasts and learners' interpretations	Morphosyntax, lexicon	49 Japanese FL adults	Tailor made immediate and delayed post. Learners' performance was significantly different depending on how they interpreted recasts.
EGI (2010)	learners' responses to feedback, specifically repair and modified output after receiving recasts		24 Learners of Japanese	Each learner watched video clips of the recast episodes and commented on them. The learners' stimulated recall reports were analyzed in relation to their responses to the recasts: uptake, repair, and modified output. Their reports indicated that they perceived the recasts as corrective feedback significantly

				more frequently compared to cases where they did not produce uptake. Modified output was also significantly related both to learners' recognition of corrective recasts and to their noticing of the gap.
Ellis (2007)	Recasts and metalinguistic feedback	Past Tense Comparative-er	34 ESL, adults	Untimed GJT, metalinguistic, oral imitation (pre-test, post-test and delayed tests).The number of feedback moves directed at past tense for the recast and metalinguistic groups, exceeded that directed at the comparative but the recast group received more feedback than the metalinguistic group for both structures.
Ellis and He (1999)	Modified output	nouns	50 ESL adult learners	Word recognition, picture matching, oral production tests (Pre-test, post-test and 4 delayed post-tests). immediate and delayed positive effect on comprehension, recognition and production of words
Ellis and Heimbach	Negotiation of meaning	nouns	10 ESL child learners	Receptive and productive vocabulary tests (pre-test and post-test). Positive effect of negotiation on comprehension
Ellis, Loewen, & Erlam (2006)	Recasts, metalinguistic feedback	Past tense	34 ESL adults	Untimed grammaticality J, metalinguistic and oral imitation (pre-test, post-test and delayed test). Greater effect for metalinguistic feedback over recasts.
Elis, Tanaka and Yamazaki (1994)	Negotiation of meaning	nouns	Study1:79EFL study2:127EFL learners	Translation task (pre-test, post-test and delayed post-test). Both studies found positive effects on immediate and delayed post-test
Erlam & Loewen	the effectiveness of	noun–adjective	50 students of	Testing instruments were administered on three occasions. They comprised a spontaneous production test, an elicited imitation

(2010)	implicit and explicit corrective feedback	agreement	French	test, and an untimed written grammaticality judgment test. Results showed no significant differences for type of feedback but significant effects for oral interaction.
Han (2002b)	Recast and non-recast	Present and Past tense-ed	8 adult ESL	Oral and written narratives (pre-test, post-test, and delayed post-test). Recasts produced positive impact on the tense consistency in L2 output.
Havranek & Cesnik (2001)	Recasts, elicitations, explicit corrections	Grammar, lexicon, and English phonological	207 university students specializing in English	Class-specific tests (written, spoken completion task; translation; correction and reading aloud) direct at correct items (data on 1700 corrective feedback episodes from normal English lessons) effectiveness of corrective feedback was in order: 1) elicited self-correction, 2) explicit rejection + recast, 3) recast alone.
Ishida (2004)	Recasts	Aspectual form(-te-i(ru))	4Japanese FL adults	Obligatory occasional analysis of target items in conversations (pre-test and post-test). Recasts produced a positive effect, but the effectiveness varied according to different target structure items.
Iwashita (2003)	Recasts, negotiation, models	Verb morpheme (-te form), construction locative	55Japanese FL adults	Oral production (pre-test, post-test, and delayed post-test). Positive effect on post and delayed post-test
Kim & Mathes (2001)	Explicit vs. implicit corrective feedback	Dative verbs	20 ESL Korean adults	Controlled production tasks (as in the treatment) without feedback. Differences between performance on first and second production tasks were not significant. Differences between

				groups for gains in production were not significant. Learners expressed preference for explicit feedback.
Leeman (2003)	Recasts, negative evidence, enhanced and unenhanced positive evidence	Gender and number agreement	74 Spanish FL adults	Picture description (Post-test -/delayed post-test). Only recast and the enhanced salience groups outperformed the control group on any post-test. No difference between recast and enhance salience.
Linnell (1995)	Interaction/ no focus	Past tense	19 adult ESL learners	Written (grammaticality judgment, sentence combination, free writing, cloze test) oral (sentence imitation and oral interview) (pre-test, post-test and delayed post-test). Positive effect on post and delayed post test
Loewen (2005)	Focus on form episodes	Various linguistic features	12 classes of young adults	Recording of communicative activities. Coding of 491 focus – on-form episodes. Tailor-made tests. Focus –on form episodes results in score gains in the tailor-made post-tests 60%of the time.
Loewen & Nabei (2007)	Metalinguistic, recasts, clarification	Questions	66 EFL, adults	Timed and untimed GJT and oral production. There was no significant increase in scores from pre-test to post test. Among feedback groups, there were no significant differences among the groups on their post test scores.
Long, Inagaki, & Ortega (1998)	Recasts	Locative construction,	74 Japanese FL, Spanish FL, adults	Oral production (pre, post and delayed post-test). positive effect of recasts on post and delayed post test

Long, Inagaki, & Ortega (1998)	Recasts	object topicalization, adverb placement	24 adult learners	Oral production (pre and post-test. No positive effect of recasts on the learning of locative instruction
Loschky (1994)	Interaction / no focus	Nouns	41 adult learners	Vocabulary recognition test (pre-test and post-test). positive effect on comprehension but no positive effect on acquisition
Lyster (2004)	Form –focused instruction(FFI)+ recasts, FFI+ Prompts, FFI only	Grammatical Gender	148 French Immersion, children	Binary choice test, text completion (oral production), object identification, picture description (2 post-tests). recasts were less effective than prompts in leading to improvements, especially on the written production tasks
Lyster & Ranta (1997)	Explicit correction, recast, clarification request, metalinguistic, elicitation, repetition	Grammar, lexis, phonology	French immersion, children	Teachers favour use of recasts but it doesn't lead to repair. Elicitation, metalinguistic feedback, clarification request and repetition lead to greater repair.
Mackey (1999)	Modified input, and feedback from negotiated meaning	Word order in WH questions	34 adult ESL learners	Oral 'spot the difference' task (pre-test, post test1, post-test2 and post-test3). Interactionally modified input produced a positive effect on the development of question formation but this is only evident for learners who are developmentally ready.
Mackey (2006)	Interactional feedback, noticing	Questions, Plurals, past tense	28 ESL adults	Learning journals, stimulated recall and questionnaire (pre/post-tests). positive relationship between reports about noticing and L2 development for questions
Mackey, Gass,	Recasts, negotiation,	Morphosyntax,	10 ESL and	Videotaped a communicative task with a native (English) or

& McDonough (2000)	learners' perception	phonology, lexis, semantics	7 Italian FL learners	near native (Italian) interviewer. Recasts mainly for morphosyntactic errors, negotiation for phonological errors. Non-native speakers noticed phonology most. Morphosyntactic feedback often not perceived. Most accurate perception of lexical and then phonological feedback, 33% of morphosyntactic feedback correctly perceived.
Mackey & Oliver (2002)	Interactional feedback(recasts, negotiation)	Questions	22 ESL, children	Oral production (pre-test, post-test, and 2 delayed post-test). Positive effect on post and delayed post-tests.
Mackey & Philp (1998)	Recasts and interaction	Questions	35 adult ESL learners	Oral 'spot the difference' tasks (pre-test, 2 post tests, and 1 delayed post-test). Positive effect of recasts for more proficient learners.
Mackey & Silver(2005)	Interactional Feedback(recasts, negotiation)	Questions	26 ESL children	Oral production (pre-test and 2 post-tests). Positive effect of interactional feedback on development of question forms.
McDonough (2005)	Interactional feedback (recasts, clarification requests	Past tense (activity verbs)	60 EFL adults	Oral production (pre/post tests and 2 delayed post-tests). clarification requests play an indirect role in ESL question development by facilitating production of modified output
McDonough & Mackey (2006)	Recasts, responses (repetitions, primed production)	Questions	46 EFL adults	Oral pre-test and 2 post-tests similar to the treatment activities.
	Interaction	Articles	91 EFL, adults	Grammaticality J, oral production, written production (2 post-

Muranoi (2000)	Enhancement (IE) and recasts, requests for repetition and IE+ meaning-focused debriefing			tests). Positive effect of interaction enhancement both on post and delayed post-test. Greater effect of interaction enhancement+ formal debriefing than interaction +meaning-focused debriefing
Nabei & Swain (2002)	Corrective feedback episodes with a tailor – made post-test with a stimulated recall interview and delayed post test	unspecified	1 EFL learners of upper-intermediate	Grammaticality judgment test (tailor made post-test, with a stimulated recall interview and delayed post-test. Positive effect evident on the delayed post-test with little effect on the post-test.
Nagata (1993)	feedback indicating what was missing or not expected and metalinguistic explanations	Japanese passive structures; verbal predicates and particles	32 L2 Japanese, adults	Written test using same format as treatment task. Group B (metalinguistic explanations outperformed group A on particles but not verbal predicates. Learners expressed preference for metalinguistic explanation.
Nassaji (2007)	Elicitation, reformulation and others	No specific linguistic focus	42 ESL adults	No measurement tests. Reformulations occurred more frequently than elicitations. Reformulation and elicitation occurred in different forms with different characteristics.
Philp (2003)	Recasts	Questions	33 adults ESL	Cued immediate recall during oral communicative tasks to

				measure noticing (5NS-NNS dyadic interaction over 2 weeks). High level of noticing of recasts. Learners' processing biases may limit noticing
Rosa & Loewen (2004)	A) explicit feedback to both correct and incorrect+ opportunity for the incorrect, B) implicit feedback indicating whether the answer was right or wrong	Contrary to the fact conditional sentences in the past	100 learners of L2 Spanish, adults	Three multiple–choice recognition tests and three written controlled production tests (immediate and delayed post-tests). Recognition tests indicate a statistically significant difference evident between groups A and B for new but not old items. Production tests indicate a statistically significant difference for old bit not new items. Both groups outperformed the control group.
Révész (2009)	task variable +/− contextual support combined with the focus-on-form technique known as recasting	past progressive form	90 adult learners of English as a foreign language	A pretest-posttest-delayed posttest design was employed to detect any improvement in participants' ability to use the linguistic target, which was the past progressive form. Results: 1) Learners who received recasts but did not view photos outperformed learners who received recasts while viewing photos. 2) The group that viewed photos but did not receive recasts achieved greater L2 gains than the group who neither viewed photos nor received recasts.
Sachs & Suh (2007)	Recasts (computer mediated) and	Sequence of tenses	30 EFL adults	Paper based multiple choice and computer mediated tests (pre/ post-tests)

	enhancement techniques			
Sagarra (2007)	Recasts (computer delivered), modified output	Gender and number agreement	82 Spanish FL adults	Screening, written, oral face to face and working memory tests. Oral recasts had an immediate and delayed positive effect on the development of grammatical accuracy in written tests and face to face interactions and on the production of modified output.
Sanz (2003)	Computer delivered implicit vs., explicit feedback in processing instruction	Position of clitic pronouns between object and verb	28 first year university learners of Spanish	Interpretation tested. Production tests: a) sentence completion and b) written video retelling. Both groups significantly increased ability to interpret and accurately produce the target with no difference between groups on any measure.
Sato (1986)	Interaction/ no focus	Past tense	2 adult ESL learners	Oral production (weekly conversational sessions over 10-month period). No development
Sheen (2006)	Recasts and metalinguistic feedback	Articles	80 ESL adults	Speeded dictation, writing, error correction, (pre/post/delayed tests). Metalinguistic group outperformed both recast and control groups but the recast group did not perform significantly better than the control group.
Sheen (2008)	recasts , language anxiety, modified output and L2 learning	Articles	61 ESL adolescents	Speeded dictation test, writing test, and error correction test (pre-test, post-test, and delayed post-test).
Silver (2000)	Interaction/ no focus	questions	32 adult ESL learners	Oral communication test, written word order test, and multiple choice preference test (pre-test, post-test, and delayed post-test).

				No greater effect of negotiation on the immediate post-test than input or output condition but greater effect of negotiation found on the delayed post-test.
Trofimovich, Ammar & Gatbonton (2007)	Recasts(computer-delivered), noticing	Morphosyntax (possessive determiners), lexis, verbs	32 ESL adults	Online picture description (pre/post/ delayed tests). When the learners made errors and then received recast, they were more likely to detect their lexical than their morphosyntactic errors. Overall, learners benefited from the recasts received.
Yang & Lyster (2010)	Prompt and recast CF in form-focused classrooms	use of regular and irregular English past tense	72 Chinese learners	Pretests, immediate posttests, and delayed posttests administered 2 weeks after the treatment assessed participants' acquisition. The effects of prompts were larger than those of recasts for increasing accuracy in the use of regular past tense forms, whereas prompts and recasts had similar effects on improving accuracy in the use of irregular past tense forms.