School of Education

Faculty of Humanities

The Effectiveness of Pre-modified Input and Interactional Feedback during Task-Based Language Teaching on the Uptake and Automaticity of English Comparative Structures

Muhlisin Rasuki

0000-0002-4216-7998

This thesis is presented for the Degree of

Doctor of Education

of

Curtin University

December 2020

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

The research presented and reported in this thesis was conducted in accordance with the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research (2007) – updated May 2015. The proposed research study received human research ethics approval from the Curtin University Human Research Ethics Committee (EC00262), Approval Number # HRE2017-0101

Signature: Muhlisin Rasuki

Date: December 15, 2020

Abstract

The present study follows the pedagogic frameworks of task-based language teaching and is primarily concerned with examining the effectiveness of pre-modified input and interactional feedback on the uptake and automaticity of English comparative structures (ECSs). In this study, the term 'pre-modified input' is used to refer to *model performances* in task-based L2 instruction that have been prepared to contain various exemplars indicating how the target L2 form (ECSs) may be used in contextualized communication, while the term 'interactional feedback' is used to refer to specific strategies such as *negotiation* that indicate to learners that their production of the target L2 form is non-target-like or less comprehensible. Furthermore, the term 'uptake' is operationally defined as learner incorporation of the target L2 form from the input or feedback into his or her L2 production, while the term 'automaticity' is operationally defined as learners' abilities in producing the target L2 form rapidly without undue pauses, repetitions and repairs.

Previous task-based L2 studies have indicated that the provision of pre-modified input or interactional feedback is effective for promoting L2 development (e.g., Ayoun, 2001; Long, Inagaki & Ortega, 1998; Lyddon, 2011; Mackey, 1999; Shintani, 2011; Shintani & R. Ellis, 2010). Nonetheless, the abstract notion of L2 development is usually operationalized in these studies as *accurate use* of specific L2 forms that learners did not know previously or knew only partially. Given that the abstract notion of L2 development is usually defined in the L2 literature as learners' increased abilities in *using* L2 forms *accurately* and *fluently* in relevant L2 communicative contexts, the current study aimed to seek further empirical evidence as to whether the provision of pre-modified input or interactional

feedback is effective in facilitating improvements in various dimensions of L2 development: *increased use*, *increased accuracy* and *increased fluency* in the use of a specific task-relevant L2 form in real-time L2 production. In this study, increased use and increased accuracy in the use of a specific task-relevant L2 form is conceptualized in terms of 'uptake of ECSs', while increased fluency in the use of such a form is conceptualized in terms of 'automaticity of ECSs'.

This study employed a pre-test, post-test, delayed post-test with comparison group design. Participants were fifty-one adult Indonesian learners of L2 English. These participants were randomly assigned into one of three conditions: pre-modified input (PI), interactional feedback (IF), and comparison. Participants in the PI condition performed input-based tasks where they needed to identify a picture of an item from a set of pictures of very similar items (distracters) based on a series of descriptions given by the researcher. In contrast, participants in the IF condition performed output-based tasks where they needed to describe a picture of an item clearly and unambiguously to the researcher so he could identify it from a set of pictures of very similar items (distracters). These participants were provided with interactional feedback in the form of 'recasts' when they produced the target L2 form inaccurately, and/or 'confirmation checks' when they continually avoided using the form when describing the item. Finally, participants in the comparison condition completed given tasks orally with no provision of input or feedback. No explicit instruction was provided throughout the treatments. In this way, each condition was designed to foster incidental as opposed to intentional learning of the target L2 form (Long, 2015, 2019; Shintani & R. Ellis, 2010).

The materials consisted of twelve sets of oral description tasks adopted and adapted from Lambert (2019). A previous study by Lambert (2019) indicated that the tasks

enabled both native and non-native speakers of English to use comparative structures naturally during communication. Nine sets of these tasks were used in the treatments, while three sets were used in the pre-test, post-test and delayed post-test to observe gains, the stability of gains and the transferability of gains to comparable or parallel tasks.

Gains in the uptake of ECSs were measured by observing two usage features: (1) the proportion of ECSs that participants produced in the pre-test, post-test and delayed post-test, as well as (2) the proportion of error-free uses of the ECSs. On the other hand, gains in automaticity were measured by observing three aspects of utterance fluency in the production of ECSs: (3) the total number of pruned syllables, including pauses, uttered per second, (4) the frequency of filled/unfilled pauses and verbatim repetitions that took place within- and between-structure boundaries per second, and (5) the frequency of within-structure repairs per second.

The results of *within*-group comparisons indicated that participants in the experimental (PI and IF) groups performed significantly better on the first three measures: (1) – (3), and that these gains were stable and transferable to the parallel tasks. In contrast, those in the comparison group did not make significant improvements in any of the measures: (1) – (5).

With regard to *between*-group comparisons, statistical analyses revealed that there were already significant differences between the experimental and comparison groups on the pre-test, especially with respect to measures (1) and (3), with the comparison group performed significantly worse than the experimental groups on these two measures. In contrast, no significant differences were identified between the two experimental groups in the pre-test. For this reason, between-group

comparisons for (1) and (3) were performed between the experimental groups only. The results of the analyses indicated that the experimental groups outperformed the comparison group on measures (2) and (4), and that these gains were stable and transferable to the parallel tasks. No significant differences were noted between the experimental groups on any of the measures of uptake and automaticity of ECSs: (1) – (5).

These results are discussed in relation to those established in the earlier L2 work. Furthermore, learner-internal and -external factors that have potentially mediated the efficacy of PI or IF in the study are also considered in providing a comprehensive account of the results. These factors include: (i) participants' prior knowledge of ECSs, (ii) the functional or cognitive demands of the tasks used in the study, and (iii) the relative frequency of which the target L2 form (ECSs) occurred in the input or feedback.

Acknowledgements

I received tremendous support from many people during the course of completing this dissertation, but first, I would like to acknowledge the funding I received from the LPDP (Indonesia Endowment Fund for Education), Ministry of Finance, Republic of Indonesia. The LPDP provided me with financial support throughout my study at Curtin University, Perth, Australia.

I am also very grateful to my doctoral supervisors: Dr Craig Lambert and Professor Rhonda Oliver. They are wonderful teachers and great mentors. It is my privilege to work with them during the completion of this dissertation.

Special thanks to Dr Craig Lambert, my principal supervisor, who had been with me since I first commenced my academic journey at Curtin University. He introduced me to the topic that I raised in this dissertation, read every draft, pointed out every drawback, and criticized every idea that I proposed in this work. The results of his efforts are apparent in every chapter of this dissertation. Needless to say, any remaining shortcomings are entirely my own.

I would also like to thank some people in my home institution, Universitas Muhammadiyah Jember, Indonesia, in particular Dr Moch. Hatip, who gave me permission to use some facilities and resources in the Faculty of Education during data collection, Ms. Anita F. Laely, who gave me permission to recruit students in the Department of Language and Arts, and all students who participated in this study.

Working on a doctoral degree can be frustrating at times and often needs a long period of staying in solitude. For these reasons, I am indebted to my wife, Pratiwi

Andriani, for her patience and willingness to take care of everything during my absence (physically and/or mentally).

Finally, I would like to dedicate this dissertation to my two young daughters, Anatasya Iftihani and Anandita Dewi, who make my life truly meaningful, and to my parents, who have loved and supported me for all my entire life.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The present study is situated within a pedagogic framework of task-based language teaching (TBLT) and is primarily concerned with examining the effectiveness of pre-modified input and interactional feedback on the uptake and automaticity of English comparative structures. Given the focus of the study, this chapter begins with a definition of 'tasks'. It then provides an overview of the use of tasks in L2 research and pedagogy, as well as the relationship between tasks, meaningful L2 use, focus on form and L2 development. Finally, a chapter-by-chapter overview of this thesis is presented.

1.1 What Is Meant by '*Tasks*'?

Various definitions of 'tasks' have been proposed in the TBLT literature (see e.g., Bygate, Skehan & Swain, 2001; R. Ellis, 2003, Long, 2016; Samuda & Bygate, 2008; Skehan, 1998a; van den Branden, 2006). In this study, following R. Ellis (2009), the term '*task*' is used to refer to L2 instructional activities which (1) encourage learners to focus on the meaningful use of an L2 at the level of comprehension and/or production, (2) involve some kind of a communicative gap which creates the necessity for learners to comprehend and/or produce meaning, (3) require learners to use their own linguistic and non-linguistic resources in the course of comprehending and/or producing meaning, and (4) have a clearly defined communicative outcome (p. 223).

In line with the definition above, tasks can be categorized into two distinct types based on the direction of information exchange between learners as they work on them: ‘one-way’ and ‘two-way’ (Long, 1989, pp. 41-42). In one-way tasks, the role of learners is usually split between being either a ‘speaker’ or a ‘listener’. The speaker is the one who holds the critical information that is to be communicated to the listener, while the listener is the one who needs to comprehend the information being given in order to arrive at the task outcome (e.g., Shintani, 2011; Shintani & R. Ellis, 2010). In contrast, in two-way tasks there is no specific role assigned to learners. That is, all learners are encouraged to engage actively in the process of exchanging information to complete a given task successfully (e.g., Ammar & Spada, 2006; Mackey & Oliver, 2002).

One-way tasks may further be categorized into two distinct types: input-based and output-based. Input-based tasks comprise activities where learners are assigned the role of ‘listener’ and thus are required to listen to and comprehend what a speaker says in order that they are able to complete given tasks successfully (e.g., Erlam & R. Ellis, 2018; Shintani, 2016). This type of tasks does not require learners to produce output (sentences or utterances) although they can do so if they wish (R. Ellis, 2001, 2017). By contrast, output-based tasks comprise activities that require learners to produce output with or without opportunities to receive feedback on form from their interlocutor (e.g., Lambert, Kormos & Minn, 2017; Mackey, 1999; Oliver, 1995; Tavakoli, Campbell & McCormack, 2016).

Tasks may also be distinguished in terms of whether they are likely to elicit the use of certain linguistic forms from learners. In a study by Shintani and R. Ellis (2010), for instance, they found that certain tasks enabled the teacher (or researcher) to elicit the use of a plural marker *–s* in English from young Japanese learners of L2

English in a relatively natural communication context. R. Ellis (2003) refers to these tasks as ‘focused tasks’ (cf., Loschky & Bley-Vroman, 1993). Although it is generally agreed that focused tasks play a pivotal role in instructed L2 development research (e.g., Long, Inagaki & Ortega, 1998; Robinson, Cadierno & Shirai, 2009; cf., Skehan, 2014), debate continues about the utility of such tasks in L2 instruction (R. Ellis, 2009, 2017; Long, 2015, 2016; Skehan, 2003, 2007). Arguments about focused tasks are discussed in more detail in Chapter 2.

Other types of tasks that have received a wider acceptance among TBLT advocates are what R. Ellis (2003) labels ‘unfocused tasks’. Unlike focused tasks, unfocused tasks allow a much wider range of linguistic forms to be used by learners as they perform the tasks. Long (2015, 2016) maintains that unfocused tasks should form the basis of TBLT, since they allow learners to practice using their L2 resources in ways that are in line with their own development and internal syllabus.

Finally, Long (1989) distinguishes between ‘closed’ and ‘open’ tasks. To complete closed tasks, learners need to reach a single correct solution that has been determined beforehand. Conversely, open tasks refer to tasks in which there is no pre-determined correct solution and require learners to reach consensus or an original solution to the tasks (see e.g., Lambert & Engler, 2007). The current study employed one-way, closed tasks. The operationalization of such tasks is discussed in detail in Chapter 6.

1.2 Task-Based L2 Research and Pedagogy

Over the last four decades, the potential use of tasks as a unit of analysis for the selection and organization of L2 instruction (Long, 1985, 2015; Long & Crookes, 1992) has generated many empirical studies. Findings from this body of research

have demonstrated that tasks are effective in fostering L2 performance and development for different groups of L2 learners, including young L2 learners at low proficiency levels (e.g., Azkari & Mayo, 2016; Erlam & R. Ellis, 2018; Oliver, 1998; Shintani & R. Ellis, 2010) as well as adult L2 learners at different proficiency levels (e.g., Ahmadian & Tavakoli, 2011; Kuiken & Vedder, 2002; Lambert, Kormos & Minn, 2017; Long, Inagaki & Ortega, 1998; Tavakoli, Campbell & McCormack, 2016). This body of research has also shown the utility of tasks in gaining control over L2 systems that learners have partially mastered (e.g., Lambert, Kormos & Minn, 2017) or in developing new L2 knowledge (e.g., Long, Inagaki & Ortega, 1998).

From a pedagogic perspective, the utility of tasks in fostering L2 performance and development largely depends on how they are designed and implemented in the classroom. For instance, output-based tasks that allow learners to 'negotiate for meaning' (Long, 1996, 2007, 2015) with a more proficient interlocutor can provide rich opportunities for them to receive personalized, well-timed feedback regarding the accuracy and/or comprehensibility of their L2 production (Iwashita, 2003; Mackey, Oliver & Leeman, 2003; Oliver, 1995; van den Branden, 1997). In contrast, input-based tasks that afford exposure to *pre-modified input* (i.e., models of language use that native or proficient non-native speakers produce when performing relevant communication tasks—Long, 2015, 2019) can facilitate incidental learning of specific forms that occur in the input (ibid.) and, thus, support learner use of the L2 and with greater frequency, accuracy and/or fluency in relevant communication contexts (N. Ellis, 2008; N. Ellis & Larsen-Freeman, 2006; Lightbown, 2008). To this end, TBLT is predicated upon the idea that L2 development is *an incidental process* and is largely driven by learners'

communicative needs as they work to achieve certain communicative outcomes (Lambert, 2019, p. 3).

However, in many L2 instructional contexts tasks are more often used as a device for practicing previously taught L2 forms rather than for getting learners engaged in meaningful L2 communication (Butler, 2011). In such instructional contexts, the teaching activities may involve (1) presenting learners with a specific L2 form explicitly, (2) giving them a series of sentence-level exercises aiming at helping them learn how to use the form accurately in sentences, and (3) giving them relevant communicative tasks aiming at helping them develop abilities in using the form accurately and fluently in communication. In this way, tasks are used to engage learners in *intentional* rather than incidental learning of L2 forms (see Chapter 4 for details). Advocates of TBLT consider that such tasks are no different from 'situated L2 learning exercises' (Bygate, 2016; R. Ellis, 2009; Lambert, 2018; Long, 2015; Robinson, 2011a; Skehan, 1998b). This is because their primary use is limited to providing learners with contexts to practice and automatize the processing of target L2 forms in rapid performance (DeKeyser, 2015, 2017).

This is not to say, however, that the teaching of L2 forms is undervalued in TBLT. Rather, L2 forms are taught in line with learners' communicative needs (Long, 2015, p. 249). As stated above, the teaching of L2 forms may be done through providing learners with interactional feedback when they fail to produce sentences or utterances in the L2 accurately (Long, 1996, 2007, 2015), or, alternatively, by exposing them to pre-modified input containing models of L2 task performances (Long, 2015, 2019; Shintani, 2016; see Chapter 4 for details). In fact, advocates of TBLT agree that only having learners engaged in the meaningful L2 use that tasks promote is insufficient to drive forward L2 development (Bygate, 2016; R. Ellis, et

al., 2020; Lambert, 2019; Long, 2015; Robinson, 2001; Skehan, 2003). L2 teachers, therefore, need to pay sufficient attention to L2 forms when designing and implementing tasks in their classroom so that task work can effectively facilitate L2 development.

1.3 Tasks, Meaningful L2 Use, Focus on Form and L2 Development

In line with the discussion above, task-based L2 research is concerned with identifying task-relevant factors that may optimize L2 development. Long has consistently argued that integrating focus on form (FonF) into task-based L2 instruction is indispensable to enable task work that effectively promotes L2 development (Long, 1991, 1998, 2007, 2015; Long & Robinson, 1998). Long defines FonF as follows:

> Focus on form refers to how attentional resources are allocated, and involves briefly drawing students' attention to linguistic elements (words, collocations, grammatical structures, pragmatic patterns, and so on), *in context*, as they arise incidentally in lessons whose overriding focus is on meaning, or communication (Long, 1998, p. 40, emphasis in original).

FonF was proposed by Long as a reaction to the ineffectiveness of two major instructional approaches dominating L2 teaching practices in the preceding years. Long labels these two approaches 'Focus on Forms' (FonFs) and 'Focus on Meaning' (FonM).

FonFs refers to an approach where L2 forms are taught explicitly and deliberately to learners in a decontextualized, discreet fashion. This approach normally assumes that L2 learning is the product of accumulating knowledge of L2 forms bit-by-bit, and that what is taught by the teacher is what will be learned by the learner. This approach underpins L2 teaching methodologies such as the Grammar Translation

Method, the Audiolingual Method, Total Physical Response, Presentation-Practice-Produce, etc. On the other hand, FonM is an approach that assumes L2 forms can be learned incidentally through meaning-focused communication alone. This approach neither involves pre-planned nor overt teaching of L2 forms. Instead, classroom L2 instruction is conducted solely through communicative activities. Theoretically, this approach assumes that L2 learning is triggered and constrained by the kinds of input and/or output (i.e., language samples) that learners receive and/or produce when communicating meaning (Krashen, 1982, 1998; Prabhu, 1987; Swain, 1985, 1993). This approach underlies L2 instructional programs such as the Natural Approach, Immersion, and an earlier version of TBLT governed by a procedural syllabus.

Long reiterates that both FonFs and FonM suffer from major limitations (Long, 1985, 1998, 2009, 2015; Long & Robinson, 1998). In the case of FonFs, it ignores empirical research findings which show that L2 learning "is rarely, if ever, a one-time, categorical event, and that learners pass through developmental stages" (Long, 1998, pp. 37-38). With respect to FonM, years of research shows that "A pure focus on meaning is inefficient", and that "Studies show rate advantages for learners who receive instruction with attention to code features" (Long, 1998, p. 40).

Today, virtually all advocates of TBLT agree that FonF constitutes a necessary component in TBLT to promote L2 development. Nevertheless, they differ in their perspectives as to how FonF may best be realized in TBLT (Bygate, 2016; Doughty, 2001; Doughty & Williams, 1998; R. Ellis, 2016; Robinson, 2001; Skehan, 2003). This disagreement is largely driven by distinct theoretical positions that they hold. Four major theoretical accounts that motivate the impetus of

different versions of TBLT (involving different types of FonF) will be discussed in detail in Chapter 2.

To sum up, interest in TBLT among L2 researchers is largely driven by theoretical and pedagogical issues associated with L2 teaching and learning processes. The present study was also motivated by these issues. In particular, the study aimed to investigate how different types of tasks (input-based and output-based) involving a specific type of FonF (pre-modified input or interactional feedback) facilitate L2 development for adult L2 learners. Chapter 4 discusses these issues in greater detail.

1.4 Overview of the Thesis

This thesis is divided into nine chapters. The present chapter has contextualized the study within the field of TBLT, provided a definition of tasks, discussed various types of tasks that have been widely used in research on task-based L2 use and L2 development. This chapter has also outlined the relationship between the use of tasks in L2 instruction, meaningful L2 use, FonF and L2 development.

Chapter 2 then reviews theoretical perspectives underpinning TBLT and describes how the perspectives have led to the development of multiple versions of TBLT (e.g., R. Ellis, 2003, 2018b; Long, 1985, 2015; Robinson, 2001, 2015; Skehan, 1998a, 2018; Swain, 2000; Swain & Lapkin, 2001). Following this, four major theoretical accounts that have given rise to multiple versions of TBLT, involving different types of FonF, are critically reviewed. These include the interactionist (e.g., Long, 2015), cognitivist (e.g., Robinson, 2001; Skehan, 1998a), socioculturalist (e.g., Lantolf, 2000; Swain, 2000), and modular perspectives (e.g.,

R. Ellis, 2018b). The relationship between these theoretical accounts and the current study are also discussed.

Chapter 3 discusses the relationship between task-based L2 use and L2 development. To begin, the construct of L2 development is first discussed. Then, an operational definition of the term L2 development adopted in the current study is presented. Following usage-based theories of L2 development (e.g., Cadierno & Eskildsen, 2015; Robinson & N. Ellis, 2008), the current study operationally defines L2 development as gradual improvements in the use of particular L2 forms by learners over time, especially as they produce the forms under real-operating conditions. A range of factors that usage-based theories posit as important in L2 development are also discussed. These include exposure to meaningful L2 use, type/token frequency, chunking, cue reliability, and perceptual saliency. The chapter concludes by specifying how these factors may be incorporated in the design of pedagogic tasks to afford optimal conditions for L2 development.

Chapter 4 and 5 then focus on the variables of the study. Chapter 4 focuses on the independent variables: pre-modified input and interactional feedback, while Chapter 5 focuses on the dependent variables related to the two constructs investigated: uptake and automaticity.

To begin, Chapter 4 first presents an operational definition of the term pre-modified input and interactional feedback as employed in the current study. In general, the term ‘pre-modified input’ is used to refer to *models of performances* in task-based L2 instruction that have been prepared to contain various exemplars indicating how a target L2 form (e.g., English comparative structures) may be used in contextualized communication. In contrast, the term ‘interactional feedback’ is

used to refer to specific strategies such as *negotiation* that indicate to learners that their production of the target L2 form is non-target-like or difficult to comprehend. Following this, an in-depth discussion of the debate concerning the relative utility of pre-modified input or interactional feedback in L2 development is reviewed, and key studies that have compared the effectiveness of these two instructional options are critiqued. The chapter concludes by outlining limitations of the studies and pointing out how these limitations have been addressed in the current study.

Chapter 5 then discusses the dependent variables of this study, or how the impact of the two instructional interventions discussed in Chapter 4 may be measured. First, the notions of uptake and automaticity and how they provide evidence of L2 development are discussed. Then, ways in which these two constructs have been measured in previous L2 development studies are examined. Problems related to measurement in those studies are identified, pointing to the need to assess learners' abilities in using task-relevant L2 forms at a certain period of time and to consider gradual improvements they achieve over time, especially as they produce the forms in real time under naturally-occurring L2 communication (Pica, Kang & Sauro, 2006; Rebuschat, 2015; Schmid, Verspoor & MacWhinney, 2011).

Chapter 6 and 7 then present the methodology and results of the main study. Chapter 6 begins by outlining the objectives of the study, followed by research questions and specific hypotheses formulated based on the literature reviewed in the previous chapters. This chapter also provides information about the participants, materials, target L2 form, tests, procedures and statistical analysis used in this study with an aim to future replication.

Chapter 7 presents the results of the study. Prior to reporting the results, information regarding the outcomes of preliminary data screening is provided. The results of a series of statistical analyses are then presented. Given that the data obtained were not normally distributed, a series of non-parametric statistical tests were employed to answer each research question proposed in the previous chapter. Finally, this chapter concludes with a summary of overall findings based on the results of the statistical analyses.

Chapter 8 then discusses the overall results of data analyses in relation to those established in the earlier work. It also discusses the implications of the study for the field of task-based L2 research. Following this, the applicability of the procedures applied in the study to classroom L2 instruction are discussed so that L2 teachers may find the outcomes of this study useful and/or relevant to their own classroom instructional practices.

Finally, Chapter 9 concludes by summarizing the overall findings of the current study, as well as outlining how the findings may contribute to knowledge of task-based L2 learning and teaching.

Chapter 2

Tasks in L2 Teaching and Research

This chapter offers a review of how 'tasks' are conceptualized. To begin, different pedagogic perspectives that have led to the development of multiple versions of TBLT are discussed. Next, four theoretical perspectives that underpin the utility of tasks in L2 pedagogy are reviewed, and their relationship to the current study is discussed. The chapter concludes by specifying the key issues addressed in the current study: the relative impact of pre-modified input and interactional feedback provided in the context of TBLT on L2 use and L2 development.

2.1 Different Pedagogic Perspectives Underpinning TBLT

As stated briefly in Chapter 1, the definition of 'tasks' vary considerably. Such differences stem from distinct pedagogic perspectives that, in turn, reflect how tasks may be designed and implemented in the classroom. The differences also reflect variation in perspectives as to how task-based L2 assessment can be conducted. These are described in detail in the next three subsections.

2.1.1 Situational vs. Interactional Authenticity in TBLT

One of the unresolved questions of TBLT concerns whether tasks should *always* be designed in relation to learners' real-world activities. Long (1985, 2015, 2016) argues in favor of such a position. He claims that a relationship between tasks and learners' real-world activities is central to TBLT. Indeed, he defines tasks as activities that learners do outside of the classroom (ibid.). These activities, as Long

puts it, may or may not involve the use of an L2. However, the goal of TBLT should be to assist the learners to develop their L2 communicative abilities that are directly applicable for use outside of the classroom. In this way, tasks should have 'situational authenticity' (R. Ellis, 2017, p. 508), that is, they should encapsulate what learners do in the real world. Furthermore, for Long (2015, 2016; Long & Crookes, 1992), tasks constitute the primary component in designing units of L2 instruction (see also Robinson, 2015, for a similar argument).

In order for tasks to be best suited to groups of target learners, Long (2015) indicates that they need to be selected based on "a carefully conducted task-based learner needs analysis (NA)" (p. 88). A task-based needs analysis is used to generate a series of 'target tasks' (i.e., genuine exemplars of what learners do, or are likely to do, in their daily life). Once the target tasks have been identified, the teacher or syllabus designer can then categorize them into a series of 'target-task types' (i.e., tasks that have common features). Once the target-task types have been established, the teacher or syllabus designer can then modify, grade and sequence them into a series of 'pedagogic tasks' (i.e., tasks that resemble target tasks in some way but are specifically designed for use in the classroom). Long claims that such an approach to task design helps learners see the relevance between the language that they learn in the classroom and what they will do with the language in the real world, leading to an increase in their motivation (Long, 2015, 2016; Robinson, 2011b, 2015).

Unlike Long, however, R. Ellis (2003, 2009, 2017) and Skehan (2003, 2007) contend that tasks *do not always* need to be established to reflect the tasks that learners will face in real-life. Rather, they argue that the emphasis should be on meaningful L2 use that takes place during L2 instruction. In this way, R. Ellis and

Skehan consider that promoting ‘interactional authenticity’ is more important than restricting tasks to what learners will do in the real world (R. Ellis, 2017, pp. 508-509). Nonetheless, as Skehan (2007) also notes:

> That does not mean, of course, that a real-world relationship is inappropriate, and indeed, many tasks will have such a relationship. But it is the acceptance, on the part of the learner, that *language should be used meaningfully that is the major factor*” (Skehan, 2007, p. 291, emphasis added).

For many teachers, Skehan’s and R. Ellis’s position may be considered more tenable than Long’s position, especially in the contexts where the L2 is learned as part of the school curriculum rather than based on an analysis of learners’ real-world communicative needs (Bygate, 2016; Lambert, 2010). This is particularly the case where tasks can be used to create conditions where leaners can practice using their L2 meaningfully, which for many educators constitutes an effective way to promote L2 use and development (Edwards & Willis, 2005).

The current study employed pedagogic tasks that were specifically designed to promote ‘interactional authenticity’ as opposed to ‘situational authenticity’. That is, the tasks were intended to encourage learners to use an L2 meaningfully to achieve certain communicative outcomes that only indirectly reflected their real-life communicative needs. Furthermore, as will be discussed in greater detail in Chapter 5, the tasks used in the current study were designed in a way that would potentially help learners consolidate their L2 knowledge so they would be able to use the knowledge more accurately and fluently in relevant communicative contexts.

2.1.2 The Role of Tasks in Promoting Learning of Specific L2 Forms

According to R. Ellis (2003) tasks may be of two types, either 'focused' or 'unfocused'. As indicated in Chapter 1, focused tasks are those that necessitate the use of specific, pre-determined L2 forms at the level of production and/or comprehension. Unfocused tasks, on the other hand, comprise tasks which allow a wider range of L2 forms to be used by learners. R. Ellis (2003, 2009, 2017) also claims that both focused and unfocused tasks have a place in TBLT in that they facilitate development and use of different domains of an L2.

For instance, whilst still maintaining a focus on meaning, focused tasks can be simultaneously used to help L2 learners distinguish between singular and plural forms in a target language, leading to their incidental learning of these forms (Erlam & R. Ellis, 2018; Shintani & R. Ellis, 2010). Alternatively, unfocused tasks may be used to provide learners with opportunities to employ their L2 resources to complete pedagogic versions of target tasks (Long, 2015, 2016; Robinson, 2015), or to ensure balanced development between complexity, accuracy and fluency in using the L2 (Skehan, 1996, 1998a, 2007, 2018).

Skehan and Long, however, caution that the use of focused tasks in TBLT may result in learners focusing on pre-determined L2 forms rather than on using their L2 meaningfully to achieve a relevant communicative outcome. Skehan (2007), for instance, writes "[in the case of focused tasks], we cannot be so sure that there is a primacy of meaning; whether this is so seems to depend a lot on teacher activity" (p. 295). For this reason, Skehan (1998b) refers to focused tasks as 'structure-trapping' tasks. Similarly, Long (2015, 2016) dismisses the use of focused tasks and contends that 'genuine' versions of TBLT do not include teaching of pre-

determined L2 forms, either overtly or covertly. Thus, both Long and Skehan insist that all tasks used in TBLT should be unfocused (Skehan, 1998b, 2014; Long, 2015, 2016).

In addition, Skehan and Long reject the use of focused tasks based on theoretical grounds (e.g., Long, 2015, 2016; Skehan, 1998a, 2014). That is, both believe that L2 development is learner-driven and, in particular, that the "*teachability* [of L2 forms] is recognized as being constrained by *learnability*" (Long, 2015, p. 13, emphasis in original). That is, having learners attend to specific, pre-determined L2 forms while attempting to comprehend and/or produce meaning in the L2 may not facilitate their L2 development, especially when they are not developmentally ready to do so (Pienemann, 1984, 1989; Pienemann & Keßler, 2012). However, other researchers have shown that the use of focused tasks in comprehension- and/or production-based lessons can be an effective tool to help learners acquire pre-determined L2 forms (e.g., Kim, 2012; Shintani, 2012). Even Long (in Long, Inagaki & Ortega, 1998) found that learners can negotiate for meaning and receive intensive interactional feedback related to particular, pre-determined L2 forms in a way that facilitates the development of their knowledge of these forms.

An alternative way to view the relationship between the use of tasks and the use of specific, pre-determined L2 forms by learners in naturally-occurring task-based L2 communication is proposed by Loschky and Bley-Vroman (1993). According to these authors, the involvement of specific L2 forms in the performance of tasks may be of three types: 'task-natural', 'task-useful', or 'task-essential' (p. 132). As defined by these authors, 'task-naturalness' refers to the condition wherein the use of specific L2 forms can arise naturally during the performance of tasks, but that these tasks can still be completed easily without the use of these forms. By contrast,

‘task-usefulness’ refers to the condition wherein the use of specific L2 forms can facilitate successful completion of tasks. In other words, although it is possible for learners to complete the tasks in question without the use of specific L2 forms, the use of these forms can make successful completions of the tasks easier. Finally, ‘task-essentialness’ refers to the condition wherein tasks cannot be completed successfully without the use of specifc L2 forms. To this end, the notion of ‘task-essentialness’ proposed by Loschky and Bley-Vroman (1993) largely corresponds to the notion of ‘focused tasks’ proposed by R. Ellis (2003, 2009, 2017)—in that both are related to the idea that certain L2 communication tasks can necessitate the use of specific, pre-determined L2 forms on the part of learners so that they are able to complete the tasks successfully.

With regard to L2 development, a number of instructed L2 development studies appear to provide empirical support to the utility of task-essentialness (or focused tasks) in facilitating L2 development (e.g., Long, Inagaki & Ortega, 1998; Mackey 1999; Marsden & Chen, 2011; Morgan-Short & Bowden, 2006; Sanz & Morgan-Short, 2004; Shintani & R. Ellis, 2010). That is, these studies indicate that the use of such tasks in comprehension- and/or production-based L2 instruction is effective in stimulating learner selective attention to target L2 forms in ways that facilitate learning of these forms.

It is worth noting, however, that the current study *did not* employ ‘focused tasks’ or tasks that necessitated the use of a specific, pre-determined (target) L2 form. That is, learners were able to complete given tasks successfully by using other forms that manifest similar pragmatic functions to those targeted in the current study (i.e., English comparative structures). For instance, instead of saying ‘*X is smaller than Y*’, learners were able to say ‘*X is small. Y is big*’ to mean the same thing.

Nonetheless, the use of the target form (English comparative structures) on the tasks often proved 'useful' for learners to meet the functional demand of the given tasks effectively (see Chapter 6, for details; see also Lambert, 2019, for empirical support).

2.1.3 Assessment of Task-Based L2 Learning

Just as there are differences of opinion about how tasks can be used in L2 teaching, so too are there a range of perspectives about how task work should be assessed. For example, there are those who suggest that learners' abilities to complete given tasks can be the basis for assessment. That is, task work needs to be assessed based on learners' ability to arrive at a specified communicative outcome, regardless of which L2 forms they use—e.g., following street directions on a map to reach a specific destination (Long, 2015, p. 330) or, in the case of the current study, helping the interlocutor choose correct items from a set of very similar items (Lambert, 2019; Lambert & Nakamura, 2019). Indeed, Long (2015) maintains that adding the criteria of appropriateness or grammaticality of the use of L2 forms into task-based assessment might compromise the integrity of tasks in TBLT. He notes:

> The danger of opting for the addition of a linguistic "caboose" to a test of task-based abilities is that it can quickly lead to difficult questions regarding the frequency and/or degree of ungrammaticality or inappropriateness that will be tolerated. . . . Worse, introduction of a linguistic caboose could eventually lead to a reorientation of a task-based course, as a result of washback, to one which devotes progressively larger segments of class time to work on language as object (Long, 2015, p. 333).

On the other hand, R. Ellis (2017) argues that "teachers often say they can't assess students in TBLT when there are no explicit linguistic targets" (p. 520). Hence, for R. Ellis it is inevitable to include learners' abilities to use L2 forms in the assessment of task-based L2 learning.

Given that TBLT aims to help learners develop abilities in *using an L2* in communication (Long, 1985, 2015), it may be too hasty to eliminate language focus from the assessment of task-based L2 learning. Nonetheless, basing task-based L2 assessment only on learners' use of specific L2 forms may also undermine the purpose of TBLT, which is to help learners use an L2 *as a tool* to achieve relevant communicative outcomes (R. Ellis, 2003, 2009). One way to reconcile Long's and R. Ellis's position with regard to task-based L2 assessment may be to examine (1) whether learners are able to complete given tasks successfully, and (2) how they use certain L2 forms to complete the tasks. In the current study, these two criteria for assessing task-based L2 learning were considered. Specifically, point (1) was used as a basis for recruiting participants. That is, only those (prospective participants) who could complete a recruitment task (see Appendix 1) which resembles those used in the main study (see Appendix 2) were recruited into this study. In this way, point (1) was used to help minimize individuals' variation with respect to their task performances. With respect to point (2), it was used as a means for examining how pre-modified input and interactional feedback, which were provided within the context of task-based L2 instruction, facilitated improvements in learners' use of specific L2 forms (i.e., forms that the learners did not use previously or used only with partial control). To do so, the current study employed 'local' rather than 'global' measures of L2 learning.

In the context of task-based L2 research, 'global measures' are typically used to examine the effects of tasks on complexity, accuracy and fluency of learners' L2 use (to be elaborated in Section 2.2.2, below). In contrast, 'local measures' are typically used to track the development of accurate use of specific L2 forms (e.g., article usage, plural marker -*s* in English, or—in the case of the current study—

English comparative structures) before and after learners engage in task-based L2 instruction. Although both global and local measures have merit, they are still considered problematic when used as a basis for measuring L2 development (Foster & Wigglesworth, 2016; Lambert & Kormos, 2014; Pallotti, 2009; Plonsky & Kim, 2016; Sanz & Grey, 2015; Skehan, 2014). The measurement issues and how they are addressed in the current study are discussed in detail in Chapter 5.

2.2 Different Theoretical Perspectives Underpinning TBLT

As discussed briefly in Chapter 1, advocates of TBLT agree on two key pedagogic principles underlying TBLT: meaningful L2 use on the part of the learner and opportunities to 'focus on form' (FonF). Notwithstanding, they do not agree as to how FonF may best be realized in TBLT.

According to Long (2015), the first proponent arguing for the prime importance of incorporating FonF in TBLT (e.g., Long, 1991, 1997), FonF is primarily reactive. That is, it constitutes a reaction to a perceived problem that occurs during L2 communication. To date, however, FonF has been extended by other task-based L2 researchers and is purported to encompass all kinds of attempts to draw learners' attention to L2 forms while engaging in meaning-focused activities (Doughty, 2001; Doughty & Williams, 1998; R. Ellis, 2016; Skehan, 2003). Such an extension to the operationalization of FonF in TBLT reflects different theoretical accounts of L2 development. For example, while Long's position represents an interactionist perspective (e.g., Long, 1996, 2015), others encompass cognitivist (e.g., Robinson, 2001, 2005b; Skehan, 1998a, 2018), socioculturalist (e.g., Lantolf, 2000; Swain, 2000), and modular perspectives (e.g., R. Ellis, 2018b, 2018c). These four theoretical accounts are reviewed in the next four subsections.

2.2.1 Tasks from an Interactionist Perspective

Long (1985)'s initial TBLT proposal was closely aligned with what was known of the processes of SLA at the time, particularly the theories expounded by Krashen (1982) and Pinemann (1984). Krashen's proposal derived largely from his standard monitor theory. In contrast, Pienemann derived his proposal from his work on the teachability hypothesis.

According to Krashen (1981, 1982, 1985), L2 development takes place mainly through exposure to comprehensible input, i.e., samples of meaningful L2 use provided to learners in a way that they can comprehend. Furthermore, for Krashen (1998), neither explicit instruction nor explicit practice in producing L2 forms plays a critical role in fostering L2 development. Rather, he claims that the ability to use an L2 develops through the process of comprehending meaning during L2 communication. Hence, according to Krashen (1981, 1982, 1985, 1998), it is crucial to provide learners with plenty exposure to comprehensible input to foster L2 development.

To promote input comprehensibility, Krashen contends that L2 input to which learners are exposed during L2 communication or L2 instruction needs to be linguistically simplified. However, such simplification also needs to incorporate L2 forms that learners have not yet acquired—i.e., what Krashen refers to as '$i + 1$' (see also Chapter 4, Section 4.2).

Pienemann, however, rejects Krashen's proposal. He argues that "The input hypothesis [which Krashen advocates] is highly speculative: It is not supported by any direct empirical evidence" (Pienemann, 1985, p. 47). Furthermore, he asserts that exposure to comprehensible input will not automatically lead to the

development of L2 productive skills. This is because the two processes (i.e., input comprehension and L2 production) rely on different cognitive processing. Finally, Pienemann (1985) maintains that linguistic simplifications, which Krashen suggests are a way to promote input comprehensibility, may, in fact, lead to early fossilization and a stigmatized L2 variety among L2 learners.

Having outlined the shortcomings of Krashen's proposal, Pienemann (1985) suggests that selecting and grading L2 forms that fall within the areas of learners' L2 processing constraints may prove more beneficial in promoting L2 development. To this end, he has proposed the teachability hypothesis (see e.g., Pienemann, 1984) to identify what is learnable at a given time and, thus, what is teachable at that time. Furthermore, Pienemann (1985) argues that it is necessary to group learners in terms of their current L2 developmental stage for the purpose of instruction.

In a review of these two proposals, Long contends that "both Krashen's and Pienemann's arguments are flawed" (Long, 1985, p. 84). Nonetheless, he also acknowledges some strengths in both of their arguments. For instance, like Krashen, Long also argues that comprehensible input is essential to facilitate L2 development. However, unlike Krashen, Long holds that comprehensible input is not simply facilitated by linguistic simplification, but rather by interactional modifications (Long, 1983b, 1983c).

Long also agrees with Pienemann in that L2 development is mainly constrained by learners' L2 developmental stage. However, Long (1985) clearly sees that it is impractical to group learners based on their current stage of L2 development for instruction. He proposes, instead, that learners perform communicative tasks to

promote their L2 development. Finally, Long (1985) suggests that the process of selecting and grading instructional materials should be done in line with the level of complexity of particular tasks, rather than the complexity of L2 forms as Pienemann suggests.

In general, Long's (1985) proposal for TBLT is underpinned by his interactionist perspective of L2 development. The *early* version of this perspective suggests that L2 development is facilitated by comprehensible input which learners obtain as they engage in negotiated interaction, i.e., interaction that allows learners and their interlocutors to modify their interactional flow to avoid or overcome communication breakdown. For instance, during negotiated interaction, learners may 'request clarification' from their interlocutor to help them (better) understand what is being said. When responding to the request, the interlocutor may then 'modify his or her output' to facilitate the learners' comprehension. The interlocutor may also 'check the learners' comprehension' by asking a simple question such as '*do you understand?*'. Long (1981a, 1981b, 1983b, 1983c) argues that this sort of 'negotiation for meaning' that takes place during negotiated interaction helps to make input more comprehensible and, thus, potentially support L2 development (see Chapter 3, Section 3.2.2, for further discussion of how this process of negotiation for meaning may facilitate incidental L2 learning).

Several L2 studies in the late 1980s and early 1990s were motivated by Long's early version of interactionist perspective and Krashen's monitor theory. In general, these studies were conducted to examine how simplified input and negotiated interaction affect learners' L2 comprehension and/or their L2 development. Results from these studies suggest that negotiated interaction is more effective than—or, at least, as effective as—simplified input in promoting L2 comprehension (e.g., Gass

& Varonis, 1994; Pica, Young & Doughty, 1987; Yano, Long & Ross, 1994). However, an early study by Loschky (1994) found no significant difference between simplified input and negotiated interaction in promoting L2 development.

In the *updated* version of his interactionist perspective, Long (1996) further takes into account other processes that are pertinent to negotiated interaction, such as the provision of 'interactional feedback' that learners receive from their interlocutors (whether native or proficient non-native speakers) concerning their non-target-like L2 use. Long (1996) also discusses the role of 'modified L2 use' that the learners produce following feedback from their interlocutors. In general, he suggests that this type of negotiation work facilitates L2 development "because it connects input, internal learner capacities, particularly selective attention, and output in productive ways" (Long, 1996, p. 452).

To this end, while the *early* version of Long's interactionist perspective assumed that 'input modifications' which learners receive during task-based L2 interaction facilitate L2 development, the *updated* version of his interactionist perspective gives a prominent role to the provision of 'interactional feedback' following the production of the learners' non-target like L2 use during task-based L2 interaction. Many task-based L2 studies conducted since the late 1990s were further motivated by this latter perspective (e.g., R. Ellis & He, 1999; Gass & Torres, 2005; Han, 2002; Iwashita, 2003; Leeman, 2003; Mackey, 1999; Mackey & Philp, 1998; Oliver, 1995, 1998, 2000, 2002; Oliver & Mackey, 2003; Saito, 2015; Strapp, Helmick, Tonkovich, & Bleakney, 2011). In general, these studies indicate that learners who have opportunities to receive interactional feedback from their interlocutors on their non-target-like L2 use outperform those who simply listen to

the speech of their interlocutors, be it simplified through linguistic modifications or interactionally-modified through negotiated interaction.

As will be discussed in greater detail in Chapter 4, the current study is motivated by Long's *updated* theoretical claims. That is, this study seeks to determine whether the provision of interactional feedback is more effective than repeated exposure to input containing models of task-based L2 performances in promoting L2 development.

2.2.2 Tasks from Cognitivist Perspectives

The cognitivist perspective underlying TBLT was first proposed by Skehan (1996, 1998a) and is primarily concerned with determining the effects of task implementation on learners' L2 use (Skehan, 2014, 2016, 2018). In his early proposal, Skehan (1996) elucidates how the primacy of meaning that tasks promote may drive learners' attention away from form and, thus, potentially hinders the development of their L2 resources. Furthermore, Skehan (1996) argues that neither linguistic simplification nor negotiation work aimed at comprehensibility can address this problem. Skehan (1996) writes:

> processing language to extract meaning does not guarantee automatic sensitivity to form, and the consequent pressures for interlanguage development which is assumed by supporters of the Input Hypothesis . . . Further, there is widespread use of communication strategies. These, too, help the learner succeed with meaning while having the consequence of sometimes bypassing form (Skehan, 1996, p. 41).

According to Skehan (1996, 1998a), learners' working memory capacity is limited in its ability to process information, and this makes simultaneous attention to meaning and forms during input comprehension difficult for learners (see also VanPatten, 1990, for similar arguments). Skehan (1996, 1998a) also notes that L2

learning that relies solely on input comprehension is insufficient to promote abilities in using an L2 productively. For this reason, Skehan proposes that having learners perform ‘output-based tasks’ (see Chapter 1) under certain performance conditions is also necessary to foster their productive L2 use as well as their L2 development (see also Swain, 1985, 1995; Lapkin, Hart & Swain, 1991, for similar arguments).

In order for output-based tasks to facilitate L2 development (and not only to stimulate productive L2 use), Skehan (1996, 1998a) argues that such tasks need to be implemented in ways that encourage dual-mode processing: *rule-based* and *exemplar-based* processing. Rule-based processing underlies learners’ abilities in using L2 forms to construct novel utterances in the L2 (i.e., utterances unlike those that learners have encountered or produced before). In contrast, exemplar-based processing underlies learners’ abilities to make use of routine constructions (or chunks) that they frequently use or encounter when attempting to comprehend and/or produce meaning in the L2.

Based on the rule-/exemplar-based processing distinction, Skehan (1996, 1998a) proposes how certain tasks may be implemented in ways that will affect different modes of cognitive processing underlying L2 use. To do so, he distinguishes three dimensions of L2 use that are susceptible to conditions under which tasks are performed. These include complexity, accuracy and fluency (CAF) (Skehan, 2009a, 2009b). Complexity refers to the extent of which learners can combine two or more clauses into a single clause as well as use a range of different words. Accuracy refers to error-free use of grammatical forms. Fluency refers to rapid production containing low frequency of pauses and repairs (Bui & Skehan, 2018;

R. Ellis & Barkhuizen, 2005, Ch. 7). Skehan (1996, 1998a, 2001, 2007) further argues that performance in each dimension of CAF requires the involvement of learners' attentional capacity. That is, allocating attentional capacity to one dimension may result in the other dimensions suffering (Skehan, 2009a, 2009b, 2014, 2016, 2018). He notes:

> there is a tension between form (complexity and accuracy), on the one hand, and fluency, on the other. Then, within form, one can contrast attention directed to using challenging language (complexity) relative to conservative, less advanced language, but greater accuracy (Skehan, 2009a, p. 511).

Throughout his work, Skehan consistently argues that learners' attentional capacity is limited, and this causes a trade-off effect between the three dimensions, particularly those concerned with form (i.e., complexity and accuracy). Thus, in relation to task-based L2 production Skehan maintains that learners may be able to perform well on one, or two (at best), of the three dimensions of L2 use (e.g., on accuracy only, or on accuracy and fluency). It is uncommon, he notes, to find that learners are able to perform satisfactorily on the three dimensions simultaneously, or that their L2 use is both accurate and complex at the same time (Skehan, 2009a, 2009b, 2016; Skehan & Foster, 2001). This theoretical position is widely known as a Trade-Off or Limited Capacity Hypothesis (see Skehan, 2018, for a recent overview of this hypothesis).

Skehan, however, is not the only advocate of TBLT suggesting a cognitivist underpinning in this instructional approach. Another cognitivist perspective for using tasks in L2 pedagogy is proposed by Robinson and is outlined in his Cognition Hypothesis (see e.g., Robinson, 2001, 2003b, 2005b, 2007, 2010, 2011a, 2015).

According to the Cognition Hypothesis, task characteristics play an important role in determining the kind of L2 use that will arise on tasks. This hypothesis also suggests how certain tasks may be designed in ways that will either stimulate rapid, automatic access to existing L2 resources under real-operating conditions, or promote incidental learning of new form-function-meaning mappings.

To guide research into the Cognition hypothesis, Robinson (2005b) proposes an operational taxonomy for distinguishing task characteristics based on three factors: (1) task condition (i.e., how tasks are done); (2) task difficulty (i.e., how learners perceive a difficulty associated with given tasks); and (3) task complexity (i.e., how certain elements embedded in tasks affect different level of cognitive and functional demands). According to Robinson (2005b, 2011a, 2015), these three factors affect the allocation of learners' attentional resources along two dimensions: *resource-directing* and *resource-dispersing*. Resource-directing dimensions refer to the extent to which tasks impose certain levels of cognitive and functional demands on the part of the learner during L2 communication. Thus, designing tasks along resource-directing dimensions can promote the allocation of learners' attentional resources to form (complexity and accuracy), thereby providing contexts for *L2 development*. In contrast, resource-dispersing dimensions refer to the extent to which tasks require rapid access to relevant linguistic resources during real-time L2 use. Designing tasks along resource directing dimensions can thus foster the *automatization* of currently emerging L2 systems (Robinson, 2003b, 2005b).

Based on this resource-directing/dispersing distinction, Robinson (2010, 2015) proposes a set of operational principles as a basis for designing and sequencing tasks in L2 instruction. First, according to Robinson (2010, 2015), tasks need to be

designed and sequenced according to their complexity. Second, tasks that are high in resource-dispersing dimensions need to be presented first to promote rapid retrieval of learners' L2 resources during real-time L2 use. Third, following successful performance of the previous tasks, learners need to be presented with tasks that are high in resource-directing dimensions to promote destabilization of their interlanguage systems, as well as to foster their incidental learning of new form-function-meaning mappings. In this respect, Robinson holds a similar position to Skehan in that tasks are needed to stimulate the development of learners' L2 resources and L2 use. Two key points of difference between Robinson's and Skehan's cognitivist perspectives concern the cognitive and linguistic factors underlying L2 use: (1) whether learners are only able to draw on a single attentional pool when processing information during real-time L2 production, and (2) whether complexity and accuracy can be raised simultaneously under the performance of complex tasks.

With regard to the attentional pool, Skehan argues that learners are only able to draw on a single attentional pool when completing cognitively demanding tasks, and this makes them unable to attend to both complexity and accuracy when completing such tasks. Robinson, on the other hand, argues that learners are able to draw on multiple attentional pools concurrently when attempting to cope with cognitively demanding tasks, and this can encourage them to attend to both complexity and accuracy when performing such tasks. Furthermore, Robinson argues that decrements in specific domains of L2 use (e.g., between accuracy and complexity) are not caused by learners' limited capacity to draw only upon a single attentional pool, but rather by breakdowns in "action-control" (Robinson, 2011b, p. 12). That is, when performing cognitively demanding tasks, learners may fail to

pay attention to specific features embedded in the tasks, or that they fail to retrieve relevant linguistic resources from their long-term memory to cope with the cognitive demands of the tasks in question (see Robinson, 1995, 2003a, for in-depth discussion of the role of attention and memory in L2 processing and development). For this reason, Robinson (2011a, 2011b) maintains that complexity and accuracy can be raised simultaneously under the performance of cognitively complex or demanding tasks. In fact, he argues that the more complex communicative tasks are, the more likely they are to elicit accurate *and* complex L2 use (Robinson, 2001, 2005b, 2010, 2015).

The current study is also informed by cognitivist perspectives, especially Robinson's Cognition Hypothesis. Specifically, this hypothesis is used to explain how the nature of the tasks selected for use in this study, and the conditions under which the tasks are performed by learners (e.g., as input-based or output-based tasks), may direct the learners' attention to a target L2 form in ways that improve their ability to use the form in their own production accurately and fluently. A more detailed discussion of this topic is presented in Chapters 3 and 4.

It is necessary to note, however, that studies testing the claims of Skehan's Limited Capacity Hypothesis and Robinson's Cognition Hypothesis have not resulted in conclusive findings (Han & Kang, 2018; Jackson & Suethanapornkul, 2013; Skehan, 2016). As argued by R. Ellis (2000), one of difficulties in testing these hypoetheses is related to the fact that task work is not only influenced by the characteristics of given tasks and/or the conditions under which tasks are performed, but rather is also dependent on learners who complete them. This

argument leads to the third theoretical perspective on task-based L2 performance and learning.

2.2.3 Tasks from the Socioculturalist Perspective

Task-based L2 researchers adopting a socioculturalist perspective put a greater emphasis on learners and the dialogic process that takes place between learners as they work on tasks (Lantolf, 2000; Swain, 2000). Specifically, this theoretical perspective holds that each learner brings his or her socio-cultural history and subjective interpretations to a task. Therefore, it is difficult to make predictions about the kind of L2 use that will arise during tasks. A study by Coughlan and Duff (1994) provides some evidence for this claim. These researchers (Coughlan & Duff, 1994) show that one learner performed the same task very differently on two different occasions. In interpreting this finding, the researchers suggest that tasks are subject to multiple interpretations by the learner. Thus, L2 use that arises on tasks cannot be controlled by the teacher, researcher or syllabus designer (Foley, 1991; Lantolf & Ahmed, 1989; cf., Chapter 5, Section 5.1).

Another central claim of a socioculturalist perspective that specifically relates to L2 development concerns the dialogic process that arises during task performance. Following Vygotsky (1978), who argued that the development of higher-order mental activities is mediated by social interaction, L2 researchers adopting a Vygotskyan paradigm when studying task-based L2 activities argued that L2 development is not merely a product of a solitary process that takes place within the constraints of the learner's mental activity. Rather, L2 development arises out the interaction between learners in a given situation (see e.g., Ammar & Hassan, 2018; Swain & Lapkin, 2001). Thus, a key difference between socioculturalist and

interactionist perspectives (as discussed in Section 2.2.1) concerns specific interactional processes that are stipulated to promote L2 development. That is, while the interactionist perspective maintains that *interactional modifications* such as the provision of interactional feedback and learner modified L2 use following the feedback promote L2 development (Gass & Mackey, 2015; Long, 1996, 2007, 2015; Mackey & Oliver, 2002; Oliver 1998, 2000, 2002), the socioculturalist perspective maintains that *dialogic processes* that take place as learners work on tasks mediate L2 development. Furthermore, central to these dialogic processes are episodes where a more capable interactional partner (e.g., a native speaker, a teacher or a more proficient non-native speaker) provides necessary "scaffolding" (i.e., assistance) to a novice L2 learner to help him or her solve linguistic problems (Lantolf, 2000, 2006; Swain, 2000; Swain & Lapkin, 2001). In this way, while the interactionist perspective suggests that L2 development takes place *through* interaction, the socioculturalist perspective argues that L2 development takes place *within* interaction.

A related theoretical rationale associated with a socioculturalist perspective, which is quite relevant to the current study, is Swain's Output Hypothesis (see Chapter 4). Swain (1993, 1995, 2000, 2005) proposes the Output Hypothesis as a reaction to inadequacy of comprehensible input in fostering the development of productive L2 use. Evidence for such inadequacy was obtained from a series of evaluative studies on content-based instruction and immersion programs in Canada (e.g., Harley, 1992; Harley & Swain, 1984; Lapkin, Hart, & Swain, 1991; Swain, 1985; Swain & Lapkin, 1982, 1989, 1995). Findings from these studies indicate that immersion students who received plenty of exposure to comprehensible input during classroom instruction continued to make grammatical errors and used imprecise

vocabulary when speaking or writing in the L2. According to Swain (1985, 1993, 1995, 2000, 2005), such failure in developing nativelike accuracy in using the L2 at productive level is due to the lack of opportunities that immersion students received to engage in meaningful L2 output production. According to Swain (ibid.), such opportunities are necessary to get the students mentally engaged in the *syntactic processing* (which is needed for productive L2 use), rather than solely rely on *semantic processing* (which underlies the ability to comprehend meaning in the L2).

To date, numerous task-based L2 studies have been conducted to test a socioculturalist perspective on task-based L2 learning. In general, these studies suggest that tasks provide a logical site for contextualized, productive L2 use to occur in a classroom in ways that enable the teacher, or peers, to "scaffold" as individual learners encounter difficulties when attempting to express ideas in the L2 clearly or accurately (Lantolf, 2000; Toth, 2011; Thoms, 2012). Furthermore, tasks also provide a logical site for pushing learners to produce comprehensible L2 output through, for instance, reformulating their initial, less comprehensible L2 utterances following the indication of a lack of comprehensibility by their interlocutor, thereby potentially stretching their knowledge of form-function-meaning mappings in the L2 (Izumi, 2003; Fortune & Thorp, 2001; Swain & Lapkin, 1995).

It is worth mentioning, however, that despite the theoretical support regarding the potential utility of pushing learners to produce comprehensible L2 output (Swain, 2000, 2005), a number of L2 researchers argue that such an option may be ineffective to help learners stretch their knowledge of L2 forms in ways that result in more accurate and fluent production of the forms (Doughty, 2001; Doughty &

Williams, 1998; Long, 1991, 1998; Long & Robinson, 1998; Izumi, 2002, Izumi & Bigelow, 2000; Mackey & Philp, 1998). In the current study, therefore, such a "pushed-output" condition (i.e., where learners simply performed tasks orally with no provision of input or feedback) was used as a baseline to determine learner improvements as a consequence of performing tasks alone without receiving input or feedback on the target L2 form (English comparative structures) (see Chapter 4 for details).

2.2.4 Tasks from a Modular Perspective

A more recent proposal for using tasks in L2 pedagogy is proposed by R. Ellis (2018a, 2018b) under a framework of a modular language curriculum. According to R. Ellis (ibid.), it is necessary to include both task-based and structure-based modules in L2 teaching to optimize L2 learning. However, he notes that the inclusion of these two types of modules in an L2 instructional design must be kept *separate* or *non-integrated*. More specifically, R. Ellis suggests presenting a task-based module first coupled with opportunities to focus on form (FonF) (Long, 1991; Long & Robinson, 1998) to promote the development of basic L2 fluency and incidental learning of L2 forms. Once these instructional goals are achieved, a structure-based module may then be presented to learners to orient their attention to specific L2 forms that remain problematic, or are difficult to learn under incidental learning conditions. To do so, R. Ellis (2018c) advocates using 'consciousness-raising tasks' as a way to promote learners' awareness of particular L2 forms. Consciousness-raising tasks refer to tasks that "function as problem-solving tasks where 'language' becomes the topic to be talked about" (R. Ellis 2018c, p. 12; see also Fotos & R. Ellis, 1991). According to R. Ellis, consciousness-raising tasks

have the potential to integrate focus-on-form and focus-on-meaning within a single L2 learning activity.

The modular perspective of TBLT that R. Ellis proposes is built on the assumption that different processes of L2 learning (i.e., incidental, implicit and intentional L2 learning—see Chapter 4, Section 4.2) are necessary to facilitate L2 development. That is, R. Ellis claims that while some forms are effectively learned under implicit or incidental learning conditions, others (e.g., L2 forms that are perceptually less salient and/or communicatively redundant—see Chapter 3) might be learned more effectively under intentional learning conditions (R. Ellis, 1994, 2002, 2006). Furthermore, R. Ellis argues that these learning conditions foster the development of different types of L2 knowledge: *implicit* and *explicit L2 knowledge*. Implicit L2 knowledge refers to the unconscious type of knowledge which underlies intuitive, effortless, spontaneous L2 performance. In contrast, explicit L2 knowledge refers to the conscious type of knowledge which is accessible under slow, effortful, carefully controlled L2 performance. Drawing on research on the development of implicit and explicit L2 knowledge, R. Ellis (2006) states that these two types of L2 knowledge are needed to develop a wider range of L2 resources, as well as to foster learners' abilities to use their resources accurately in communication. Furthermore, in line with what is originally claimed by DeKeyser (2003), R. Ellis (2018a, 2018b) holds that although explicit L2 knowledge is not automatically accessible under real-operating conditions, it may eventually be automatized to result in rapid retrieval during real-time L2 processing and, thus, might be indistinguishable to the processing governed by implicit L2 knowledge.

The current study does not investigate the effectiveness of a modular perspective of using tasks. Nonetheless, its theoretical underpinning (which acknowledges the role of implicit, incidental and intentional learning to promote different types of L2 knowledge) does inform the current study. This is discussed in more detail in Chapter 4.

2.3 Summary

This chapter has discussed different perspectives of the use of 'tasks' in L2 research and pedagogy. It has been argued that tasks sould be designed and implemented in ways that inspire selective attention to form, i.e., focus on form (FonF), so that task work will not only stimulate meaningful L2 use but also promote L2 development. There are many ways as to how this may be done. Depending on a theoretical perspective, advocates of TBLT have proposed that FonF may be most effective when it is realized through (1) rich exposure to *input* coupled with opportunities to *negotiate for meaning*, especially when learners encounter difficulties in comprehending the input (Long, 1983b, 1983c, 1985; Shintani, 2016; Shintani & R. Ellis, 2010); (2) the provision of *feedback* when learners produce non-target-like or less comprehensible L2 output (Gass, 1997, 2003; Gass & Mackey, 2015; Long, 1996, 2007, 2015); (3) having learners *produce output* and providing '*interactional scaffolding*', especially when they face problems while attempting to produce comprehensible L2 output (Lantolf, 2000; Swain, 2000, 2005; Swain & Lapkin, 1995); (4); manipulating *task designs* (e.g., their relative complexity—Robinson, 2001; 2003b, 2005b, 2010, 2011a, 2015) or *implementations* (e.g., by giving planning time—Foster & Skehan, 1996, 1999; Skehan, 1996, 1998a, 2007, 2009a, 2009b, 2014, 2016, 2018); and (5)

implementing a separate, non-integrated task-based *and* structured-based syllabus (R. Ellis, 2018b, 2018c). Among these distinct pedagogic and theoretical proposals, the current study aims at examining the key aspects of (1) and (2), namely, rich exposure to input coupled with opportunities to negotiate for meaning, and the provision of feedback following learners' production of non-target-like or less comprehensible L2 output. Before reviewing the role of input and feedback in L2 development further, the next chapter reviews the linguistic, cognitive and contextual factors that can mediate (or inhibit) L2 learning from input or feedback.

Chapter 3

Task-Based L2 Use and L2 Development

This chapter discusses theoretical perspectives on the relationship between task-based L2 use and L2 development. It begins by problematizing the way that the abstract notion of L2 development has been commonly operationalized in instructed L2 development studies. Following this, the definition of L2 development adopted in the current study is presented. Next, a range of factors that are argued to play an important role in facilitating L2 development are identified. The chapter concludes by specifying how pedagogic tasks may be designed and implemented to optimize L2 development.

3.1 Instructed L2 Development

Over the last 40 years, L2 developmental research has produced some empirical evidence suggesting that L2 learners both in naturalistic and instructed settings may pass through the same L2 developmental orders and sequences (Laufer, 1990; Lenzing, 2015; Ortega, 2009; VanPatten, 2011). This research has also shown that such developmental orders and sequences cannot be easily altered by formal L2 instruction (Eskildsen, 2015; Pienemann, 1987; cf., Zhang & Lantolf, 2015). Nonetheless, this line of research also shows that instruction is effective in speeding up the rate of L2 development (DeKeyser, 2017; Doughty, 1991; R. Ellis, 1989; Long, 1983a, 2015; Norris & Ortega, 2000; Rodgers, 2006). Question remains, however, as to the type of instruction that might prove optimal in driving forward L2 development (R. Ellis, 2002; Norris & Ortega, 2000; Shintani, Li & R.

Ellis, 2013). This is because early studies in this area often did not devote adequate attention to describing the specific details of the instruction that took place (de Bot, Lowie & Verspoor, 2005).

Numerous instructed L2 development studies have been conducted to examine the effectiveness of specific instructional conditions on L2 development. The results, however, are still far from being conclusive. For instance, while a number of studies suggest that comprehension-based instruction is both effective and sufficient to promote L2 development, others suggest that such instruction is insufficient—although effective—to drive forward L2 development (for reviews see e.g., DeKeyser & Botana, 2015; Shintani, Li & R. Ellis, 2013). One reason for this discrepancy might be that the construct of L2 development has been defined and measured differently in the studies.

With regard to measurement, some studies have employed free language production tests to measure L2 development (e.g., Iwashita, 2003; Shintani & R. Ellis, 2010), while others have used only controlled production and/or sentence or word recognition tests (e.g., R. Ellis, Loewen, & Erlam, 2006; Gass & Torres, 2005; Loschky, 1994; Marsden, 2006; Morgan-Short & Bowden, 2006; Robinson, 2005a; VanPatten, Collopy, Price, Borst & Qualin, 2013). Hence, although numerous studies have been conducted to examine how certain types of L2 instruction affect L2 development, generalizing from these results is difficult (R. Ellis, 2018a, Long, 2017; Norris & Ortega, 2003).

From a theoretical standpoint, the use of tests that reflect natural language use is not always deemed necessary (Bowden, 2016; Bowles, 2011; R. Ellis, 2005; R. Ellis & Roever, 2018; Erlam, 2006; Godfroid, 2016; Gregg, 2001, 2003;

Rebuschat, 2013; Rothman & Slabakova, 2018; Schwartz, 1999; Suzuki, 2017b; Suzuki & DeKeyser, 2015; Suzuki & Sunada, 2018; White, 2003, 2015). On the other hand, the use of such tests might provide empirical evidence about whether certain instructional conditions are effective in helping learners develop new knowledge of form-function-meaning mappings, or consolidate partially learned ones, in ways that enable them to use the L2 accurately and fluently in real time. Indeed, L2 researchers generally agree that the abstract notion of L2 development may best be characterized *not* in terms of a static product of accumulating knowledge of L2 forms, but rather as a dynamic process of developing the ability to use these forms accurately and fluently in real-time communication (Bialystok, 2011; Bygate, 2016; Chater & Christiansen, 2018; Christiansen & Arnon, 2017; de Graaff & Housen, 2009; DeKeyser, 2017; N. Ellis, 2002, 2009a; R. Ellis, 2005; Hulstijn, 2002; Jiang, 2007; Krashen, 1982; Kroll, Michael, Tokowicz & Dufour, 2002; Lim & Godfroid, 2015; Long, 2015; MacWhinney, 2001b; Rebuschat, 2015; Segalowitz, 2010; Skehan, 1996).

In the current study, the term L2 development is thus used to refer to gradual improvements in *producing* L2 forms under real-operating conditions. As will be discussed in Chapter 5, these gradual improvements will involve increased *frequency*, *accuracy* and *fluency* of particular L2 *morphosyntactic structures* in relevant contexts. In this way, the current study adopts usage-based perspectives on L2 development.

3.2 Usage-Based Perspectives on L2 Development

Usage-based perspectives draw on a number of related theoretical perspectives, including cognitive linguistics (e.g., Croft & Cruse, 2004; Lakoff, 1987, 2012;

Langacker, 1987a, 2008; Talmy, 2000), construction grammar (e.g., Eskildsen, 2008; Goldberg, 1995, 2003, 2006), functional linguistics (e.g., Givón, 1995; Bates & MacWhinney, 1989), dynamic system theory (e.g., de Bot, Lowie & Verspoor, 2007; Herdina & Jessner, 2002), connectionism (e.g., Christiansen & Chater, 1994, 1999; Redington & Chater, 1998), emergentism (e.g., N. Ellis & Larsen-Freeman, 2006; MacWhinney, 1999), and complexity theory (e.g., Cameron & Larsen-Freeman, 2007; Larsen-Freeman, 2002). Two fundamental claims that are common to these theorietical perspectives are that: (i) language (L1 and L2) development constitutes a gradual process of developing mental representations of constructions (form-function-meaning pairings) through experience with meaningful, contextualized language use; (ii) the ability to use constructions in real-time communication results from the strength of the mental representations. These two fundamental claims are reviewed in detail in the four subsections below.

3.2.1 Units of L2 Development

Usage-based perspectives on L2 development (N. Ellis & Wulff, 2015; Robinson & N. Ellis, 2008) view the basic units of L2 representations are constructions, that is, form–function–meaning associations that are conventionalized in the speech community and entrenched as language knowledge in the speaker's mind (N. Ellis, 2003, p. 66). Constructions exist in multiple forms, ranging from morphemes to a fully productive syntactic pattern (S = NP + VP).

To illustrate, a short sentence in English like '*they look smaller*' consists of multiple constructions. These include (1) a function-word construction: '*they*', (2) two content-word constructions: '*look*' and '*small*', (3) a morphological construction: '*–er*', (4) a subject-verb-argument construction: *they* (subject) *look*

(verb) *smaller* (argument), and (5) an abstract-syntactic-frame construction: [S [NP [Pro(*they*)]] [VP [V(*look*)] [Adj(*small*) [morph (*-er*)]]]]. Hence, each sentence that speakers produce during natural communication normally involves multi-layered constructions. Some are quite simple and concrete; others are quite complex and abstract (N. Ellis & Wulff, 2015; Goldberg, 2003, 2006).

The degree of simplicity and concreteness of constructions is important for constructional learning (Boyd & Goldberg, 2009; see also DeKeyser, 2005; R. Ellis, 2006; Goldschneider & DeKeyser, 2001, 2005). That is, the simpler and more concrete constructions are, the easier they are to learn. It is worth noting, however, that the degree of simplicity and concreteness of particular constructions is not simply determined by their superficial linguistic features, but rather by (1) the 'communicative value' that particular constructions contribute to the meaning of sentences or utterances; (2) the 'perceptual saliency' of linguistic forms that are used to represent particular constructions, especially when these forms are used in natural communication; and, (3) the 'reliability' of these forms in representing particular constructions (N. Ellis, 2006, 2008; N. Ellis & Sagarra, 2011; MacWhinney, 2001b).

By way of illustration, although a morphological form like *–er* in English is superficially simple, this morphological form may be considered complex when being assessed using the reliability criterion of difficulty. This is due to the fact that the morpheme *–er* in English can be used with different word classes (Verb and Adjective, in particular) to create a new word class and/or new meaning of target words: *teach* (V) → *teach<u>er</u>* (N), *write* (V) → *writ<u>er</u>* (N), *small* (descriptive adj) → *small<u>er</u>* (comparative adj), *young* (descriptive adj) → *young<u>er</u>* (comparative adj). In other words, the morpheme *–er* in English is used to represent different types of

construction (in the sense of having multiple meanings and functions depending on context). Given such ambiguity, the use of this morphological form may not be easily learned by learners of L2 English (Bates & MacWhinney, 1989; DeKeyser, 2005; N. Ellis, 2008; Goldberg & Casenhiser, 2008; Goldschneider & DeKeyser, 2001, 2005; MacWhinney, 1987, 1997, 2001b).

Furthermore, the ambiguity of the morpheme *–er* in English is amplified by the fact that it cannot be instantly used with any verbs or adjectives in the language: *study* (V) ≠ **studier*, *bring* (V) ≠ **bringer*, *expensive* (descriptive adj) ≠ **expensiver*, *fun* (descriptive adj) ≠ **funner*, etc. Hence, although the use of the morpheme *–er* in English appears to follow a certain rule as described or prescribed in pedagogic English grammar books, the information provided in those books simply cannot capture the internal complexity or subtlety of which this morphological form is actually used by native speakers of English in natural communication. The same case may be true of other grammatical constructions.

That does not mean that constructions are just random combinations of letters or speech sounds. On the contrary, corpus-based linguistic research shows that the bulk of constructions existing in natural language are highly systematic and regular (Goldberg, Casenhiser & Sethuraman, 2004; McEnery, Xiao & Tono, 2006). Furthermore, the systematicity and regularity play an important role in constructional learning (Boyd & Goldberg, 2009; Dingemanse, Blasi, Lupyan, Christiansen, & Monaghan, 2015; N. Ellis, O'Donnell & Römer, 2013, 2015; Goldberg & Casenhiser, 2008; MacWhinney, 2001b; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). Nonetheless, the systematicity and regularity of constructions cannot be captured by specifying rules (Bybee, 2010; Croft & Cruse, 2004; Culicover, Jackendoff & Audring, 2017; Goldberg, 1995, 2003, 2006,

Lakoff, 1987; Langacker, 1987a, 2008). Thus, L2 instruction that emphasizes rule-based practice to help learners develop knowledge of L2 constructions is likely to be insufficient for constructional learning. This is because knowledge of constructions that native speakers possess typically consists of multiple associative networks that emerge via a process of abstraction from many exemplars that they process when engaging in meaningful, contextualized communication (Barlow & Kemmer, 2000; Behrens, 2009; Bybee, 2010; Bybee & Hopper, 2001; Christiansen & Chater, 1994, 1999; N. Ellis & Larsen-Freeman, 2009; N. Ellis & Ogden, 2017; N. Ellis, O'Donnell & Römer, 2013; Goldberg, 1995; Hopper, 1998; Jurafsky, 2002; Langacker, 1987a, 2008; MacWhinney & O'Grady, 2015; Taylor, 2002; Tomasello, 1998, 2003). For this reason, usage-based approaches to L2 development posit exposure to such rich input as crucial in approximating native-speaker knowledge of a language (Archard & Niemeier, 2004; Beckner et al., 2009; Bybee, 2008; Cadierno & Eskildsen, 2015; N. Ellis, 2003, 2009a, 2013; N. Ellis & Larsen-Freeman, 2009; N. Ellis & Wulff, 2015; Eskildsen 2008; Langacker, 2008; Robinson & N. Ellis, 2008; Tyler, 2012).

Information about the use of constructions in the input, however, is normally available tacitly. Furthermore, in the case of oral or interactional input, the information is typically only available within a fraction of a second, making processing difficult (Baddeley, 2003; Bybee, 2010; Bybee & Hopper, 2001; Christiansen & Chater, 2016; N. Ellis, 2002, 2009a, 2019; Langacker, 2008; Saffran, & Wilson, 2003; Sharwood-Smith, 2008). For these reasons, the facilitative effects of input on constructional learning typically take place gradually, especially in contexts where the input contains ambiguous, complex or

abstract constructions (as in the case of the use of morpheme *–er* in English, as discussed above).

3.2.2 Input in the Initial Stage of Learning L2 Constructions

Usage-based perspectives thus hold that L2 development constitutes the outcome of the information processing that takes place when the learner's cognitive systems (i.e., perception, attention, working memory, and long-term memory) interact with input: perceiving, tallying, categorizing, associating and abstracting regularities and constraints of linguistic patterns embedded in the input (Bybee, 2008, 2010; N. Ellis, 2002, 2009a; N. Ellis, O'Donnell & Römer, 2013, 2015; Goldberg, 2003, 2006; Goldberg & Casenhiser, 2008; Hulstijn, 2002; Langacker, 2008; Lieven & Tomasello, 2008; MacWhinney, 2001a; Tomasello, 2003). To illustrate, learners of L2 English who already know the meaning of the words *house*, *window*, *door*, *big*, *small* and have already learned a basic English sentence structure (e.g., *this house is big*, *the window is small*), but who have never been exposed to the use of English comparative structures (ECSs) in naturalistic or instructed settings, may experience difficulty in comprehending the following expression: (1) *this house is bigger than the other* (see also Section 3.2.4, below). If this is true, the difficulty that these learners experience may stimulate the *initial stage* of learning of the target construction (ECSs). That is, upon experiencing the difficulty, the learners may *perceive* gaps in their L2 resources. At the same time, they may *attend* to part of the expression that creates the difficulty, leading them to mentally *register* that part of the expression in their short-term (working) memory (N. Ellis, 2001, 2005, 2011; Robinson, 1995, 2003a; Schmidt, 2001). Furthermore, depending on the learning context, the learners may 'negotiate for meaning' to overcome

comprehension difficulty (Long, 1983a, 1983b, 1996). In doing so, they may request clarification from their interlocutor (see also Chapter 2, Section 2.2.1 and Chapter 4, Section 4.5). In responding to the request, the interlocutor may give an equivalent expression in the learners' L1 (i.e., translating). Alternatively, the interlocutor may reformulate the expression by using a more elaborate structure that is potentially more comprehensible, as in *I mean this house looks big in comparison to the other house that I mentioned to you just now* (see Long, 2015, 2019, for details). In the case where such clarification (translating or elaboration) helps learners comprehend the meaning of the expression in (1), their initial stage of learning the target construction (ECSs) may be strengthened (in the sense that the learners are capable of assigning meaning of the expression following clarification or *interactional feedback*) (N. Ellis, 2005, 2009b, 2011; Gass, 1997; Gass & Mackey, 2015; Long, 1996, 2015, 2019; Mackey, Oliver & Leeman, 2003; Pica, 1994).

Negotiation for meaning, however, is not the only strategy that learners can use to overcome comprehension difficulties. Another strategy that can serve the same purpose is 'semantic bootstrapping' (i.e., inferring meaning of certain expressions based on context and/or prior knowledge of lexical items used in the expressions) (see Anderson & Lynch, 1988; Bates & MacWhinney, 1989; Clark & Clark, 1977; Crossley, Subtirelu & Salsbury, 2013; N. Ellis, 2019; Gass, 1989). Indeed, research on implicit and incidental L2 learning indicates that learners are often capable of assigning accurate meaning and function to new linguistic forms that they have never known without the need to negotiate for meaning (see e.g., Casenhiser & Goldberg, 2005; Denhovska, Serratrice & Payne, 2016; N. Ellis, O'Donnell &

Römer, 2013, 2015; Pham, Kang, Johnson & Archibald, 2020; Rebuschat, 2015; Rogers, 2017; Ruiz, Tagarelli & Rebuschat, 2018; Williams, 2005).

Note, however, that regardless of whether this initial stage of learning the target construction takes place through negotiation for meaning or semantic bootstrapping, knowledge of the problematic construction that the learners develop at this stage might be (i) 'highly conservative', that is, it is limited only to target exemplars that the learners encountered recently (Goldberg & Casenhiser, 2008; MacWhinney, 2001a); (ii) 'highly unrepresentative', that is, it does not encompass all the necessary information about the target construction (see Sections 3.2.3 and 3.2.4 below); and (iii) weakly encoded in their memories and, thus, may easily suffer from memory decay (Baddeley, 2003; N. Ellis, 2001; Sharwood-Smith, 2008; Williams, 2012). For these reasons, usage-based theories of language development hold that *recurrent* exposure to input is necessary (Bybee, 2006, 2008, 2010; Bybee & Hopper, 2001; N. Ellis, 2002, 2009a, 2019) to facilitate further development and entrenchment of the knowledge of a construction (e.g., ECSs).

3.2.3 Type/Token Frequency in L2 Development

To account for how recurrent exposure to input is likely to affect further development and entrenchment of learner knowledge of constructions, it is necessary to consider two kinds of input frequency that usage-based theories argue to contribute to the learning process. These are *type* and *token frequency*.

Type frequency refers to the number of distinct lexical items that can be substituted in a particular position of a certain linguistic construction (N. Ellis, 2009a, p. 143). For instance, the use of a morpheme *–er* in ECSs has a high type frequency,

because this morphological unit can be used with a large number of adjectives in English, especially those consisting of one or two syllables (e.g., *big* → *bigger*, *small* → *smaller*, *pretty* → *prettier*, *simple* → *simpler*). In contrast, the irregular comparative form involving sound or lexical changes (e.g., *good* → *better*, *bad* → *worse*, *little* → *less*, *far* → *further*) has a much lower type frequency, because this form only applies to a limited number of adjectives in the language. Nonetheless, the use of the morpheme *–er* in ECSs may still be considered a lower type frequency compared to the system of comparison that employs a grammatical particle: '*more*'. This is because this grammatical particle can be used with a larger class of words, including adjectives (e.g., *expensive* → *more expensive*, *beautiful* → *more beautiful*), nouns (e.g., *viewers* → *more viewers*, *water* → *more water*), and adverbs (e.g., *quickly* → *more quickly*, *easily* → *more easily*). In this way, type frequency is mainly concerned with the structural pattern of a linguistic construction, "whether it is a word-level construction for inflection [as in the case of the use of morpheme *–er* in ECSs] or a syntactic construction specifying the relation among words [as in the case of the use of grammatical particle *more* in ECSs]" (N. Ellis, 2009a, p. 143).

In contrast, token frequency refers to the number of lexical items that actually occur in a given slot of a particular linguistic construction, especially when this construction is used for real-life communication (Casenhiser & Goldberg, 2005; N. Ellis, 2009a; McDonough & Kim, 2009). For instance, although the morpheme *–er* in ECSs can be used with either *long* or *lengthy* to mean the same thing, as in *the movie was {longer, lengthier} than we thought*, it is possible that in real-life use the word *longer* is produced more frequently by English speakers than the word

lengthier. If this is true, it may be argued that the word *longer* has a higher token frequency than the word *lengthier*.

According to Bybee (2008, 2010; Bybee & Hopper, 2001), learner experience with high token frequency is likely to have a 'conserving effect' where target linguistic exemplars are encoded and stored as single linguistic units (or chunks) in the learner's long-term memory, leading to efficient and fluent use in future usage (see also Skehan, 1998a). To illustrate, learners who receive abundant exposure to ECSs containing a high token (e.g., '*is longer than*') at the initial stage of learning ECSs (as in, *this {movie, meeting, trip, work, road, scarf, story, etc.} is longer than the {other, previous} one*) may know what the phrase '*is longer than*' means. These learners may also be able to use this phrase accurately and fluently in real time, although they might not know the inherent linguistic information underlying the construction of this phrase (e.g., the tense, aspect, subject-verb agreement, inflection in the word *longer*: *long* + *er*, etc.).

Over time, as these learners receive exposure to various exemplars demonstrating the same structural pattern (as in, *this book is thicker than that one*, *this room is cleaner than the other*, *this building is bigger than it looks in the photo*, etc.), their knowledge of ECSs might increase, making them capable of using this structural pattern to generate different expressions by filling slots with distinct lexical items (e.g., *X is {long, short, small, fast, slow} + er than Y*). During this stage of development, restructuring in the learners' L2 resources may also occur, resulting in decreased abilities in using the structure fluently and/or accurately in real time (de Bot, Lowie & Verspoor, 2007; N. Ellis & Schmidt, 1998; Larsen-Freeman, 2002, 2013a; McLaughlin, 1987; van Geert, 1991). After repeated use, the

restructured knowledge again becomes stabilized (or entrenched), and the learners again develop greater control in using the structure in real time (ibid.).

Eventually, after encountering numerous exemplars indicating different types of ECSs (i.e., Adj + –er and more + {Adj, N, Adv}) in contextualized communication, the learners may successfully develop abstract schematic knowledge consisting of [NP] [VP [V + comparative argument]] (PP than). Such knowledge can help the learners produce ECSs with distinct lexical items (as in *this house {looks higher, has more rooms, was sold more quickly} (than the other house)*). Again, exposure to these new exemplars usually lead to further restructuring in the learners' L2 resources, making them less capable of using the structure accurately and/or fluently in real time. With repeated use, this restructured L2 knowledge becomes stabilized, resulting in the ability to use the structure accurately and fluently in real time (ibid.).

To this end, usage-based theories hold that language development is facilitated by a combination of both type and token frequency (Boyd & Goldberg, 2009; Bybee, 2010; Bybee & Hopper, 2001; Collins, Trofimovich, White, Cardoso & Horst, 2009; N. Ellis & Wulff, 2015; Rowland, 2007; Tomasello, 2003). Specifically, exposure to high type frequency promotes the productivity of language use (N. Ellis, 2002). This is because the more distinct lexical items are perceived in particular slots of a linguistic construction, the more likely that it will be deemed applicable to be used with various lexical items of the same category (N. Ellis, 2002, p. 166). In contrast, "considerable practice with a particular token … [typically] results in automaticity of production" (N. Ellis, 2009a, p. 145). This is because repeated processing of particular tokens usually leads to chunking, where target linguistic exemplars are encoded and stored as single linguistic units in the

learners' long-term memory (Bybee, 2003; 2008, 2010; N. Ellis, 2001; Langacker, 1987a).

It is necessary to note that L2 researchers differ in their positions on whether the chunks that learners use in natural communication, especially those that emerge after the learners have already developed abstract schematic knowledge of constructions, is governed by fast application of rules or by retrieval of item-based exemplars (for an overview of automaticity in L2 learning see e.g., DeKeyser, 2001; Segalowitz, 2003). Notwithstanding, they generally agree that chunks emerge as a result of consistent mapping of form, function and meaning, making effective and efficient L2 processing possible. For this reason, chunking may constitute a true condition underlying the development of automaticity (rather than a mere speed-up in language processing—see Chapter 5 for details). Despite this general agreement, questions remain as to how automaticity may be validly measured in L2 processing research (see e.g., Hulstijn, van Gelderen & Schoonen, 2009; Lim & Godfroid, 2015; Philips, Segalowitz, O'Brien & Yamasaki, 2004; Segalowitz & Segalowitz, 1993; Suzuki & Sunada, 2018). Issues concerning specific measures of automaticity that have been proposed in the L2 literature are discussed in detail in Chapter 5.

3.2.4 Cue Reliability, Input Comprehension and L2 Learning

Although many linguistic constructions are used in highly regular, systematic fashion in natural language, making inductive constructional learning possible (see Section 3.2.2 above), different levels of ambiguity also exist. In English, for instance, the same forms may be used to realize different meanings (e.g., *bark* of a tree or of a dog) or different grammatical functions (e.g., I *am* work*ing* now vs. I

am work*ing* tomorrow). Likewise, different forms are also often used to realize the same meaning (e.g., *many* or *a lot of* Xs) or the same grammatical function (e.g., X is easi*er* than Y, but X is *more* difficult than Z). Furthermore, a number of well-formed sentences may also be ambiguous in different ways, especially when they are presented out of context (e.g., *the officer threatened a man with a gun*; *he saw her duck*). To account for how these kinds of ambiguity affect input comprehension and language learning, usage-based perspectives draw largely on the Competition Model proposed by MacWhinney and associates (e.g., Bates & MacWhinney, 1987; MacWhinney, 1987, 1997, 1999, 2001a, 2001b; MacWhinney & Bates, 1989; MacWhinney, Bates & Kliegl, 1984).

The Competition Model attempts to account for how the existence of multiple cues in language use affect the ease or difficulty of input comprehension and language learning. The central claim of the Competition Model is that cue reliability (i.e., when a formal linguistic feature entails a single interpretation) facilitates input comprehension and language learning. In contrast, less reliable cues (i.e., when a formal linguistic feature is associated with multiple interpretations) are likely to slow down these processes. This is because less reliable cues are likely to attract competition during processing for meaning (MacWhinney, 2001b; for empirical support, see, e.g., Brooks, Braine, Catalano, Brody & Sudhalter, 1993; Kempe & MacWhinney, 1998; Taraban & Roark, 1996).

To illustrate, two sentences like (i) *Indonesians eat rice* and (ii) *Indonesians eat more rice than Australians* clearly differ in their syntactic and semantic complexity. Such differences might impose different levels of processing demands on the part of listeners (see Clark, 1970; Kennedy, 2007). From the perspective of the Competition Model, the first sentence is likely to be easier to process than the

second. This is because the first sentence contains highly reliable and non-competing cues. In contrast, the cues in the second sentence compete. In terms of 'agent identification', for instance, the first sentence clearly indicates which NP can take the role of 'agent', whereas the second sentence does not. Based on a connectionist model (MacWhinney, 2001b, pp. 70-71; see also N. Ellis, 2002, pp. 157-160), the first NP in the first sentence fulfils all necessary conditions required for 'agency' (e.g., preverbal positioning [*Indonesians* before *eat*] and subject-verb agreement [*eat* agrees with *Indonesians*, rather than with *rice*]). In contrast, agent identification in the second sentence is more complex. That is, although preverbal positioning is clear (i.e., *Indonesians* before *eat*), the subject-verb agreement is ambiguous (*eat* agrees in number with both *Indonesians* and *Australians*). Based on this cue, both NPs could take the role of 'agent'. Given that the second sentence consists of two agents, the learner might need to refer to more specific cues that can indicate the relationship of these two agents to the verb (*eat*) and, thus, to be able to arrive at an accurate interpretation. With regard to the second sentence, the 'scalar cue' appears to provide the necessary information for this purpose. This clue is indicated by a comparative structure (more + N + PP (than-phrase)) (see also Clark, 1969, 1970; Clark & Card, 1969; Kennedy, 2001).

According to the Competition Model, there are four dimensions of cue distribution that determine cue reliability, input comprehension and language learning. These include *task frequency*, *cue availability*, *simple reliability* and *conflict reliability* (MacWhinney, 2001b).

'Task frequency' counts how often particular tasks are performed or experienced by learners (e.g., narrating stories verbally, or—in the case of the current study—

describing or listening to descriptions of very similar items and comparing their features in order to identify them).

‘Cue availability’ concerns (i) how often a particular cue occurs in the input, and (ii) how contrastive the cue is in relation to other cues in a given context. To illustrate (i), morpheme *–er* or *more* is always used in the ‘regular form’ of ECSs to indicate comparative arguments, whereas the PP (than-phrase) is largely optional. That is, the comparative intent manifested in sentences containing the use of ‘Adj + –er’ or ‘more + {Adj, N, Adv}’ can still be understood even without the use of the ‘than-phrase’ (PP), as in *this house* *<u>looks {bigger, more beautiful}</u> in the photo*. With respect to (ii), the existence of multiple NPs in sentences containing ECSs is contrastive only when the NP(s) relate directly to the comparative argument mentioned in the sentence. For instance, two phrases like *this photo* and *that photo* in a sentence like ‘*the house in this photo looks {bigger, more beautiful} (than the one in that photo)*’ tell nothing about the comparative argument in the sentence. That is, the two phrases (*this photo* and *that photo*) are only used in the sentence to provide more information about target referents being compared, namely, *two houses* depicted in two photos.

‘Simple reliability’ refers to the condition where a particular cue conveys only one meaning or function. For instance, the use of ‘Adj + –er’ or ‘more + {Adj, N, Adv}’ in English sentences reliably indicates that the sentences entail comparative arguments (e.g., *this house is bigger, but it is more expensive*).

Finally, ‘conflict reliability’ refers to the condition where a particular cue is contrastive in relation to other cues. For instance, the use of ‘Adj + –er’ reliably contrasts with the use of ‘V + –er’ in terms of their meaning and function.

However, the use of morpheme '–er' and 'more' in ECSs is not so reliable for the reasons noted earlier (see Section 3.2.1, above).

With these distinctions of cue-reliabity dimensions in mind, learners' attention to and processing of a comparative marker, especially 'Adj + –er', may be enhanced when the input does not explicitly compare two or more referents. By way of illustration, learners who receive input stating that '*X is bigger*' in the context where they need to identify (subtle) differences manifested in two very similar items (e.g., X and Y) will likely need to attend to or process the use of the morpheme *–er* attached to the adjective '*big*' in order to grasp the meaning manifested in the input accurately (e.g., X and Y are the same, except for their size). In contrast, when the input explicitly compares differences of these two referents, as in *X is bigger than Y*, it is possible that the use of morpheme *–er* is 'blocked' or 'overshadowed' by the presence of the word '*than*' (see N. Ellis, 2006; N. Ellis & Sagarra, 2010). This is because the literal meaning of the word '*than*' clearly entails a comparison. Furthermore, the word '*than*' is perceptually more salient than the morpheme '*–er*' (see also O'Grady, 2015; Talmy, 2008; VanPatten, 2002, 2015a, 2015b).

Notwithstanding, the processing condition illustrated above may not hold in the case of the grammatical particle '*more*'. This is because this grammatical particle is perceptually salient. Furthermore, this grammatical particle manifests a literal meaning which clearly indicates comparison. Hence, regardless of whether the input contains an expression like '*X is more upright*' or '*X is more upright than Y*', the online processing of the particle '*more*' in the two expressions may largely be the same.

Given that the use of particle '*more*' and morpheme '*–er*' in ECSs requires somewhat different levels of cognitive processing, it is possible that, *ceteris paribus*, the use of particle '*more*' will be learned more quickly than the use of morpheme '*–er*'. This prediction, however, may not hold with learners in the current study. This is because the learners might have already had (partial) explicit knowledge of the target form (ECSs) prior to participating in this study (see Chapter 6, Section 6.3, for details). Because of this, the present study *did not* examine whether pre-modified input and interactional feedback were effective in facilitating learning of ECSs by learners of L2 English with zero knowledge of this morphosyntactic structure, but rather whether these two instructional options were effective in facilitating the consolidation of knowledge of this structure in ways that enabled the learners to use it more accurately and fluently in spontaneous L2 production.

In a review of how the four cue-distribution dimensions discussed above (MacWhinney, 2001b) are specifically related to task-based L2 pedagogy, Lambert (2019) argues that only the first dimension (i.e., task frequency) may be manipulated in L2 task design to promote different kinds of experience with L2 use. In contrast, the other dimensions (i.e., cue availability, simple reliability, and conflict reliability) are more difficult to manipulate because they relate to fixed language features (Lambert, 2019, p. 13).

Lambert (2019), however, discusses the four cue-distribution dimensions in relation to monologic (one-way) output-based tasks, where the control over L2 use totally rests on the part of the learner. With respect to the input-based and dialogic (two-way) output-based tasks (such as those employed in the current study—see Chapters 4 and 6), the other three dimensions may be more easily manipulated. For

instance, cue availability associated with the use of particular L2 forms may be manipulated by increasing the relative frequency with which these forms occur in the input or during task-based L2 interaction (see Chapter 4, Section 4.1, for illustrations). In addition, the forms can also be used with distinct lexical items (i.e., high-type/low-token frequency) to espouse simple and conflict reliability of the forms (see e.g., Denhovska, Serratrice & Payne, 2016; Goldberg & Casenhiser, 2008; McDonough & Kim, 2009). Finally, these three dimensions of cue-distribution (i.e., cue availability, simple reliability, and conflict reliability) may further be espoused by increasing the saliency during which target L2 forms occur in the input or feedback (see e.g., Shintani, 2011, 2012, 2016; Shintani & R. Ellis, 2010). In short, these three dimensions of cue distribution (cue availability, simple reliability, and conflict reliability) may be manipulated in input-based and dialogic output-based tasks by the way that the L2 is used during tasks (see also Loschky & Bley-Vroman, 1993; Long, 2019; Robinson, 2015).

To date, however, studies which examine the effects of simple/conflict reliability through manipulating type/token frequency are relatively few (N. Ellis, 2009a; N. Ellis & Collins, 2009; Robinson, 2015). Furthermore, extant studies on these effects indicate that high-type/low-token frequency facilitated only learners' L2 comprehension, rather than production of target L2 forms (McDonough & Kim, 2009; McDonough & Nekrasova-Becker, 2014; Reali & Christiansen, 2007; Révész, Sachs, & Hama, 2014; Rott, 1999; Wiener, Lee, & Tao, 2019). In contrast, studies which examined the effects of frequency (regardless of the type/token distinction) indicate that input frequency does play a role in facilitating language development (e.g., Arnon & Snider, 2010; Crossley, Skalicky, Kyle, & Monteiro, 2019; Janssen & Barber, 2012; for a review see e.g., N. Ellis, 2002). For this

reason, the present study considers only frequency as a factor that potentially mediates the effects of input or feedback. Chapter 4 discusses this issue in greater detail.

3.3 Summary

This chapter has defined the construct of L2 development as conceptualized in the present study. It also has discussed a range of factors that potentially play an important role in promoting and/or driving forward L2 development. In line with usage-based perspectives on L2 development, it has been argued that L2 development constitutes the outcome of information processing that takes place as learners' general cognitive systems interact with input. Further, abundant exposure to L2 forms is necessary to strengthen learners' L2 knowledge so they can use it accurately and fluently in real time. Questions remain, however, as to whether recurrent exposure to particular L2 forms provided through pre-modified input or interactional feedback differ in their effects in facilitating learning and/or use of the forms in spontaneous L2 production. This issue is discussed in Chapter 4.

Chapter 4

Pre-modified Input and Interactional Feedback

This chapter discusses the theoretical foundation for the current study with respect to the independent variabels in the study ('pre-modified input' and 'interactional feedback') in L2 development. To begin, an operational definition of the terms 'pre-modified input' and 'interactional feedback' as they are used in the study are provided. Following this, distinct theoretical perspectives that underpin the relative utility of pre-modified input and interactional feedback in L2 development are reviewed. Furthermore, criticisms that have been levelled against those perspectives are discussed. Key empirical studies that have been conducted to investigate the effectiveness of pre-modified input and interactional feedback are then critiqued. The chapter concludes by outlining how certain pedagogic tasks may enhance L2 learning by way of pre-modified input and/or interactional feedback.

4.1 Defining 'Pre-modified Input' and 'Interactional Feedback'

As briefly noted in Chapter 1, the term 'pre-modified input' is used in this study to refer to *model performances* in task-based L2 instruction which have been prepared to contain various exemplars indicating how a target L2 form—namely, English comparative structures (ECSs), may be used in contextualized communication, as illustrated in Excerpt 4.1.1.

Excerpt 4.1.1:

Researcher : ***I want you to get me a girl's hat***

Learner : *ok*

Researcher : ***the brim of the hat slopes down from the left to the right***

: ***so the left brim of this hat IS HIGHER THAN the right one***

Learner : there are two [items that look like that]

On the other hand, the term 'interactional feedback' is used to refer to specific strategies such as *negotiation* that indicate to learners that their production of the target L2 form (ECSs) is non-target-like, as illustrated in Excerpt 4.1.2. Such *negotiation* can also show how the form may be used to express the same intended meaning, as illustrated in Excerpt 4.1.3.

Excerpt 4.1.2:

Learner : **the left side is long than right side*

Researcher : ***ok, so the left side IS LONGER THAN the right side***

Learner : *yes*

Excerpt 4.1.3:

Learner : *the left side is small, and the right side is big*

Researcher : ***do you mean the left side IS SMALLER THAN the right side?***

Learner : *yes*

As illustrated in Excerpt 4.1.1, pre-modified input was provided under a condition where the researcher took the role of 'speaker' (information provider), while the learner was assigned the role of 'listener' (information receiver). The tasks used to provide pre-modified input in the current study were 'input-based tasks'. As noted in Chapter 1, Section 1.1, these tasks require the learner to correctly process

meaning associated with the use of specific L2 forms in order to achieve the task goal (e.g., identifying a picture from a set of pictures). Input-based tasks do not require learners to produce output although they are allowed to do so if they wish (R. Ellis, 2001; R. Ellis & He, 1999; R. Ellis, Tanaka & Yamazaki, 1994; Erlam & R. Ellis, 2018; Gass & Varonis, 1994; Long, Inagaki & Ortega, 1998; Loschky, 1994; Mackey, 1999; Shintani, 2012, 2016; Shintani & R. Ellis, 2010). More detailed discussion about the pre-modified input materials used in the present study is presented in Section 4.4, below.

On the other hand, interactional feedback was provided under a condition that was opposite to the premodified input condition. That is, learners in the interactional feedback condition were assigned the role of 'speaker' (information provider), while the researcher took the role of 'listener' (information receiver). The tasks used to provide interactional feedback in this study were thus 'output-based tasks'. As discussed in Chapter 1, Section 1.1, these tasks require the learner to produce output in line with the task goal (e.g., describing a picture clearly and unambiguously so a listener can identify it from a set of pictures). In the L2 literature, the type of interactional feedback as exemplified in Excerpt 4.1.2 is commonly referred to as '*recasts*', that is, reformulations of the whole or part of learners' non-target-like L2 uses with more target-like usage and without changing the original meaning. By contrast, the type of feedback as exemplified in Excerpt 4.1.3 is commonly referred to as *confirmation checks*, that is, expressions used to elicit confirmation from the speaker that his or her utterance is correctly heard or understood (Leeman, 2003; Long, Inagaki & Ortega, 1998; Mackey, 1999, 2006; Mackey & Oliver, 2002; Nobuyoshi & R. Ellis, 1993; Oliver, 1998, 2000, 2002; Oliver & Mackey, 2003; Pica, Kanagy & Falodun, 1993; Shintani, 2011; Shintani

& R. Ellis, 2010). More detailed discussion of these two types of interactional feedback is provided in Section 4.5, below.

To this end, there are at least two key features that distinguish 'pre-modified input' from 'interactional feedback' as operationalized in the current study. First, pre-modified input affords exposure to the target L2 form *proactively* (i.e., non-contingent on learners' L2 use), whereas interactional feedback affords exposure to the target form *reactively* (i.e., contingent on learners' L2 use) (Gass, 2003; Iwashita, 2003; Long, 2015, Ch. 3; Mackey, 1999; Shintani, Li & R. Ellis, 2013; Strapp, Helmick, Tonkovich & Bleakney, 2011). Second, pre-modified input only provides '*positive evidence*' (i.e., information of what is possible in the L2), whereas interactional feedback potentially affords both '*positive*' and '*negative evidence*' (i.e., information of what is and what is not possible in the L2, or how particular expressions may be expressed more clearly or less ambiguously in the L2) (Gass & Mackey, 2015; Long, 1996, 2007, 2015; Mackey, Oliver & Leeman, 2003). Despite these differences, both pre-modified input and interactional feedback were realized in the context of this study with oral enhancement or rising intonation when the target L2 form was presented as either input or feedback (see, e.g., the words written in capitals in Excerpts 4.1.1 – 4.13; see also Shintani, 2011, 2016; Shintani & R. Ellis, 2010).

4.2 Pre-modified Input and Interactional Feedback: The Debate

Discussions of the role of pre-modified input and interactional feedback in L2 development are normally associated with research on how L2 forms are best introduced to learners to facilitate L2 development within instructional contexts. Depending on theoretical perspectives, researchers have examined the utility of

pre-modified input and/or interactional feedback in facilitating implicit, intentional or incidental L2 learning. In the current study, the relative utility of these two types of L2 instructional options were specifically investigated within incidental L2 learning paradigms (to be elaborated in Section 4.3, below). However, since the discussion of incidental L2 learning by way of pre-modified input or interactional feedback is likely to be obscure without first discussing how it differs from both implicit and intentional L2 learning, it is necessary to discuss key differences of these three types of L2 learning conditions (implicit, intentional and incidental) along with theoretical and pedagogic perspectives that underpin them.

Krashen (1981, 1982) introduced the notion of 'comprehensible input' as a means for promoting L2 development. As noted in Chapter 2, Section, 2.2.1, the term 'comprehensible input' refers to samples of meaningful L2 use provided to learners in ways that they can comprehend the meaning that is conveyed by the L2. According to Krashen (1981, 1982, 1985, 1994), L2 development refers to what learners have "acquired" (i.e., *learned implicitly*) from exposure to comprehensible input. He also notes that other processes such as explicit instruction, output production, and the provision of interactional feedback play no role in facilitating "acquisition" or implicit L2 learning (Krashen, 1998).

In order to make input comprehensible and useful for acquisition, Krashen (1982, 1985) maintains that samples of meaningful L2 use to which learners are exposed in the classroom, or in naturalistic settings, need to be 'linguistically simplified'. This may be done by controlling the frequency of unfamiliar words and complex sentence structures. Nonetheless, specific L2 forms that learners have not (fully) acquired need to be incorporated in the input (i.e., $i + 1$ in Krashen's term). In the literature, this kind of input is usually referred to as (*pre-*)*modified input* (de la

Fuente, 2002; R. Ellis & He, 1999; R. Ellis, Tanaka & Yamazaki, 1994; Gass, 2003; Gass & Varonis, 1994; Long, 1983b; Loschky, 1994; Mackey, 1999; Mackey & Abbhul, 2005; Pica, Young & Doughty, 1987).

As described in Chapter 2, Section 2.2.3, the first empirical evidence revealing the insufficiency of comprehensible input in fostering L2 development came from a series of evaluative studies of French immersion programs in Canada (e.g., Harley & Swain, 1984; Swain, 1985; Swain & Lapkin, 1982, 1989). In general, these studies revealed that students who already spent years in the immersion programs (and thus, by assumption, having already received a large amount of comprehensible input containing $i + 1$ as a result of engaging in various comprehension-based activities) still fell short of attaining native-like abilities when using the L2 in spontaneous production. In interpreting these findings, Swain (1985, 1995, 2000, 2005) argues that the students may not have had sufficient opportunities to use the L2 in production. According to Swain (ibid.), L2 production facilitates L2 development because it induces deeper syntactic processing which is often unnecessary during input comprehension (cf., R. Ellis, 1994, 1997; Gass & Mackey, 2015).

4.3 Pre-modified Input from Attentional Perspectives

An alternative account regarding the insufficiency of comprehensible input in L2 development has been put forward by VanPatten (1990, 2002, 2015a, 2015b 2015c, 2017). VanPatten argues that exposure to comprehensible input alone will not automatically lead to L2 development, especially with adult L2 learners. This is because these learners usually do not attend to, or mentally process, formal

linguistic features manifested in the input, especially when those features are perceptually non-salient and/or communicatively redundant (see also Chapter 3).

To illustrate, beginning-level adult learners of L2 English who listen to an utterance like (1) *he cleaned the room yesterday* may fail to attend to the function of the morpheme *–ed* attached to the verb '*clean*' (which indicates past tense). This is because the same function is more clearly indicated by the word '*yesterday*' (VanPatten, 2002, 2015b). Likewise, beginning-level adult learners of L2 English who listen to an utterance like (2) *the left sleeve looks slightly longer than the right sleeve* may fail to attend to the function of the morpheme *–er* attached to the adjective '*long*' (which indicates comparison), since the same function is more clearly indicated by the literal meaning of the '*than*-phrase' (see Chapter 3, Section 3.2.4, for details).

Given that exposure to comprehensible input does not always require learners to attend to, or to mentally process, relevant L2 forms in the input, VanPatten (1996, 2002, 2015a) asserts that it is necessary to get learners engaged in *intentional L2 learning* (see also Schmidt, 1990, 2001; Sharwood-Smith, 1993, for similar arguments). Notwithstanding, although VanPatten disagrees with Krashen that L2 development constitutes the outcome of implicit learning by way of comprehensible input, he agrees that output production plays no role in L2 development. In fact, VanPatten (2002) argues that requiring learners to produce output in the early stage of L2 development can inhibit "input processing" which is necessary for learning of form-function-meaning mappings.

To help learners process input accurately and effectively, VanPatten (1996, 2002, 2004; VanPatten & Cadierno, 1993) devised an instructional model called

Processing Instruction (PI). As described by VanPatten, PI aims to alter learners' default, inappropriate mental processing strategies with more appropriate ones to facilitate learning of form-function-meaning mappings. To accomplish this, PI includes three stages of instructional activities: the provision of *explicit information*, *structured-input activity*, and *affective-input activity*.

In practice, 'explicit information' is usually realized in the form of a brief explanation provided to learners as to how a specific L2 form relates to particular meaning and function. Following this, learners engage in a 'structured-input activity' to help them grasp form-function-meaning relationships. This activity normally includes 'pre-modified input' in the form of short sentences exemplifying the use of a target L2 form and other similar forms as distracters. The use of distracters in this activity aims to encourage learners to engage in active mental processing and avoid rote learning (VanPatten, 2002). Finally, learners are provided with an 'affective-input activity' to help them reinforce their previous input processing. This activity also comprises the use of pre-modified input in the form of short sentences exemplifying the use of the target L2 form (for a concise review of PI, see, e.g., Rasuki, 2017). According to VanPatten (2015a), this type of input-based L2 learning activities promotes the development of knowledge of form-function-meaning mappings that can be tapped during real-time comprehension and production.

Although VanPatten's input-based L2 instructional model (PI) has been highly influential, criticisms have been leveled against: (i) the theoretical perspective that becomes the foundation of PI (see e.g., DeKeyser, Salabery, Robinson & Harrington, 2002), (ii) the validity and generalizability of research findings that support the efficacy of PI (see e.g., DeKeyser & Botana, 2015; Rasuki, 2017;

Shintani, 2015), and (iii) the instructional components in PI that are argued to affect L2 development.

With respect to point (iii), controversy exists regarding the role of explicit information in PI. That is, researchers argue that the efficacy of PI might not solely be due to the results of input-processing activities (i.e., structured-input and affective-input activities). This is because PI also provides learners with opportunities to develop 'explicit' or 'declarative knowledge' of target L2 forms by way of receiving explicit information about the forms (DeKeyser & Botana, 2015; Morgan-Short & Bowden, 2006). From the perspective of Skill Acquisition Theory (DeKeyser, 2007, 2015, 2017), learning will be highly facilitated when learners already developed declarative knowledge of target L2 forms, regardless of whether the instruction is input-based or output-based, although—as DeKeyser (ibid.) suggests—the facilitative effects of declarative knowledge will be most evident in the skill that has been practiced. That is, declarative knowledge used in the context of input-based L2 instruction will facilitate the development of L2 comprehension skills better than L2 production skills. By contrast, declarative knowledge used in output-based L2 instruction will facilitate the development of L2 production skills better than L2 comprehension skills (DeKeyser, 1996, 1997; DeKeyser & Sokalsky, 1996).

Informed partly by VanPatten's proposal for designing and using pre-modified input in L2 instruction (PI), Shintani (2011, 2012, 2016, Shintani & R. Ellis, 2010) devised an instructional model for teaching specific L2 forms through the use of 'input-based tasks' (IBT) (see Chapter 1). Unlike PI which aims to foster *intentional learning*, IBT aims to foster *incidental learning* of form-function-meaning mappings. To do so, IBT does not involve 'explicit information'

(Shintani, 2011, 2016). Furthermore, learners are not made aware of the presence of specific L2 forms in the input. Rather, they are encouraged to focus on meaning (i.e., comprehend what the teacher says) so they are able to complete given tasks successfully.

Incidental learning also differs from implicit learning as described by Krashen (1981, 1982). That is, incidental learning does not preclude the possibility that learners manifest some level of awareness of forms when attempting to comprehend input (Shintani & R. Ellis, 2010, p. 609). By way of illustration, beginning-level young learners of L2 English who are asked to identify and select pictures of particular items based on a series of descriptions given by the teacher may eventually become aware that an *–s* sound articulated by the teacher (e.g., *squirrels, bears, batteries, bananas*) means pictures containing multiple referents. In contrast, when the teacher does not articulate an *–s* sound (e.g., *squirrel, bear, battery, banana*), it means pictures containing only a single referent (Shintani & R. Ellis, 2010). For more detailed discussions of 'implicit', 'intentional' and 'incidental L2 learning', see DeKeyser (2003), Hultijn (2003), and Leow and Zamora (2017).

Finally, IBT also differs from PI in that IBT allows, but does not require, learners to produce output while attempting to comprehend input (R. Ellis, 2001; Erlam & Ellis, 2018; Shintani, 2011, 2012, 2016; Shintani & R. Ellis, 2010). Following the *early version* of Long's Interaction Hypothesis (Long, 1981, 1983b, 1983c, 1985; see Chapter 2, Section 2.2.1), IBT predicates that allowing learners to produce output will provide opportunities for 'negotiation for meaning' to occur, especially when learners fail to comprehend the input being given. Such negotiation might prove useful to help learners comprehend the input better and, at the same time,

attend to specific forms that they have not (fully) learned, thereby facilitating L2 development (Gass, 1997; Long, 1996; Pica, 1994, 1996; see also Chapter 3, Section 3.2.2).

The current study used pre-modified input following the pedagogic framework as outlined in IBT. That is, rather than presenting pre-modified input to promote implicit or intentional L2 learning (Krashen, 1982, 1985; VanPatten, 1996, 2002), pre-modified input was designed to foster incidental L2 learning (Shintani, 2016). Chapter 6, Section 6.6.1, outlines the procedures of how pre-modified input was realized in the current study.

The efficacy of IBT has been examined in a series of classroom-based studies with beginner-level young L2 learners (e.g., Erlam, 2019; Erlam & R. Ellis, 2018; Shintani, 2011, 2012; Shintani & R. Ellis, 2010). In general, these studies indicate that IBT is effective in helping these learners use target L2 forms (vocabulary and grammar) accurately for both comprehension and production. It is worth noting, however, that given the low proficiency level of the learners in the studies, the researchers often attribute the efficacy of IBT to the 'negotiation for meaning' which frequently occurred during the tasks. Hence, it remains to be seen whether the same case is true when IBT (involving the use of pre-modified input) is used with adult L2 learners who have larger receptive L2 knowledge, but who are still struggling with producing sentential utterances fluently and accurately. It is possible, as Mackey (1999) suggests, that these learners will hardly negotiate for meaning when receiving pre-modified input, and this will potentially make specific L2 forms manifested in the input simply go unattended (see also Long, 2015, 2019).

4.4 Criticisms against the Use of Pre-modified Input in Task-Based L2 Instruction

The instructional model for using tasks to expose learners to specific L2 forms through comprehension- or input-based activities like those in Shintani (2011, 2012, 2016; Shintani & Ellis, 2010) is not new. Similar models have been employed in a number of task-based L2 studies conducted in the 1990s (see, e.g., R. Ellis & He, 1999; R. Ellis, Tanaka & Yamazaki, 1994; Gass & Varonis, 1994; Loschky, 1994; Mackey, 1999; Pica, 1994; Pica, Young & Doughty, 1987). These earlier L2 studies were concerned with investigating how different types of input provided through task-based L2 instruction affect learners' L2 comprehension and/or L2 development. To do so, researchers distinguished between two types of input: *baseline* and *modified*. Baseline input refers to samples of language use that native speakers (NSs) produce when communicating with other NSs or highly proficient non-native speakers (NNSs). On the other hand, modified input refers to samples of language use that NSs or highly-proficient NNSs produce when communicating with low-proficient NNSs.

Within the modified-input type, a distinction has also been made between *pre*-modified and *interactionally*-modified input. Pre-modified input refers to samples of L2 use that have been prepared prior to being given to low-proficient NNSs with an aim to avoiding or minimizing comprehension difficulty, thereby making negotiation for meaning unnecessary (see Excerpt 4.1.1 in Section 4.1, above, for an example). In contrast, interactionally-modified input refers to samples of L2 use that arises as a result of negotiation for meaning as triggered by NNSs' difficulties

in comprehending utterances produced by a NS or highly-proficient NNS interlocutor.

Earlier task-based L2 studies on these three types of input generally indicate that baseline input is less beneficial than pre-modified and/or interactionally-modified input in promoting L2 comprehension and L2 development for low proficient NNSs (Gass & Varonis, 1994; Pica, 1994; Pica, Young & Doughty, 1987). With respect to the utility of pre-modified and interactionally-modified input, however, the results are somewhat mixed. For example, while some studies suggest that pre-modified and interactionally-modified input do not differ significantly in facilitating L2 comprehension and L2 development (R. Ellis, Tanaka & Yamazaki, 1994; R. Ellis & He, 1999; Loschky, 1994), others indicate significant differences (Gass & Varonis, 1994; Long, Inagaki & Ortega, 1998; Mackey, 1999).

A few studies have also been conducted to examine the effects of two types of *pre*-modified input (i.e., *simplified* and *elaborated*) on L2 comprehension (e.g., Oh, 2001; Parker & Chaudron, 1987; Yano, Long & Ross, 1994). In these studies, 'simplified input' was usually realized by controlling the frequency of complex sentence structures and unfamiliar words. In contrast 'elaborated input' was achieved by adding redundancy, regularity and explicitness to complex sentence structures and unfamiliar words that commonly occur in natural communication between NSs (for a comprehensive review of simplified and elaborated pre-modified input, see Long, 2015, Ch. 9). Table 4.1 provides examples of simplified and elaborated input in relation to baseline NS data on tasks similar to those used in the present study.

In general, studies comparing the utility of simplified and elaborated input indicate that elaborated input is more effective than, or (at least) as effective as, simplified input in facilitating L2 comprehension (Oh, 2001; Parker & Chaudron, 1987; Yano, Long & Ross, 1994). Nonetheless, although a study by Kim (2006) indicates that elaborated input also facilitates incidental L2 vocabulary learning, relatively little is known as to whether elaborated input facilitates learning of L2 morphosyntax (Long, 2015, 2019).

Table 4.1. Baseline, simplified and elaborated input

Baseline version (data from Lambert's (2019) study) *The left sleeve is further from the body of the jacket than the right sleeve*.
Simplified input *The left sleeve is far from the body of the jacket, and the right sleeve is near the body of the jacket.*
Elaborated input (taken from pre-modified input materials in this study) *Now look at the sleeves of the jackets. They look quite different, right? Especially the distance between the right sleeve and the body part. My jacket is the one where the right sleeve is slightly* ***further away from the main part of the body than the left sleeve****. Or, the left-hand sleeve of my jacket is* ***closer to the main part of the body than the right sleeve.*** *So, you can see that* ***the armpit on the right side of my jacket is wider than the one on the left side.***

As indicated in Table 4.1, the pre-modifed input in the current study was elaborated rather than simplified. To do so, three main principles for elaborating input (summarized in Long, 2015, p. 252) were employed: *adding redundancy*, *adding regularity* and *adding explicitness of thematic structure*.

Using the example of elaborated input in Table 4.1, redundancy was achieved by means of repeating certain words rather than using pronouns (see, e.g., the use of the words *jackets*, *sleeves*, *the right/left sleeve*, *the body part*). Regularity was

achieved by means of maintaining parallelism (e.g., *Now look at… They look …*, etc.) and by using full NPs instead of anaphors (e.g., *my jacket … my jacket* instead of *it*). Finally, explicitness of thematic structure was achieved by means of employing intra- and inter-sentential linkers (e.g., *now*, *right?*, *especially*, *or*, *so*).

Following Yano, Long and Ross (1994, the primary purpose for using elaborated pre-modified input was to avoid or minimize comprehension difficulty on the part of learners, thereby reducing the need for negotiation for meaning to occur (see also Oh, 2001; Parker & Chaudron, 1987). Nonetheless, for the reasons noted in the previous section, designing pre-modified input in this way may reduce the need for learners to attend to specific L2 forms that they have not (fully) learned (Mackey, 1999; Skehan, 1996; Swain, 1995; VanPatten, 1996, 2002). Furthermore, although the input designed in this study affords relatively high frequency of exposure to the target L2 form (i.e., ECSs) (see Chapter 3; see also Appendix 3), some researchers argue that there is no guarantee that such exposure alone is sufficient to promote incidental learning of the form (Gass & Mackey, 2002; Granena & Long, 2013; Izumi, 2002, 2003; Swain, 1995, 2000; Swain & Lapkin, 2001; cf., Chapter 3). Finally, although learners may eventually learn the target L2 form as a result of attending to its multiple occurrences in the input, they may still use the form less accurately in their production. This is because, as noted in Section 4.1, pre-modified input affords only ‘positive evidence’, or information about grammatical forms in the L2, but does not provide ‘negative evidence’, or information about those forms that are ungrammatical (Gass & Mackey, 2015; Long, 1996, 2007, 2015, Ch. 3; Mackey, Abbuhl & Gass, 2013; Strapp, Helmick, Tonkovich & Bleakney, 2011; White, 1987). Given these possible limitations of the utility of pre-modified input in L2 development, many L2 researchers claim that providing

learners with interactional feedback at the time when they produce ill-formed sentences or encounter difficulties in conveying meaning in the L2 might prove more beneficial in facilitating L2 development. This is because such feedback might provide learners with relevant input at the time when they are likely to be most receptive (Adams, 2003; Doughty, 2001; Gass, 1997; Gass & Mackey, 2002, 2015; Long, 1996, 2007 2015; Long & Robinson, 1998; Mackey, Abbuhl & Gass, 2013; Pica, 1994; van den Branden, 1997). The claim of the efficacy of interactional feedback is discussed in greater detail in the next section.

4.5 Interactional Feedback and L2 Development

As noted in Section 4.1, the term interactional feedback is used in the context of this study to refer to two forms of implicit corrective feedback: '*recasts*' and '*confirmation checks*'. Recasts refer to reformulations of the learner's non-target-like L2 uses with more target-like uses without changing the original meaning, as illustrated in Excerpt 4.1.2. On the other hand, confirmation checks refer to expressions used to elicit confirmation from the learner that his or her utterance is correctly heard or understood, as illustrated in Excerpt 4.1.3 (see also R. Ellis, Basturkmen & Loewen, 2001; Lyster & Ranta, 1997; Sheen, 2004).

4.5.1 Recasts as Interactional Feedback

The theoretical underpinnings of the role of recasts in L2 development are based on Long's (1996) updated version of the Interaction Hypothesis. As noted in Chapter 2, Section 2.2.1, while the initial form of the Interaction Hypothesis (Long, 1981, 1983b, 1983c) holds that comprehensible input that learners receive from engaging in interaction is sufficient to promote L2 development, the updated form of the

Interaction Hypothesis (Long, 1996) describes how the exposure to comprehensible input alone as insufficient. That is, Long (1996) claims that learners also need to receive implicit corrective feedback regarding the accuracy and/or comprehensibility of their output to drive forward their L2 development. Long also notes that although 'explicit corrective feedback' such as overt error correction or metalinguistic explanation rarely occurs in natural L1 and L2 communication, 'implicit corrective feedback' such as recasts is omnipresent (Long, 1996, p. 444).

Drawing on earlier studies of L1 learning, Long (1996, pp. 430-445) provides an in-depth review indicating that parents or adults normally recast children's erroneous speech (Bohannon & Stanowicz, 1988). Furthermore, recasts appear to affect children's L1 development (Baker & Nelson, 1984; Farrar, 1990). Likewise, early studies on L2 learning in naturalistic or communicative classroom settings show that NSs (teachers or peers) recast NNSs' erroneous L2 uses (Doughty, 1994; Oliver, 1995). An early study by Richardson (1993) also indicates that recasts appear to have a facilitative role in L2 development (see also Ayoun, 2001; Mackey & Oliver, 2002). Despite this, Long (1996) is careful to point out that empirical studies that examined the utility of recasts in L2 development at that time was relatively scarce (p. 455). Nonetheless, the claim that recasts play a facilitative role in L2 development seems reasonable on logical learnability grounds (ibid.).

Among the first L2 researchers who questioned the facilitative role of recasts in L2 development was Lyster (1998a, 1998b). He argues that recasts might be ineffective to facilitate L2 development because recasts do not overtly indicate to learners that their L2 uses are problematical in some way. Thus, learners may fail to grasp the corrective intent manifested in recasts. Instead, they may perceive recasts as either confirmations or alternative ways of expressing the same meaning.

The skepticism expressed by Lyster (1998a, 1998b) regarding the utility of recasts is based on the results of a descriptive study on oral corrective feedback that he conducted with Ranta in four immersion classrooms in Canada (Lyster & Ranta, 1997). In general, this study reveals that although recasts constitute the most frequent type of corrective feedback that teachers use when responding to learners' erroneous L2 speech, recasts rarely lead to "learner uptake" (i.e., learner *immediate* response to feedback'—Lyster & Ranta, 1997, p. 49). On the other hand, explicit corrective feedback such as overt error correction (e.g., *you should say 'I went', not 'I goed'*), metalinguistic focus (e.g., *you need past tense*), elicitation (e.g., *no, it's not I goed, it's I …*), and repetition (e.g., *I goed?*) are more likely to elicit uptake from learners. For this reason, Lyster (1998a, 1998b; Lyster & Ranta, 1997) suggests that explicit corrective feedback might be more effective than recasts in facilitating L2 development, at least in the immersion context.

Many L2 researchers, however, criticize Lyster's (1998a, 1998b) and Lyster and Ranta's (1997) work. For instance, Goo and Mackey (2013) note that it might be unwarranted to compare the efficacy of recasts and explicit corrective feedback based on the extent of which they encourage learners to produce *immediate* uptake. This is because recasts do not always require learners to produce immediate uptake (Oliver, 1995, 2000). A simple acknowledgement such as '*yes*' is often pragmatically appropriate (Egi, 2010; see Excerpt 4.1.2 in Section 4.1 for an example). In the same vein, Pica (1988) notes that the fact that recasts already provide a correct form of learners' erroneous L2 use often makes it unnecessary for the learners to reformulate their erroneous L2 use. In fact, Mackey and Philp (1998) maintain that learner immediate uptake following recasts may be "red herrings" (p. 338).

Many L2 researchers also argue that the notion of uptake, as operationalized by Lyster (1998a; Lyster and Ranta, 1997), does not necessarily indicate L2 learning, especially from recasts. This is because learners may imitate modeled utterances without attending to critical linguistic information embedded in them. In contrast, learners may attend to critical linguistic information embedded in recasts without attempting to produce immediate uptake (Bao, Egi & Han, 2011; Carpenter, Jeon, Macgregor & Mackey, 2006; Li, 2018; Loewen & Philp, 2006; Long, 2007; Mackey, Gass & McDonough, 2000; McDonough & Mackey, 2006; Nicholas, Lightbown & Spada, 2001). It is thus possible that learners can be learning from recasts in spite of the fact that they do not demonstrate uptake immediately. Studies by Shintani and R. Ellis (2010) and Mackey and Philp (1998) provide relevant empirical support for the claim that lack of immediate uptake following recasts does not equate with lack of learning.

Shintani and R. Ellis (2010) document how recasts accompanied with oral enhancement provided to young L2 learners following their erroneous production of a specific L2 form (English plural *–s*) *during* output-based task instruction led to both immediate uptake and incidental learning of this form. They found that recasts did not lead to immediate uptake of the form. Nonetheless, the learners performed significantly better in using the form in both comprehension and production *after* the instruction. In this way, this study confirms that absence of *immediate* uptake from recasts does not necessarily indicate absence of L2 learning.

In an earlier study, Mackey and Philp (1998) also examined how recasts led to immediate uptake and L2 learning. Thirty-five adult learners of L2 English at different proficiently levels (beginner or low-intermediate level) were randomly assigned to three groups: interaction with recast (I +R), interaction without recast (I

–R), and control. Both the I +R and I –R groups performed the same information-gap tasks orally with a NS interlocutor, but only the I +R group received intensive recasts when producing the target L2 form (question formations in English) inaccurately (e.g., **what the cat doing?*, **is in the picture there is a cat?*, etc.). The control group only participated in pre- and post-tests with no treatments.

The results of Mackey and Philp's (1998) study suggest that recasts rarely elicited immediate uptake of the target form from learners in the I +R group *during* the instruction. Nonetheless, these learners performed significantly better *after* the instruction than those who only participated in interaction (i.e., I –R group) or performed tasks in the pre- and post-tests only (i.e., control group).

Given that *immediate* uptake does not provide an accurate measure of learning, L2 researchers now agree that it is necessary to examine the efficacy interactional feedback such as recasts through pre-, post-, and delayed post-test designs (Goo & Mackey, 2013; Iwashita, 2003; Long, Inagaki & Ortega, 1998; Mackey & Oliver, 2002; Mackey, Oliver & Leeman, 2003). The present study employed such a design to examine how the provision of interactional feedback and pre-modified input affected learners' L2 developing systems (see Chapter 6 for details).

4.5.2 Recasts vs. Pre-modified Input

Although Long (1996, 2007, 2015) proposes recasts as an alternative option to address the insufficiency of comprehensible input in facilitating L2 development, comparative studies on recasts have mostly been concerned with examining the efficacy of recasts as compared with explicit corrective feedback (e.g., Ammar & Spada, 2006; R. Ellis, Loewen & Erlam, 2006; Erlam & Loewen, 2010; Gooch, Saito, & Lyster, 2016; Lyster & Mori, 2006; van de Guchte, Braaksma,

Rijlaarsdam, & Bimmel, 2015). Given that the current study *does not* aim to compare how recasts and explicit corrective feedback affect learners' L2 developing systems, claims that recasts are more or less effective as compared with explicit corrective feedback will not be discussed further (however, see Goo and Mackey (2013) and Lyster and Ranta (2013) for in-depth discussion, and debate, of this issue).

In more recent work, Long (2015) argues that there are at least four key features that make recasts potentially more effective than mere exposure to pre-modified input. First, drawing on Saxton's (1997) "Direct Contrast Hypothesis", Long contends that recasts afford both positive and negative evidence. According to Long, these two types of evidence are necessary to help learners develop elaborate knowledge of what is and what is not grammatical in the L2. In contrast, pre-modified input affords only positive evidence and, thus, may fail to help learners learn what is not grammatical in the L2 (Long, 1996, 2007, 2015). Second, given that recasts typically occur immediately following learners' non-target-like L2 output and retain the original meaning of the output, recasts allow learners to allocate more attentional resources to relevant L2 forms. In this way, recasts potentially afford linguistic input at the time when learners are likely to be most receptive. Third, immediate occurrences of recasts following learners' non-target-like L2 use also mean that the incorrect and correct use of L2 forms are juxtaposed. Hence, the provision recasts during task-based L2 interaction can make it easier for learners to cognitively compare their non-target-like use and target-like use of L2 forms by the interlocutor. Finally, recasts can help L2 learners, especially adults, counteract L1-tuned processing systems that are caused by entrenched L1 processing routines. Drawing on the literature of cognitive perspectives of L2

development (e.g., N. Ellis, 2005, 2008), Long (2015) argues that L1-tuned processing systems constitute a major factor that makes adult L2 learners unable to attend to and process certain linguistic features in the input appropriately. Long notes,

> [the L1-tuned processing systems] lead to adults adversely "filtering" L2 input to L1-established attractors … [Put another way,] the L1-tuned processing systems work against the adult [L2 learner], who proceeds to apply them to a new language whose parameters differ (Long, 2015, p. 41).

To illustrate, Indonesian adult learners of L2 English learning ECSs for the first time may assume that these grammatical structures are easy to learn. This is because the same structural pattern also exists in Indonesian: NP V **{more, *lebih (banyak)*} {Adj, N, Adv}** (PP). Thus, English sentences like (i) *they look* ***more beautiful*** *(than before)*, (ii) *you need* ***more money*** *(than me)*, and (iii) *she finished* ***more quickly*** *(than you)* can be translated directly into Indonesian as (i.a) *mereka terlihat* ***lebih cantik*** *(daripada sebelumnya)*, (ii.a) *kau membutuhkan* ***lebih banyak uang*** *(daripada aku)*, and (iii.a) *dia selesai* ***lebih cepat*** *(daripada kamu).* Given that this structural pattern constitutes a prototype for constructing comparative structures in Indonesian (Sneddon, Adelaar, Djenar & Ewing, 2012), Indonesian adult learners of L2 English with no or partial knowledge of these grammatical structures may fail to notice, or mentally process, the other structural pattern which is also commonly used to construct comparative structures in English: NP V **Adj + -er** (PP), as in (iv) *they look* ***bigger*** *(than before)*. Hence, theoretically at least, Indonesian learners of L2 English who have been recurrently exposed to this structural pattern (NP V **Adj + -er** (PP)) in natural communication may still fail to learn it. Instead, they may rely on using the former pattern (NP V **more {Adj, N, Adv}** (PP)) when constructing comparative structures in English, leading them to

produce a corresponding expression in (iv) as **they look **more big** (than before)*. According to Long (1996, 2007, 2015), a recast provided to these learners at the time when they produce such an ill-formed utterance might help them counteract their L1-tuned processing systems, making them realize that they cannot solely rely on their L1 grammatical systems when constructing particular structures in the L2. This processing condition, according to Long (2015), is unlikely to occur when the learners are only exposed to pre-modified input. He writes,

> Adult L2A [i.e., L2 acquisition] can be done incidentally and implicitly up to a point … from exposure to, and communicative use of, the L2, *but the results always fall short, usually far short, of native-like abilities*, and—of considerable importance for LT—take much more time than teachers and learners typically have at their disposal. (Long, 2015, p. 38, emphasis added).

Despite this, Long (2015) is careful to point out that his claim supporting the efficacy of recasts over pre-modified input is "still an open question" (p. 49). Thus, further research is needed to examine whether recasts are indeed more effective than pre-modified input in facilitating or driving forward L2 development (see also Long, 2019).

4.5.2.1 Previous Task-Based L2 Studies on the Efficacy of Recasts vs. Pre-modified Input

Among the first empirical evidence supporting the efficacy of recasts over pre-modified input was reported in Long, Inagaki and Ortega (1998). These researchers conducted two studies to examine how recasts and pre-modified input affected learners' abilities to use novel L2 forms accurately. The two studies were conducted separately, but they followed the same procedures. The first study focused on the use of adjective ordering and a locative construction in Japanese.

The second focused on the use of direct object topicalization and adverb placement in Spanish.

Twenty-four young adult learners of L2 Japanese and thirty young adult learners of L2 Spanish were recruited into the studies. These learners were randomly assigned into one of three conditions: pre-modified input, recast, or control. Learners in the pre-modified input group listened to a series of sentences provided through pre-recorded utterances. When listening to each sentence, these learners were asked to perform a certain action based on the meaning of each sentence that they heard (e.g., positioning or choosing certain objects). The learners were also asked to tell their interlocutor what they had heard from the recording so the interlocutor could perform the same action. Furthermore, the learners were not made aware of the presence of the target forms in each sentence. In this way, the instructional condition aimed to foster incidental learning of the forms.

Learners in the 'recast' group worked with an interlocutor to perform a series of the same actions. Unlike those in the pre-modified input group, learners in this recast group were required to produce their own utterances and not simply repeat a series of pre-recorded utterances. Prompts consisting of one or two-word clue was provided for these learners through recording. The prompts were used to elicit the target forms from the learners in every utterance that they produced. The interlocutor provided a recast when the learners failed to produce the target forms accurately. Nonetheless, the learners were not made aware that the recast that they received was intended to correct their inaccurate production of the target forms. In this way, this instructional condition also aimed to foster incidental learning of the forms.

Finally, learners in the control condition carried out activities that were not related to the tasks given in the post-test. Results of Long et al's (1998) studies indicated that both pre-modified input and recast groups outperformed the control group in their abilities to use the target forms accurately (except for the object topicalization in Spanish). Nonetheless, comparisons between the two experimental groups revealed that learners who received recasts outperformed those who only received pre-modified input. Based on these results, Long et al. (1998) concluded that although recasts and pre-modified input are effective to promote L2 development, recasts appear to be more effective than mere exposure to pre-modified input.

Subsequent studies on the efficacy of recasts and pre-modified input, however, provided only partial support to those reported on by Long, Inagaki & Ortega (1998). Studies by Ayoun (2001) and Lyddon (2011), for instance, indicated that recasts and pre-modified input were equally effective in promoting accurate use of specific L2 grammatical forms, and that no significant differences were noted between these two types of L2 instructional option. One possible explanation for these mixed findings may be that exposure to target L2 forms provided by way of pre-modified input or recasts in the studies was less comparable. That is, unlike pre-modified input which affords exposure to target L2 forms proactively (i.e., independent of learners L2 use), recasts can only occur when learners produce target L2 forms inaccurately (see Section 4.1 for details). Given this difference, it is possible that some learners in the studies did not attempt to produce target L2 forms, especially when they had no knowledge or had only partial knowledge of target L2 forms (Butler & Zeng, 2014; Shintani & R. Ellis, 2010). For this reason, some researchers argue that in investigating the efficacy of recasts and pre-modified input, it is crucial to employ recasts not only as a response to learners'

non-target-like use of target L2 forms, but also as a means to expose them to the forms more generally during task-based interaction, thereby making exposure to the forms during input comprehension and interaction comparable (Long, 2009; Mackey, 1999; Mackey & Oliver, 2002; Shintani & R. Ellis, 2010). In the present study, the strategy employed to afford exposure to the target L2 form more generally during task-based L2 interaction is realized through the use of 'confirmation checks' (see Excerpt 4.1.3 in Section 4.1 for an illustration).

4.5.3 Confirmation Checks as Interactional Feedback

In natural communication, confirmation checks are commonly used to elicit confirmation from a speaker that his or her utterance is correctly heard or understood. In the context of L2 learning, confirmation checks may be used to draw learners' attention to particular L2 forms *in context*. In this way, some confirmation checks share the same characteristics as recasts (Iwashita, 2003; Long, 2009; Mackey & Oliver, 2002; Nassaji, 2007; Shintani & R. Ellis, 2010).

By way of illustration, in the context where an instructional goal is to help learners of L2 English learn ECSs incidentally through performing relevant communication tasks, it is possible that these learners continually avoid using the structures while performing the tasks orally. For instance, instead of producing output like '*X is smaller than Y*', learners may continually use simple, early emerging structures like '*X is small*; *Y is big*'. To address such avoidance, the teacher can use the structures (ECSs) while seeking a confirmation from the learners (e.g., '*do you mean X is smaller than Y?*'). According to Gass (1997), this type of interactional feedback can serve as a *priming* device that potentially prompts the learners to produce the target form (ECSs) in their subsequent output (see also Branigan, Pickering &

Cleland, 2000; Pickering, Branigan, Cleland, & Stewart, 2000). For this reason, confirmation checks may prove useful for overcoming learner avoidance of particular L2 forms in ways that potentially lead to improved L2 resources (Long, 2009). A study by Mackey (1999) provides empirical support for the utility of confirmation checks in overcoming learner avoidance of particular L2 forms (see also Ferreira & Bock, 2006; Gass & Varonis, 1994; Iwashita, 2003; McDonough & Mackey, 2008; Polio & Gass, 1998; Strapp, Helmick, Tonkovich & Bleakney, 2011).

Mackey (1999) examined the effects of pre-modified input and interactional feedback involving 'recasts' and 'confirmation checks' on the use of English question formations by adult learners of L2 English. Basing her work on Processability Theory (e.g., Pienemann, 1984, 1989), Mackey (1999) argues that learners' abilities to produce specific question forms in English largely reflects their current L2 developmental stage. That is, according to Processability Theory, naturalistic learners of L2 English largely follow fixed developmental stages when producing question forms in the language, and it is hard for them to skip stages (Pienemann, 1984, 1987, 1989; Pienemann, Johnston, & Brindley, 1988). Because of this, Mackey contended that conditions that enabled learners to produce question forms at the stages *above* their current L2 developmental stage were effective in facilitating L2 development (*you have a cat?* → *do you have a cat?*; *where is the cat?* → *can you tell me where the cat is?*, etc.).

To examine the effects of pre-modified input and interactional feedback on L2 development, Mackey (1999, p. 567) categorized the target L2 form (English question formations) into six developmental stages, ranging from simple to complex. Thirty-four adult learners of L2 English from various first language

backgrounds were recruited into the study. These learners were allocated into five groups: interactor, interactor unreadies, observer, pre-modified input, and control. The main difference between the interactor and the interactor-unreadies groups was that the former consisted of learners who were assumed to be developmentally ready to acquire the target L2 form, while the latter were not (Mackey, 1999, p. 569).

With respect to instructional treatments, learners in the interactor and interactor-unreadies groups performed interaction tasks with a native-speaker interlocutor. The tasks required that these learners ask questions to the interlocutor to be able to complete given tasks successfully. When these learners produced questions that were grammatically less accurately, the interlocutor provided a recast, but when their questions were less comprehensible, the interlocutor sought confirmation from the learners.

Learners in the pre-modified input group performed the same tasks as those given to the interactor and interactor-unreadies groups. Nonetheless, rather than being required to ask questions to their native-speaker interlocutor, learners in this group only received a series of questions from their interlocutor. The questions were prepared in advance and designed in a way that would make communication breakdowns and negotiation for meaning highly unlikely. Hence, learners in this group received very limited opportunities to receive interactionally-modified input (Mackey, 1999, p. 570).

Learners in the observer group only observed the interaction taking place in the interactor group without participating in the interaction. In contrast, learners in the control group only performed tests with no treatment.

The treatments comprised three-day instructional sessions between the pre-test and immediate post-test. Learners also performed two delayed post-tests. The first was given a week after the immediate post-test; the second was given three weeks after the first delayed post-test.

Results of Mackey's (1999) study revealed that only three groups (the interactor, interactor unreadies and pre-modified input) performed significantly better than the control group in the post-test, whereas the observer group did not. Gains that the three groups made were also maintained throughout the delayed post-tests. Nonetheless, within-group comparisons indicated that only the interactor and interactor-unreadies groups significantly improved from the pre-test to the last delayed post-test.

Mackey (1999) argued that significant gains made by the two interactor groups were due to the nature of the treatments they received. In particular, interaction which affords opportunities for learners to receive interactional feedback on their erroneous and/or less comprehensible uses of the target form was argued to facilitate L2 development. Mackey concludes that this is because such feedback allowed learners to *notice mismatches* between their non-target-like and/or less comprehensible uses of L2 forms and target-like and/or more comprehensible uses of the forms modeled in the feedback, leading them to *notice gaps* in their L2 resources (see also Mackey & Oliver, 2002).

4.6 Noticing and Learning from Pre-Modified Input and Interactional Feedback

Supporters of Long's *updated* Interaction Hypothesis (hence, *the interaction approach*—Gass & Mackey, 2015) attribute the utility of interactional feedback in

facilitating L2 development to the concept of 'noticing'. This concept was first introduced by Schmidt in his Noticing Hypothesis (Schmidt, 1990, 1993). According to this hypothesis, learners' *attention to* and *awareness of* some aspects of the surface structure in the input or feedback constitute a necessary condition for L2 development to take place (cf., Schmidt, 2001).

Although advocates of the interaction approach generally agree with Schmidt (1990, 1993) that noticing plays an important role in L2 development, they disagree that it constitutes a necessary condition. Rather, they argue that L2 development can still occur subconsciously or implicitly during tasks (Doughty, 2001; Gass & Mackey, 2015; Long, 2015, 2016; Mackey, 2006, 2007; Mackey & Polio, 2009; Mackey, Abbuhl & Gass, 2012; see also N. Ellis, 2005, 2009a, 2009b, 2011). Furthermore, while Schmidt (1990, 1993) asserts that noticing necessarily involves attention *and* awareness, advocates of the interaction approach—following Tomlin and Villa (1994)—put a greater role of attention, rather than awareness, as a primary component of noticing.

According to Tomlin and Villa (1994), attention in L2 learning is governed by three main cognitive processes: *alertness* to incoming stimuli, *orientation* of attentional resources to the incoming stimuli, and *detection* of specific aspects of the stimuli. They further argue that none of these processes require awareness, but awareness may arise as a product of these processes. In this way, although the term 'noticing' is often used in the interaction literature to explain the underlying cognitive mechanisms that mediate interactional processes and L2 development, the term is often used in a slightly different way from that proposed by Schmidt (1990, 1993). In particular, the term 'noticing' is used to refer to the allocation of learners' attentional resources to certain aspects of linguistic stimuli that occur

during interaction with or without conscious awareness (Gass & Mackey, 2015; Gass & Torres, 2005; Long, 2015, 2016; Philp, 2003).

In a review of how learners' attentional resources relate to 'noticing', Robinson (1995) discusses the role of memory. According to Robinson, noticing involves "detection plus rehearsal in short-term memory, prior to encoding in long-term memory" (Robinson, 1995, p. 296). Furthermore, these processes of detecting, rehearsing and encoding target linguistic information in memory are regulated by the allocation of learners' attentional resources (ibid.).

Robinson (1995, 2003a) also relates 'noticing' to task-based L2 learning. He notes that certain characteristics of pedagogic tasks play an important role in directing learners' attentional resources to specific L2 forms during tasks, thereby inducing noticing of the forms. More specifically, he argues that tasks that are relatively high in cognitive or functional demands are likely to attract learners' attention to and noticing of specific L2 forms that are made available during tasks (e.g., by way of pre-modified input or interactional feedback). In contrast, tasks that are relatively low in their cognitive or functional demands may not be effective in doing so (Robinson & Gilabert, 2007; see also discussion of the Cognition Hypothesis in Chapter 2, Section 2.2.2).

By way of illustration, when learners of L2 English listen to a description about a picture of an item to choose from a set of pictures of dissimilar items (e.g., 'jacket', 't-shirt', and 'scarf'—see Figure 4.1, below), they might not face difficulty in selecting the correct picture provided that they know the L2 word for naming the target item (e.g., 'jacket'). In this way, the task may be considered relatively low in its conceptual demands in that this task does not require that the learners process

meaning associated with the use of L2 syntax in order to be able to complete the task successfully. Hence, in the case where the teacher attempts to use specific L2 syntactic structures (e.g., past participle form in English) in giving the instruction in hope that the learners attend to and subsequently notice the use of the structures *in context* (e.g., *have you seen my jacket?*), the learners may fail to do so. This is because the cognitive or functional demand of this task does not require the learners to allocate sufficient attentional resources to the structures to be able to complete the task successfully. Rather, the learners can complete the task by attending to the word 'jacket' alone.

Figure 4.1: A set of pictures of dissimilar items (pictures from Lambert, 2019)

In contrast, when learners listen to a description about a picture of an item to choose from a set of pictures of very similar items (as illustrated in Figure 4.2 below), it is unlikely that they will be able to complete the task successfully based on the name of the item only (i.e., 'jacket'). The task might thus be considered relatively high in its cognitive or functional demands, especially when being compared to the former task in Figure 4.1. That is, the task in Figure 4.2 requires learners to attend to the use of syntactic structures embedded in the input so they are able to complete the task successfully (e.g., *this jacket <u>looks slightly smaller than</u> the other two jackets, and the body part of this jacket <u>looks slightly slimmer than</u> the body part of the other two jackets*).

Figure 4.2: A set of pictures of similar items (pictures from Lambert, 2019)

The efficacy of pre-modified input in inducing attention to specific L2 forms is thus affected by the level of cognitive or functional demands that particular pedagogic tasks impose on learners (Robinson, 2001, 2003a, 2003b, 2005b, 2010, 2011b, 2015; Robinson & Gilabert, 2007). Given that the present study employed tasks that are relatively high in their cognitive or functional demands (see Appendix 2; see also Lambert, 2019, for empirical support), the tasks may encourage learners to attend to the use of task-relevant forms (ECSs) embedded in the input (see Appendix 3). More detailed discussion about these tasks is presented in Chapter 6, Section 6.4.

However, it is also possible that learners fail to comprehend the meaning associated with the use of the form while completing the task. As discussed in Chapter 3, Section 3.2.4, the internal structure underlying the construction of ECSs is quite complex both syntactically and semantically. Such complexity may impose processing difficulty on the part of learners, especially when they have only partial knowledge of these morphosyntactic structures (ECSs), as in the case of Indonesian learners whose L1 contains '*more*' but not '*-er*' (see Section 4.5.2). For this reason, allowing learners to negotiate for meaning (by means of requesting for clarification) may also be supportive in facilitating incidental learning of the structures (Erlam & R. Ellis, 2018; Long, 1983a, 1983b; Shintani, 2011, 2012, 2016; Shintani & R. Ellis, 2010; see also Chapter 3, Section 3.2.2, for more

elaborate discussion on this topic). In the present study, therefore, learners were allowed, but *not required*, to ask for clarification when they failed to comprehend the whole or part of the expression used in the pre-modified input materials. Despite this, all expressions in the pre-modified input were designed, following Mackey (1999), Long (2015, Ch. 9, 2019) and Yano, Long and Ross (1994), in a way that would potentially make communication breakdowns and negotiation for meaning unlikely (see Section 4.4, above).

Although learners may request for clarification from the interlocutor when they fail to comprehend the whole or part of the interlocutor's utterance, there is no need for them to produce the target L2 form (ECSs). Instead, they may only say *sorry?*, *pardon me*, *come again*, *what did you say?*, *I don't understand,* etc. For this reason, the provision of pre-modified input in the present study may only facilitate the development of receptive skills, rather than productive skills, of the form (ECSs) (DeKeyser, 2007, 2015, 2017; Izumi, 2003; Skehan, 1996, 1998a; Skehan & Foster, 2001; Swain, 1985 and elsewhere). In contrast, the provision of interactional feedback during task-based L2 interaction has potential benefits not only in fostering facilitating incidental learning of the form (ECSs) but also in fostering increased frequency, accuracy, and fluency in learners' use of this form during spontaneous production (see Section 4.5.2 for details).

To sum up, although pre-modified input and interactional feedback (as operationalized in the current study) may both be effective in inducing attention to and noticing of the target L2 form (ECSs), learners who receive interactional feedback may be able to use the form more frequently, accurately and fluently in relevant L2 communication contexts than those who only receive pre-modified input.

4.7 Summary

This chapter has provided an operational definition of the terms ‘pre-modified input’ and ‘interactional feedback’ (independent variables) as employed in the current study. It also has reviewed relevant theoretical underpinnings of these two types of L2 instructional option in L2 development. As noted in Chapter 3, the abstract notion of L2 development is operationally defined in this study as the gradual improvements in the use of particular L2 forms by learners over time, especially as learners *produce* these forms under real-operating conditions. The next chapter will further discuss various measurements that have been commonly used to assess learners’ use of particular L2 forms in spontaneous production as a means to providing evidence of L2 development, and outline specific ones that are employed in the current study.

Chapter 5

Uptake, Automaticity and Evidence of L2 Development

This chapter argues for specific measures of uptake and automaticity as measures of L2 development. It begins by outlining a set of processes that underpin L1 and L2 speech production. Specific differences in L1 and L2 speech production are also defined in terms of the type of processing that underpin them: parallel vs. serial, respectively. The relationship between the type of processing and L1/L2 knowledge is then explicated, leading to an argument that L2 speakers' abilities in formulating and executing speech in real time are largely determined by their developing L2 knowledge. Following this, the notions of 'uptake' and 'automaticity' are reviewed from Chapters 3 and 4, and key L2 studies that have employed or examined specific measures of uptake and automaticity are critiqued. The chapter then explicates how uptake and automaticity are measured in the current study.

5.1 Cognitive Underpinnings of L1 and L2 Speech Production

As will be discussed in Sections 5.5 and 5.6, below, uptake and automaticity of a specific L2 form (ECSs) were measured in this study within the context of *naturally-occurring task-based L2 production*. Given this specific context of measurement, this section starts with elucidating cognitive underpinnings of L1 and L2 speech production with an aim to establishing the relationship between measures of uptake and automaticity employed in the current study and L2 development, or how the proposed measures of uptake and automaticity might truly

reflect the status of and access to learners' developing knowledge of a target L2 form (ECSs).

It has long been acknowledged that L1 and L2 speech production is governed by complex cognitive processes which include recalling relevant information (linguistic and conceptual) from long-term memory and processing the information in working memory to meet functional demands of certain communication contexts. Levelt (1989, 1999) proposes three modules to explain how such complex cognitive processes operate during L1 speech production. These include *conceptualization* (i.e., a process of creating a preverbal message that speakers will express verbally), *formulation* (i.e., a process of converting the preverbal message into an internal speech plan), and *articulation* (i.e., a process of realizing the internal speech plan into overt speech).

According to Levelt (1989, 1999), these three speech processing modules operate *in parallel* in L1 speech production. That is, all relevant information generated in each module is processed simultaneously, provided that the information processing that takes place in each module is sufficiently automatic to allow parallel processing. In the case of many L2 speakers, however, it has been proposed that the three modules operate *serially*. That is, the output from each module forms the input for the next module and only one module operates at a time (de Bot, 1992; Kormos, 2006). Such serial processing normally occurs until L2 speakers develop a level of automaticity in the *formulation* stage to allow them to parallel process (Segalowitz, 2010, Ch. 4; Lambert, Kormos & Minn, 2017). These two types of processing underlying L1 and L2 speech production are illustrated below.

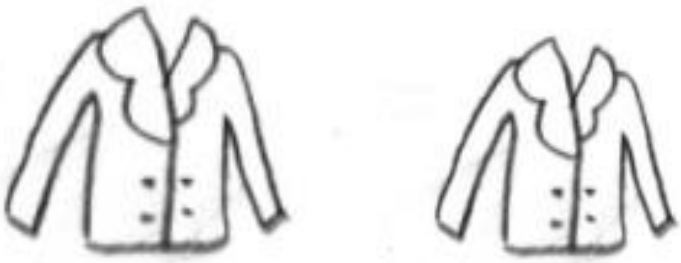


Figure 5.1: Jackets with identical features except for their sizes (pictures adopted from Lambert, 2019)

In the case of *parallel processing*, when L1 speakers are asked to disambiguate two items with very similar features, like those in Figure 5.1, all relevant conceptual information associated with the items that the speakers have as part of their world knowledge is recalled simultaneously (e.g., *these items are jackets; jackets are a piece of clothing worn by men and women to cover the upper part of their body; jackets are usually worn when the weather is cold; jackets typically have long sleeves, pockets, lapels, lining, buttons or a zip; jackets may be made of different materials and may be of different sizes, colors, prices,* etc.). This recalling process normally takes place implicitly (i.e., without the speakers' conscious control). Furthermore, all the recalled information of the target concept (i.e., '*jackets*') will usually remain active until the speakers opt to focus on specific information that matches the goal of communication (Barsalou, 2003, 2008; Lakoff, 2004; Langacker, 1987b, 2008; Slobin, 1996; Talmy, 2008). For example, in the case where the speakers choose to focus on the size of the jackets because this feature is what clearly distinguishes the two jackets (see Figure 5.1), this feature will be emphasized and other information, which is deemed irrelevant with the goal of communication, will be dropped (Slobin, 1996; Talmy, 2000, 2008). According to Levelt (1989, 1999), such a process of selecting particular information to be

communicated verbally leads to a creation of a ‘preverbal message’ during the *conceptualization* stage of speech production.

It is important to note that preverbal messages are conceptual in nature. That is, preverbal messages only contain conceptual contents that have not been put into words (Allwood, 2003; Lakoff, 1987, 2012; Langacker, 1987a, 1990, 2008; Slobin, 1996; Talmy, 2000). With respect to the two jackets in Figure 5.1, for instance, the preverbal message that the speakers create may be equivalent to (1) ‘*two jackets of different sizes*’. However, as soon as such a preverbal message is created, *formulation* takes place. That is, the preverbal message is converted into an internal speech plan (Levelt, 1999, p. 8).

Levelt (1989, 1999) argues that there are two processes by which an internal speech plan is formulated in speakers’ mind: lexical selection or proposition formulation (see also Segalowitz, 2010, Ch. 1). In the case of lexical selection, specific words that can be used to express preverbal messages are retrieved from speakers’ linguistic resources. For instance, in the case where speakers of L1 English intend to express a preverbal message like (2) ‘*a domesticated carnivorous animal that has a long snout, an acute sense of smell, and a barking, howling or whining voice*’, the word ‘*dog*’ is likely to be retrieved automatically from their long-term memory (Jurafsky, 1996; Levelt, 2001). On the other hand, in the case where a preverbal message cannot be expressed in specific words, a proposition will be formulated to provide an idea about it. For instance, given that English language does not have specific words to refer to the preverbal message in (1): ‘*two jackets of different sizes*’, English speakers will need to formulate a proposition to express it in words (a phrase or clause).

There are, of course, multiple ways in which a relevant proposition can be formulated (Segalowitz, 2010, p. 11). Levelt (1999), however, argues that the way propositions are formulated is largely determined by the *perspective* that speakers take. For example, when attempting to describe the two jackets in Figure 5.1, a proposition that proficient speakers of English formulate may be equivalent to (1.a) or (1.b).

(1.a) *The jacket on the **left** side looks quite **large** in comparison to that on the **right** side*.

(1.b) *The jacket on the **right** side looks quite **small** in comparison to that on the **left** side*.

That is, in formulating the preverbal message in (1), these speakers may opt to "highlight" specific features of the items that are located on the left *or* right side of the picture (see Figure 5.1). Hence, although the speakers may have similar ideas about the items, or what they can communicate about the items like those in Figure 5.1 (see also Appendix 2), they can differ in organizing their mental images of the items. Such differences are, again, determined by the perspectives that the speakers take (see also Langacker, 1987b, 1990, 2008; Talmy, 2000, 2008).

The process of generating a relevant proposition—as exemplified in (1.a) or (1.b)—provides a cognitive basis for linguistic encoding during the *formulation* stage. According to Levelt (1989, 1999), the process of encoding propositions linguistically is largely determined by speakers' mental lexicon. This mental lexicon contains a range of information underlying language use, such as *semantic information* (e.g., what words need to be used to express specific propositions), *pragmatic information* (e.g., whether those words are appropriate to be used in a

given communication context), *syntactic information* (e.g., how those words may be put together into meaningful, grammatical sentences), and *phological information* (e.g., how those words may be pronounced in overt speech) (Levelt, 1999, pp. 87-88; see also Segalowitz, 2010; Skehan, 2009a, 2009b). In this way, depending on speakers' mental lexicon, the proposition like (1.a) may be encoded linguistically by speakers of L1 English into (1.a.i), (1.a.ii), (1.a.iii), or something similar.

(1.a.i) *The jacket on the left side {is, looks} {large, big}, while the one on the right side {looks, is} small.*

(1.a.ii) *The jacket on the left side {is, looks} (rather) {large, big} in comparison to that on the right side.*

(1.a.iii) *The jacket on the left side {looks, is} (slightly) {larger, bigger} than that on the right side.*

To this end, the cognitive processing underlying linguistic encoding typically requires deeper computational processing than that of lexical selection (Clark, 1999; de Bot, 1992; Kormos, 1999, 2000, 2006; Paradis, 2004, 2009; Segalowitz, 2010, 2016).

The process of creating an internal speech plan (by means of lexical selection or linguistic encoding, as described above) is simultaneously accompanied with the *articulation* process in L1 speech production (Levelt, 1989, 1999). That is, as L1 speakers create an internal speech plan mentally, the resulting outcome is also verbally expressed by means of the phonetic encoding taking place at the motor level (i.e., involving the speech organs).

Finally, the whole working mechanisms of the three processing modules (*conceptualization*, *formulation* and *articulation*) are simultaneously monitored (*parsed*) to see whether the articulation process matches the internal speech plan and the preverbal message.

5.2 The Relationship between Linguistic Resources, Speech Processing and Speech Behavior

The assumption underlying parallel processing in L1 speech production proposed by Levelt (1989, 1999) is associated with the fact that linguistic resources that L1 speakers possess are typically very large. Furthermore, these L1 resources are usually stored *implicitly* in the speakers' long-term memory, making parallel processing possible. In contrast, linguistic resources that many L2 speakers possess are usually more limited. Furthermore, their L2 resources are usually stored *explicitly* in their long-term memory, making parallel processing difficult (de Jong, Groenhout, Schoonen & Hulstijn, 2015; N. Ellis, 2002, 2005; R. Ellis, 2005; Krashen, 1981, 1982; Paradis, 2009; Rebuschat, 2013, 2015; Segalowitz, 2010). Given these differences, learners' L2 speech production is usually characterized by *serial processing*. That is, unlike parallel processing underlying L1 speech production (and fluent L2 speech production), where the successful creation of preverbal message during *conceptualization* is simultaneously accompanied with *formulation*, *articulation* and *parsing*, leading to rapid and smooth speech production, serial processing is typically characterized by the need of L2 speakers to allocate their attentional resources during *conceptualization*, *formulation*, *articulation* and *parsing* (de Bot, 1992; Kormos, 2006, 2011; Segalowitz, 2010, 2016; Skehan, 2009a, 2009b). This stepped approach normally leads to varying

degrees of breakdowns and repairs during real-time L2 production (de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Lambert, Kormos & Minn, 2017).

With respect to speech behavior, therefore, while L1 speakers normally pause *between clauses* for the purpose of *conceptualization*, L2 speakers might also need to pause *within clauses* as a result of breakdown in *formulation* (Davies, 2003; Kormos, 1999, 2000; Lambert, Aubrey & Leeming, 2020; Lambert, Kormos & Minn, 2017; Skehan, 2009a; Skehan & Foster, 2008). Furthermore, L2 speakers might also need to *re*formulate their utterances before or after they are completely produced due to breakdown in phonetic encoding during *articulation* (de Jong, 2016; Kormos, 2006; Segalowitz, 2010, 2016).

To this end, speech behavior can indicate the level of automaticity of cognitive processes that govern L1 and L2 speech production (de Bot, 1992; Howell & Au-Yeung, 2002; Kormos, 1999, 2006; Levelt, 1989, 1999; Pütz & Sicola, 2010; Segalowitz, 2010, 2016; Tavakoli & Skehan, 2005). With regard to L2 speech production, decreases in the frequency of mid-clause pausing and reformulations produced by L2 speakers accompanied with increased speech rate might well indicate increased automaticity in L2 speech processing (see Section 5.6.2 for details).

5.3 Automatic L2 Speech Processing and L2 Development

It is worth noting, however, that increased automaticity in L2 speech processing (as described in Section 5.2, above) does not necessarily correlate with L2 development. That is, when a group of L2 learners are asked to perform certain oral communication tasks multiple times, they may be able to use their L2 with a faster speech rate along with lower frequency of within-clause pauses and/or within-

clause reformulations (as most research on task repetition indicates—e.g., Ahmadian, 2011; Lambert, Kormos & Minn, 2017; Wang, 2014). Nonetheless, it is possible that such increases in automaticity may be the result of *avoiding* specific L2 forms that learners initially used with partial control and *maximizing* forms that they are able to use with greater control.

For instance, when learners have only partial mastery in using English comparative structures (ECSs), they may opt to use simpler structures to convey the same intended meaning in subsequent task performance. That is, rather than producing a sentence like '*X is bigger than Y*', learners may continually produce sentences like '*X is big*. *Y is small*' (see also Givón, 1985, 1989; Robinson, 2001, for similar arguments). In this case, task repetition does not afford conditions which allow greater control (i.e., more accuracy and/or more fluency) over L2 forms which learners initially used with partial control, but rather affords conditions for developing strategies of using early emerging L2 knowledge to cope with the task demands. In order for increased automaticity to constitute L2 development, therefore, it is necessary to examine whether learners develop greater control in using L2 forms over time (see also de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Norris & Ortega, 2003, 2013; Sanz and Grey, 2015; Suzuki, in press).

Notwithstanding, it is also easonable to expect that task repetition facilitates L2 development, since, as noted in Chapter 3, every time learners engage in contextualized L2 use their L2 knowledge undergoes change (Beckner et al., 2009; de Bot, Lowie & Verspoor, 2007; N. Ellis & Larsen-Freeman, 2006; Langacker, 2008; MacWhinney, 2015). One process underlying the change is the consolidation of partially-learned knowledge of L2 forms (Bygate, 1999, 2001; N. Ellis, 2008; Read, 2004), resulting in faster or more automatic processing as learners attempt to

use these forms in subsequent task performance (Haastrup & Henriksen 2000; McDonough & Trofimovich, 2015; Pickering & Ferreira, 2008; Suzuki, in press).

For example, when learners of L2 English are asked to identify one of the jackets in Figure 5.1 with the same procedure as that performed by native speakers of English (see Section 5.1), these learners may create the same preverbal message as in (1): '*two jackets of different sizes*'. Unlike the native speakers, however, who are usually able to encode this preverbal message linguistically and realize it in overt speech at the same time (as a result of parallel processing), L2 learners might not be able to do so. Rather, the learners might still need some time to encode this preverbal message linguistically and to realize it in overt speech. In the case where these learners already learned the use of ECSs and attempt to use this morphosyntactic structure to encode the message linguistically, they may eventually produce a sentence like '*the jacket on the left side looks slightly bigger than that on the right side*'. During the encoding process, however, the learners may rely on using rules or schematic knowledge of the structure, and this typically leads to relatively slow, effortful production of the sentence (DeKeyser, 2015, 2017; Segalowitz, 2003). Hence, although the learners may be able to produce the sentence accurately, their production will be less fluent due to the result of serial or non-automatic processing (see also Suzuki, 2017a; Suzuki & DeKeyser, 2017a).

Over time, when the learners have a chance to perform the task and/or other similar tasks multiple times and produce numerous samples of ECSs as a result, they may become *less* reliant on using rules or their schematic knowledge to generate sentences containing the use of ECSs. In other words, having produced sentences containing the use of ECSs frequently as to coping with the same or similar functional demands of certain L2 communication tasks, the learners may then be

able to retrieve familiar expressions spontaneously with little attentional control from their long-term memory (Bygate, 1999, 2001; Robinson, 2001; Skehan, 1998a). In this way, the processing of ECSs becomes *more automatic*. In usage-based literature, this phenomenon is referred to as *chunking*: It is the process where certain parts of exemplars are abstracted and stored as semi-constructed units in the learner's memory, making them readily accessible during spontaneous performance (Bybee, 2008, 2010, N. Ellis, 2001, 2002, 2005, 2009a; Pawley & Syder, 1983; Tomasello, 2000, 2003; see Chapter 3 for details).

5.4 Learners' Use of Specific L2 Forms during Spontaneous L2 Production as an Indicator of L2 Development

Based on the discussion above, the current study stipulates, following Levelt (1989, 1999) and usage-based accounts of L2 learning (see Chapter 3), that learners' abilities in using L2 forms in spontaneous L2 production is largely determined by their (developing) knowledge of the forms. Furthermore, as discussed in previous chapters, pre-modified input and interactional feedback can help learners develop knowledge of a specific task-relevant L2 form (i.e., ECSs) in ways that can lead them to be able to use the form (more) accurately in naturally-occurring communication. Nonetheless, given that the provision of pre-modified input or interactional feedback as employed in the current study did not require learners to produce the form (ECSs) immediately following the provision of the input or feedback (see Chapter 4, Section 4.1, for illustrations), knowledge of this form that learners develop as a result may be stored only receptively. That is, the learners may not be able to use this form accurately and/or fluently during spontaneous L2 production (N. Ellis, 2019; Goldberg & Casenhiser, 2008; Skehan & Foster, 2001;

Swain, 2005). However, in the case where the learners are able to do so, it can be argued that the provision of pre-modified input or interactional feedback helps the learners consolidate their knowledge of the form (ECSs) in ways that can help them use it more accurately and/or fluently at productive level (for similar arguments, see, N. Ellis, 2005, 2009a, 2019; Mackey, 1999; McDonough & Mackey, 2008; McDonough & Trofimovich, 2015; Pickering & Ferreira, 2008). In the current study, such indicators of L2 development (i.e., *increased use*, *increased accuracy* and *increased fluency* of learners' production of L2 forms) are conceptualized in terms of 'uptake' and 'automaticity' of ECSs.

5.5 Uptake of a Specific Task-Relevant Form as an Indicator of L2 Development

Various definitions of uptake exist in the L2 literature (e.g., Egi, 2010; R. Ellis, Basturkmen & Loewen, 2001; Heift, 2004; Lyster & Ranta, 1997; Panova & Lyster, 2002; Robinson, 2007; Sheen, 2004). Most of the definitions, however, do not necessarily relate to L2 development (see Chapter 4, Section 4.5.1, for details). In the current study, the term uptake refers to learners' incorporation of a specific task-relevant L2 form (ECSs) from pre-modified input or interactional feedback into their own L2 production. Furthermore, in order for such uptake to constitute L2 development, the study examines learners' production of the form in naturally-occurring task-based L2 communication *before* and *after* they received pre-modified input or interactional feedback which afforded exposure to the form in contextualized ways (Gass, 1997; Mackey, 1999; Mackey & Oliver, 2002; Révész, 2009). In this way, the term 'uptake' used in the current study largely corresponds

to the term ‘acquisition’ used in other L2 studies (see Norris & Ortega, 2003, 2013; Skehan & Foster, 2001).

Even so, employing ‘uptake’ of a specific task-relevant L2 form as a measure L2 development entails more than just ‘acquisition’, particularly with respect to how it has been commonly measured in instructed L2 development studies. For instance, while it is possible to measure acquisition through the use of recognition and/or controlled production tests (e.g., R. Ellis, Loewen & Erlam, 2006; Loschky, 1994; Révész & Han, 2006; Révész, Sachs & Hama, 2014), uptake can only be measured through the use of performance tests which require free L2 production on the part of learners (e.g., Mackey & Oliver, 2002; McDonough & Kim, 2009). Furthermore, while measures of acquisition normally involve (only) determining the ratio of learners’ accurate use of target L2 forms during recognition and/or production (e.g., Révész, 2009; Shintani & R. Ellis, 2010), measures of uptake also require consideration of the amount of which the forms that are *voluntarily* attempted by learners during contextualized communication (e.g., Gries & N. Ellis, 2015; Robinson, 2007; Year & Gordon, 2009). As with measuring acquisition, however, uptake needs to be measured under relatively natural conditions where all aspects of L2 production are fully under the control of the learner so it can clearly reflect L2 development (Schmid, Verspoor & MacWhinney, 2011; Skehan & Foster, 2001).

5.5.1 Specific Measures of Uptake Employed in the Current Study

Procedures for measuring uptake, or learner incorporation of target L2 forms into their own L2 production, as described above, have been used in a number instructed L2 development studies. Mackey (1999), for instance, in her study on the

effects of pre-modified input and interactional feedback on L2 development operationalized L2 development as learners' ability to *produce* question forms in English in a given communication context. Adopting the developmental framework proposed by Pienemann (1984, 1989; Pienemann & Johnston, 1987), which describes different question forms that English language learners at different developmental levels are likely to produce, Mackey (1999) contended that the production of certain question forms by learners of L2 English reflected their current L2 developmental level. Furthermore, she noted that instructional treatments that helped learners produce question forms at the stages *above* their current developmental level were effective in promoting L2 development (see Chapter 4, Section 4.5.3, for a more detailed review of the study).

With regard to measurement, Mackey (1999) maintained that, following Pienemann and colleagues, *two* exemplars of question forms that learners of L2 English produced while completing given tasks orally in the post-tests was sufficient to provide evidence of their L2 development (p. 567). In other words, learners who were able to produce higher levels of question forms in the L2 in at least two different linguistic or communicative contexts were argued to have successfully acquired the form (ibid.).

To date, this approach for measuring L2 development (as per Mackey,1999) has been used by a number of L2 researchers as a means for examining the effectiveness of task-based L2 interaction (e.g., Kim, Payant & Pearson, 2015; Mackey & Oliver, 2002; McDonough, 2005; McDonough & Mackey, 2006, 2008; Révész, 2009, 2012). It is worth noting, however, that this approach to measuring L2 development predicates that L2 development constitutes a matter of acquiring *or* not acquiring L2 forms, rather than as a dynamic process of developing L2

resources (Larsen-Freeman, 2015). Given that the current study adopts a different view of L2 development—in that the notion of L2 development is conceptualized as *a gradual process* of mapping linguistic forms onto meaning and function (see Chapter 3, Section 3.2.3, for details), the approach to measuring L2 development proposed by Mackey (1999) might be irrelevant to use. Moreover, the selection of *two* exemplars containing targeted linguistic forms used by Mackey and colleagues as a basis for determining acquisition also appears arbitrary. That is, there is relatively little empirical evidence that supports the claim that two exemplars of learners' production of target L2 forms could indeed be used as a cut-off for determining L2 acquisition (see Spinner & Jung, 2018).

An alternative approach to measuring L2 development which resembles that employed by Mackey (1999), but is more relevant to the current study, was proposed by McDonough and Kim (2009). Informed partly by usage-based accounts of L2 development, McDonough and Kim (2009) conducted a study to examine whether variation in the type/token frequency (see Chapter 3, Section 3.2.3, for details) of particular verbs used during task-based L2 communication affected learners' subsequent production (uptake) of a target L2 form (i.e., *wh*-question formations in English with inversion, such as *what books did you buy?*, *how many cups of coffee do you drink every day?*, etc.). Eighty-five adult Thai learners of L2 English were assigned to either a *low-type/high-token* or *high-type/low-token* frequency condition. To measure the effects of these conditions, McDonough and Kim (2009) compared the proportion of learners' accurate production of the target form as they completed tasks in the pre-test and post-test. The proportion was obtained by dividing the total number of the form that learners produced correctly by the total number of obligatory contexts (p. 391). Results of

McDonough and Kim's (2009) study indicated that high-type/low token input frequency did not facilitate accurate production of the target L2 form, and thus failed to provide empirical support to the claim that high-type/low token frequency facilitates L2 development (cf., Chapter 3).

Nonetheless, McDonough and Kim (2009) were careful to point out that the findings of their study might be due to the limited treatment period (i.e., two sessions lasting for approximately 10 minutes each). Furthermore, they maintained that "Including other measures in the study, such as reaction times, might have revealed possible benefits for input type frequency, such as facilitating the ability to process *wh*-questions more quickly and/or accurately" (pp. 394-395). Put another way, McDonough and Kim (2009) suggest that employing specific measures of accuracy alone might be *in*sufficient to gauge effects of certain conditions on L2 development. Other measures are also needed to ascertain whether particular conditions affect L2 development (see also Norris and Ortega (2003) and Sanz and Grey (2015) for similar arguments).

In addition to reaction times (see Section 5.6.1.1, below), learner production (uptake) of a target L2 form may also be measured in terms of *the number of times* the form is attempted during meaningful, contextualized communication (Crossley, Salsbury, McNamara & Jarvis, 2011; Horst & Collins, 2006; Robinson, 2007). This measure is based on the idea that, although certain instructional treatments may not directly assist learners to use the target L2 form accurately in spontaneous L2 production (especially when the form is linguistically complex as in the case of *wh*-question formations in English), the treatments may still successfully stimulate increases in learners' L2 resources (N. Ellis, 2019, p. 49). For instance, in the case where the treatments involve the use of tasks, learners may incidentally notice that

specific L2 forms are recurrently used by their interlocutor to convey particular meanings and functions. Such noticing may then inspire the learners to use the forms, albeit inaccurately, as they attempt to cope with the same or related functional demands of certain communicative tasks. Such recurrent use coupled with opportunities to receive input or interactional feedback as learners use the forms inaccurately may eventually help the learners develop accuracy in using the forms (for similar arguments see de Bot, Lowie & Verspoor, 2007; N. Ellis, 2019; Gries & N. Ellis, 2015; Larsen-Freeman, 2013a; Robinson & N. Ellis, 2008; Skehan, 2014, Ch. 1; Slobin, 1997; Swain, 1995; Verspoor, Lowie & van Dijk, 2008).

To date, task-based L2 development studies that examine differences in *the number of times* specific L2 forms are attempted or produced by learners before and after they engage in certain task-based instructional conditions are relatively few. Most have been concerned with the proportion of learners' *accurate* use of specific L2 forms (e.g., Long, Inagaki & Ortega, 1998; Shintani & R. Ellis, 2010). Nonetheless, as outlined above, measuring the number of times specific L2 forms are produced by learners during tasks can provide additional evidence as to whether certain instructional conditions are effective in facilitating L2 development and use. For this reason, specific measures of uptake are operationalized in the current study in the following ways: (1) the number of times a task-relevant L2 form (i.e., ECSs) was attempted by learners as they completed given tasks orally in the pre-test and post-tests, and (2) the proportion of accurate attempts in using this form. Chapter 6, Section 6.9, outlines the procedures used for measuring uptake of the form in the current study.

5.6 Automaticity of Specific L2 Forms as an Indicator of L2 Development

While uptake operationalized in the current study is primarily concerned with learners' abilities in mapping a task-relevant L2 form (ECSs) onto meaning and function in ways that lead to *increased use* and/or *increased accuracy* of this form in spontaneous L2 oral production, automaticity is concerned with their abilities in using the form *rapidly without undue pauses, repetitions and repairs*. In cognitive terms, automaticity refers to mental processes that allow learners to recall, access and/or process certain information (linguistic and/or conceptual) rapidly with minimum attention and effort, thereby allowing their performance to be fast, stable and effortless (Moors & De Houwer, 2006; Segalowitz & Segalowitz, 1993).

Traditionally, automaticity was viewed as a dichotomous phenomenon—in that information processing that takes place internally (i.e., in the mind) is viewed as either automatic or non-automatic (controlled). Subsequently, researchers generally agree that automaticity constitutes a continuum of gradual improvements in the efficiency of processing (Anderson, 1992; Logan, 1985), and that such improvements largely follow the power law of practice (DeKeyser, 2001, 2015, 2017; N. Ellis, 2001, 2005; Segalowitz, 2003, 2010, 2016). In the literature, gradual improvements that lead to automaticity are usually referred to as 'automatization'.

It is worth noting that both automaticity and automatization constitute mental phenomena. Thus, they cannot be observed directly in research practices. Hence, studies on L2 automaticity and/or L2 automatization normally require researchers to make inferences from observing specific L2 performances (de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Lambert, Kormos & Minn, 2017; Segalowitz, 2010, 2016; Skehan, 2009a, 2014). Nonetheless, any inferences drawn on from

observing relevant L2 performances should be guided by relevant theoretical accounts (R. Ellis, 1990; Norris & Ortega, 2003, 2013; Skehan, 2009a, 2014).

Furthermore, studies on L2 automaticity and/or L2 automatization need to be conceptualized, or defined, in relation to the tasks used. This is because different tasks normally impose different levels of processing demands on the part of learners (e.g., word/sentence-level comprehension vs. discourse-level production) (Hui & Godfroid, 2020; Lim & Godfroid, 2015; Robinson, 2003b, 2005b, 2010, 2011a; Segalowitz, 2016; Segalowitz & Segalowitz, 1993). With this consideration in mind, the term automaticity is thus used in the present study to refer to cognitive processes that underlie learners' ability in producing a task-relevant L2 form (ECSs) rapidly without undue pauses, repetitions and repairs in naturally-occurring task-based L2 communication. Following Segalowitz (2010, 2016), the current study emphasizes two key processing features to characterize automaticity, namely, *fast* and *stable* L2 processing.

5.6.1 Methods for Measuring L2 Processing Speed

By far the most prevalent methods for measuring L2 processing speed (i.e., fast vs. slow L2 processing) include 'reaction times' and/or 'speed fluency'. Each of these two methods is discussed in detail in four subsections below.

5.6.1.1 Reaction Times

In the context of L2 research, reaction times (RTs) are commonly used to observe *the amount of time* (typically operationalized in milliseconds) that individual learners need to respond to particular stimuli in the L2. The use of RTs in L2 research is predicated upon the idea that L2 processing speed may be observed

through examining response latencies (i.e., the time that elapses between the presentation of particular stimuli and the response given by learners). Specifically, shorter response latencies (as indicated by smaller RTs) are usually regarded as an index of faster L2 processing. In contrast, longer response latencies (as indicated by larger RTs) are usually considered as an index of slower L2 processing (DeKeyser, 1997, 2001; Robinson, 1996, 1997; Robinson & Ha, 1993; Suzuki, 2017a; Suzuki & DeKeyser, 2017a). To this end, measures of L2 processing speed, such as RTs, normally require the researcher to make explicit judgements about relatively fast and slow processing. This is because any given feature underlying one type of processing (e.g., fast L2 processing) can only be understood in relation to a corresponding contrasting feature underlying the other type of processing (e.g., slow L2 processing) (Segalowitz, 2003, p. 384).

When making judgements about relatively fast and slow L2 processing, researchers usually consider RTs below 350 or 500 milliseconds as extremely fast. In contrast, RTs that are higher than three standard deviation above the mean as extremely slow. Furthermore, researchers usually discard these two types of extreme RTs from the analysis. This is because these extreme RTs do not necessarily reflect L2 processing speed (Ewoldsen, 2017; Marinis, 2010). That is, RTs that are extremely fast are usually due to anticipatory responses (i.e., responses that occur prior to the presentation of target stimuli), whereas those that are extremely slow are typically due to interference that occurs during trial (de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Hulstijn, van Gelderen & Schoonen, 2009; Suzuki, 2017a; Suzuki & DeKeyser, 2017a; cf., Lim & Godfroid, 2015).

5.6.1.2 The Use of RTs in L2 Research

RTs have been commonly used in L2 research to examine the relationship between learners' L2 proficiency profiles and their L2 processing speed. In general, research in this area suggests that learners' L2 proficiency profiles are positively correlated with L2 processing speed: the more proficient learners are in the L2, the faster they are in processing specific L2 forms in real time (Bird, 2012; Jiang & Nekrasova, 2007; Schmitt, 2004; Shantz, 2017; Yamashita & Jiang, 2010; Wolter & Gyllstad, 2013). A line of L2 research has also indicated the utility of RTs in examining the type of L2 knowledge that learners draw on under certain performance conditions. A study by Suzuki and DeKeyser (2015), for instance, suggested that RTs are indicative of whether learners drew on explicit or implicit L2 knowledge when completing certain tasks or tests (see also Godfroid, 2016; Jiang, 2007).

It is worth noting, however, that RTs have been commonly used in L2 studies within the context of L2 *recognition*, rather than L2 production. Given that the current study is concerned with learners' *production* of a specific task-relevant L2 form in task-based L2 communication, the applicability of such measures to examine L2 processing speed underlying such production remains an open question (Segalowitz, 2016).

Recent exceptions to this approach are those studies on instructed L2 development conducted by Suzuki and associates (e.g., Suzuki, 2017a; Suzuki & DeKeyser, 2017a, 2017b; Suzuki & Sunada, 2019; see also Sato & McDonough, 2019). In a study by Suzuki (2017a), for instance, RTs were used to gauge the effects of "massed" or "distributed" practice on L2 processing speed as reflected in learners' real-time production of specific (newly learned) L2 forms. The notions of massed

and distributed practice were defined by Suzuki (2017a) as relatively short (3.3-day) and long (7-day) learning intervals, whereas gains in L2 processing speed were defined as shorter response latencies manifested in learners' real-time production of a target L2 form in the post-tests, relative to those in the pre-test.

Sixty Japanese university students were recruited for the study. These learners were explicitly taught the use of a morphological form (*–ando*) in a semi-artificial language called Supurango under the massed or distributed practice condition. Suzuki (2017a) noted that the use of this morphological form indicated the present progressive in the language, but that its use was largely governed by multilevel rules (e.g., *gav**ar*** + *–ando* → *gav**iando***, *ton**er*** + *–ando* → *ton**iando***, *pont**as*** + *–ando* → *p**a**nt**iando***, *beten**es*** + *–ando* → *b**i**ten**iendo***).

After receiving explicit instruction on the target form, learners in each condition engaged in production practice to help them use the form accurately in spontaneous L2 oral production. In this study, all the treatments and the tests were delivered through personal computers using a specialized software called DMDX.

Suzuki (2017a) employed RTs to observe changes in the processing speed of the target L2 form in real-time production. Oral description tasks consisting of pictures of a man performing various activities were used to elicit the target form from the learners. Suzuki (2017a) noted that these tests were used to assess the extent to which the learners were able to use their explicit or declarative knowledge of the target form rapidly without repairs or reformulations (p. 523). For this reason, only learners' accurate production of the target form that did not contain repairs or reformulations was submitted to RTs (p. 528).

Following de Jong, Steinel, Florijn, Schoonen and Hulstijn (2013), Suzuki (2017a) measured RTs from the onset of the prompt (target picture) to the end of the utterance that each learner produced (p. 528). Results of this study indicated that massed practice was more effective than distributed practice in facilitating accurate production of the target form, but that these two practice conditions did not significantly differ in facilitating processing speed of the target form in spontaneous oral production.

The approach to using RTs by Suzuki (2017a) is informed by Skill Acquisition Theory (Anderson, 1992, 1993). In the context of L2 learning, this theory suggests that learners' abilities in using the L2 accurately and fluently are initiated by the development of 'declarative knowledge' (i.e., knowledge about target L2 forms). Following this, learners develop 'procedural knowledge' (i.e., knowledge of how to use the forms in contexts) through practice. Finally, learners develop 'automatized knowledge' (i.e., knowledge that can be accessed rapidly with low error rates) as a result of repeated practice in using the forms under real-operating conditions (DeKeyser, 2015, 2017). To this end, Skill Acquisition Theory of L2 learning holds that learners' abilities in using L2 forms accurately precedes their abilities in using the forms fluently. In cognitive terms, learners first need to develop knowledge of target L2 forms before being able to recall, access and/or process these forms in real time (DeKeyser, 1996, 1997; Suzuki 2018; Suzuki & DeKeyser, 2017a, 2017b; Suzuki & Sunada, 2019).

To some extent, such a theoretical claim—that the development of learners' knowledge, or mental representations, of specific L2 forms precedes the development of their fast or automatic processing of the forms—is similar to that proposed by usage-based accounts of L2 learning (Norris & Ortega, 2003; Ortega,

2015). Nonetheless, there are differences. For instance, while Skill Acquisition Theory appears to suggest that L2 development constitutes a *linear* process of acquiring knowledge of form-function-meaning mappings, usage-based accounts suggest that L2 development constitute a *dynamic* process of mapping L2 forms onto meaning and function (see Chapter 3). Therefore, while Skill Acquisition Theory stipulates that fast or automatic processing of specific L2 forms may be developed *after* learners acquire knowledge of the forms, usage-based accounts stipulate that fast or automatic processing of specific L2 forms can take place *even before* learners develop elaborate knowledge of the forms (e.g., as a result of recurrent exposure to 'high token input frequency'—see Chapter 3 for details).

To this end, measures of L2 processing speed as employed by Suzuki (2017a) might not be relevant for use in the current study. This is because the current study adopts a distinct theoretical account from that adopted by Suzuki (2017a). Specifically, while Suzuki (2017a) considered only samples of learners' accurate use of the target L2 form in measuring L2 processing speed (p. 528), the current study will consider both learners' accurate and less accurate use of the target L2 form (ECSs). Furthermore, while Suzuki (2017a) excluded learners' accurate use of the target L2 form that contained repairs or reformulations from the analysis, the current study will also analyze the frequency of repairs or reformulations that learners produce when attempting to use the form (ECSs) in real-time production.

The rationale for considering both learners' accurate and less accurate production of the target L2 form in measuring L2 processing speed is predicated upon the idea that learners usually develop "exemplar-based" and "rule-based" knowledge of the form during certain stages of L2 learning. Furthermore, these two types of L2 knowledge typically provide a cognitive basis for exemplar-based and rule-based

processing to occur during real-time L2 production (Skehan, 1996, 1998a). As described in Ch. 2, Section 2.2.2, exemplar-based processing underlies learners' abilities to make use of routine constructions in comprehending and/or conveying meaning in the L2. In contrast, rule-based processing underlies learners' abilities in using L2 forms to construct novel utterances in the L2 (i.e., utterances unlike those that learners have encountered or produced before).

To illustrate, learners of L2 English who have frequently used ECSs containing a high token (e.g., '***is longer than***') in authentic communication might be able to use these structures accurately and fluently in various contexts insofar that this token applies (e.g., *this {movie, meeting, trip, work, road, scarf, story, etc.}* ***is longer than*** *the {other, previous} one*). That is, the use of these structures (ECSs) with the high token ('***is longer than***') is governed by exemplar-based processing. By contrast, in the case where these learners are required to use the structures (ECSs) in new linguistic contexts unlike those they have produced or encountered before, they might not be able to do so accurately and fluently. That is, their use of the structures may be less accurate but fluent, or it may be fluent but less accurate. Such a trade-off between accuracy and fluency is normally due to the results of 'rule-based processing'. As noted in Ch. 2, Section 2.2.2 and Ch. 3, Section 3.2.3, rule-based processing typically requires deeper computational demands than exemplar-based processing. To this end, accurate/inaccurate production of L2 structures does not necessarily correlate with fast/slow processing (at least as viewed from usage-based or cognitivist perspectives). Rather, such processing may be due to distinct types of knowledge (i.e., exemplar-based/rule-based knowledge) that give rise to distinct types of processing (i.e., exemplar-based/rule-based processing). With respect to measuring L2 processing speed, therefore, considering

both accurate and less accurate production of the target L2 form might provide a better account in identifying the average speed of processing a target L2 form in given contexts. In the literature, such a method for measuring L2 processing speed is referred to as 'a measure of L2 speed fluency' (see Section 5.6.1.3, below, for more detailed discussion of this measure).

With regard to analyzing repairs or reformulations, L2 researchers generally agree that this hesitation phenomenon largely corresponds to the level of attentional control underlying L2 processing (de Bot, 1992; Kormos, 2006; Segalowitz, 2010, 2016; Skehan, 2014). That is, the more frequent repairs or reformulations occur during the production of specific L2 forms, the more likely that the processing of the forms during production is governed by attentional control (Kormos, 1999, 2000, 2006, 2011; Skehan, 2009a, 2014). In contrast, the less frequent repairs or reformulations occur during the production of specific L2 forms, the more likely that the processing of the forms during production takes place with little attentional control (Segalowitz, 2016). This measure of hesitation phenomenon as a potential measure of L2 automatization will be discussed in detail in Section 5.6.2.2.

5.6.1.3 L2 Speed Fluency as a Measure of L2 Processing Speed

As indicated in the preceding section, the current study employed a measure of L2 speed fluency, rather than RTs, to measure gains in L2 processing speed. Unlike RTs which are mainly concerned with examining *the time* taken by speakers to produce certain utterances or to respond to particular stimuli in the L2 (de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Suzuki, 2017a), a measure of L2 speed fluency is mainly concerned with examining *temporal features* (e.g., speech rate

and pauses) manifested in speakers' L2 utterances (R. Ellis & Barkhuizen, 2005, Ch. 7; Michel, 2017).

To measure L2 speed fluency, researchers thus usually calculate *speech rate*, i.e., the total number of pruned syllables, including pauses, divided by the total time that speakers need to produce certain utterances in the L2 (Ahmadian, Tavakoli & Dastjerdi, 2015; Derwing, Munro, Thomson & Rossiter, 2009; Lambert, Kormos & Minn, 2017; Lennon, 1990; Michel, 2011; Towell, 2002). The use of speech rate as a measure of L2 processing speed is predicated upon the idea that the speed of cognitive processing underlying learners' L2 use is reflected in the number of syllables that they produce in a given time. That is, the more syllables that learners produce in a given time, the faster their L2 processing speed operates (de Bot, 1992; Kormos, 2006; Segalowitz, 2010, 2016; Skehan, 2009a, 2014; Tavakoli & Skehan, 2005).

Speech rate has been used in task-based L2 performance studies to examine how tasks and their implementations affect learners' L2 processing speed as reflected in their *overall* L2 speech production (for a review, see, e.g., Robinson, 2011b; Skehan, 2016). In the current study, speech rate will be used to examine changes or improvements in learners' processing speed of ECSs. For this reason, the study employs a *local*, rather than global, measure of speech rate.

5.6.1.4 Local Measures of Speech Rate and L2 Development

In the L2 literature, local measures are usually defined as measures that focus on examining learners' use of specific L2 forms during L2 comprehension and/or L2 production (e.g., article usage, relativization, or, in the case of the current study, ECSs) (see R. Ellis & Barkhuizen, 2005; Foster & Wigglesworth, 2016). In line

with this definition, local measures of speech rate may thus be defined as measures that focus only on examining the rate/speed of learners' production of a specific L2 form.

The utility of local measures in studying L2 development is widely accepted in that such measures allow the researcher to track changes or improvements in learners' use of specific L2 forms over time (R. Ellis, 1994; Long, Inagaki & Ortega, 1998; Norris & Ortega, 2003). However, the use of local measures in L2 development studies has mostly been concerned with learners' *accurate use* of target L2 forms. Hence, studies on instructed L2 development have been predominantly based on accuracy data (Norris & Ortega, 2000, 2001; Shintani, Li & Ellis, 2013). Nonetheless, as Sanz and Grey (2015) noted, although accuracy data can provide fruitful information about progression that potentially takes place in learners' developing L2 knowledge, this type of data

> cannot address the status of or access to knowledge resulting from the process under study (p. 303) …. Consequently, a great deal of information is left unattended when using accuracy data as the only dependent measure of effects and then drawing conclusions about the relative effectiveness of instructional conditions (p. 304).

Given the limitation of accuracy data in studying L2 development, there is a growing consensus among L2 researchers that the notion of L2 development might need to be extended to encompass processing dimensions in addition to accuracy (Bialystok, 2011; Bygate, 2016; de Graaff & Housen, 2009; de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; DeKeyser, 2015, 2017; R. Ellis, 2005; Goldberg & Casenhiser, 2008; Jiang, 2007; Kawaguchi & Di Biase, 2012; Lim & Godfroid, 2015; Long, 2015; McDonough & Kim, 2009; Rodgers, 2011; Sanz & Grey, 2015; Skehan, 2014; Suzuki & DeKeyser, 2015). That is, as discussed in Chapter 3, the abstract notion of L2 development may best be characterized *not* in

terms of a static product of accumulating knowledge of L2 forms, but rather as a dynamic process of detecting, encoding and storing particular linguistic information in memory, and subsequently retrieving it to cope with relevant communicative or functional demands. In this regard, employing local measures of speech rate might prove beneficial to provide additional evidence as to how instructional conditions under study affect L2 development (e.g., whether the conditions differ in facilitating faster processing speed of a specific task-relevant L2 form). Chapter 6, Section 6.9, outlines step-by-step procedures for employing local measures of speech rate.

5.6.2 Methods for Measuring L2 Processing Stability and Automatization

As noted earlier in this section, faster L2 processing alone (as indicated, for instance, by a faster rate of production of specific L2 forms by learners) is *in*sufficient to indicate L2 automatization or gains in L2 automaticity. This is because automatization entails not only faster but also more stable L2 processing.

The importance of considering the stability of online L2 processing in investigating L2 automatization was first proposed by Segalowitz and Segalowitz in the early 1990s (Segalowitz & Segalowitz, 1993). These researchers argue that in studying L2 automatization it is first necessary to distinguish between *speed-up* and *automatic processing*. Broadly defined, speed-up processing refers to fast *but effortful* L2 processing (in the sense that although online L2 processing occurs rapidly, the processing is largely controlled by learners' attentional resources). In contrast, automatic processing refers to fast *and effortless* L2 processing (in the sense that online L2 processing not only occurs rapidly, but also takes place with little attentional control). To this end, a key feature that distinguishes speed-up

from automatic processing concerns the level of attentional control that are involved during online L2 processing. Furthermore, according to Segalowitz and Segalowitz (1993), observing the stability of online L2 processing can validly indicate whether online L2 processing takes place with a high or low level of attentional control (Segalowitz, 2003, Segalowitz, Segalowitz & Wood, 1998).

Although L2 researchers generally agree with Segalowitz and Segalowitz (1993) that these two processing features (i.e., speed and stability) can provide valid measures of automaticity, questions remain as to how L2 processing stability may be validly measured in L2 research.

5.6.2.1 The Coefficient of Variation as a Measure of L2 Processing Stability and Automatization

Segalowitz and Segalowitz (1993) proposed the coefficient of variation (CV) as a measure of L2 processing *stability*. CV may be computed by dividing mean standard deviation of reaction times (RTs) by mean RTs. In L2 research, a small CV is usually used to indicate that online L2 processing occurs with a high degree of stability, whereas a large CV is used to suggest that such processing is unstable (Hulstijn, Van Gelderen & Schoonen, 2009; Lim & Godfroid, 2015; Rodgers, 2011; Suzuki, 2018).

With regard to measuring L2 *automatization*, Segalowitz and Segalowitz (1993) argued that three conditions should be met in order to provide reliable evidence of L2 automatization: (1) faster L2 processing speed (as indicated by decreases in RTs), (2) more stable L2 processing (as indicated by decreases in CV), and (3) a significant positive correlation between RTs and CV. Hence, decreases in RTs alone without proportionate decreases in the CV will never yield a significant

positive RT-CV correlation, thereby constituting evidence of a mere speedup, rather than automatic processing.

Motivated by Segalowitz and Segalowitz's (1993) proposal of the utility of CV in measuring L2 automatization, Hulstijn, Van Gelderen and Schoonen (2009) conducted a validation study. Data of Hulstijn et al.'s (2009) validation study were obtained from two different studies conducted earlier: a two-year longitudinal study with 200 participants, and an L2 word training study with 41 participants.

The participants in the longitudinal study were asked to perform a series of tests in L1 Dutch and L2 English. The tests included lexical decision, lexical retrieval (picture naming), sentence verification, and controlled sentence production. On the other hand, the participants in the L2 word training study were asked to perform a visual lexical decision test in the L2 before and after receiving certain instructional treatments.

The mean of RTs and CV for each participant's tests performance were calculated, and their relationship was analyzed. In general, results of Hulstijn, et al.'s (2009) validation study did not provide convincing evidence of the utility of CV as a measure of L2 automatization (p. 555). That is, although the participants' performance on each test became faster as their language proficiency increased, the CV did not always correlate positively with RTs (see Hulstijn, et al., 2009, pp. 565-574, for details).

Lim and Godfroid (2015) conducted a partial, conceptual replication of Hulstijn, et al.'s (2009) validation study with forty Korean learners of L2 English (20 intermediate and 20 advanced) and twenty English native speakers. These participants were asked to perform three speeded tests in English. These included

semantic classification, sentence verification, and controlled sentence construction. Lim and Godfroid (2015) also analyzed the relationship between mean RTs and CV of each participant's performances on each of these tests. Results of Lim and Godfroid's (2015) study indicated that the utility of CV largely depended on the level of processing that particular tests impose on the participants. In particular, these researchers noted that "the CV may be a valid measure of automatization at *the sentence level*, provided the tasks [or tests] used target *lower-level processes* such as word recognition, parsing, and semantic proposition formation" (Lim & Godfroid, 2015, p. 1247, emphasis added).

To this end, the extant validation studies on the utility of CV as a measure of L2 automatization have largely yielded inconsistent findings. Furthermore, findings of the validation study by Lim and Godfroid (2015) suggest that CV *may not* be a valid measure of L2 automatization *at the discourse level*. This is because discourse-level processes normally require high level processing demands in comparison to that required by word- and sentence-level processes (N. Ellis, 2019; R. Ellis, 2002; Paradis, 2004, 2009; Robinson, 2003b, 2005b, 2010, 2011a). Given that the current study employed discourse-level L2 production tests to examine automatization (i.e., gains in automaticity) from the provision of pre-modified input or interactional feedback (see Chapter 6, Section 6.7 for details), examining both *temporal features* and *hesitation phenomena* manifested in learners' L2 production might better reflect the level of automaticity underlying such production (Segalowitz, 2010, 2016). In the L2 literature, temporal features and hesitation phenomena are usually examined through the use of 'measures of L2 utterance fluency' (Skehan, 2003, 2009a, 2014; Tavakoli & Skehan, 2005).

5.6.2.2 Measures of L2 Utterance Fluency as a Measure of L2 Automatization

Measures of L2 utterance fluency are informed by the *adapted* model of L1 speech processing proposed by Levelt (1989, 1999). As discussed in Sections 5.1, features that distinguish L1 from L2 speech production concern the type of processing that governs L1 or L2 speech production (i.e., *parallel* vs. *serial*) and the internal process of linguistic encoding that occurs during 'formulation' in the L1 or L2 (i.e., *automatic* vs. *controlled*) (de Bot, 1992; Kormos, 1999, 2000, 2006, 2011; Segalowitz, 2010, 2016; Skehan, 2009a, 2014). Furthermore, features that determine whether the internal process of linguistic encoding occurs automatically or less automatically (controlled) concern both the *size* and the *type* of linguistic resources that L1 and L2 speakers normally possess. That is, while speakers normally possess a *large* amount of linguistic resources for their L1, speakers normally possess a *limited* amount of linguistic resources in their L2. Moreover, while linguistic resources that L1 speakers possess are normally stored *implicitly* in their long-term memory, linguistic resources that many L2 speakers possess are usually stored *explicitly* in their long-term memory. According to many L2 researchers, this range of differences affect online linguistic processing during real-time L1/L2 production (e.g., de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Lambert, Kormos & Minn, 2017). That is, while L1 speakers are usually capable of accessing their linguistic resources automatically (i.e., without the need of attentional control), L2 speakers are usually less able to do so. Rather, access to linguistic resources by L2 speakers during real-time L2 production normally requires certain levels of attentional control (see also N. Ellis, 2005; R. Ellis, 2005; Kormos, 2006; Muranoi, 2007; Paradis, 2009; Rebuschat, 2013, 2015; Skehan, 2009a; Segalowitz, 2010).

With respect to speech behavior, therefore, while L1 speakers are normally able to deliver their speech *fluently*, L2 speakers are usually *less fluent* in doing so. To this end, fluency manifested in observable speech might well reflect the fluency of online linguistic processing that gives rise to the speech (see Section 5.2 for details). Segalowitz (2010, 2016) referred to these two types of fluency as 'utterance fluency' and 'cognitive fluency' respectively.

In L2 research, utterance fluency is usually observed by examining temporal features and hesitation phenomena that are manifested in observable speech. According to Segalowitz (2010, 2016), both temporal features and hesitation phenomena are *intimately linked* to specific dimensions of cognitive fluency, especially those that Skehan (2003) has referred to as *speed*, *breakdown* and *repair fluency* (Segalowitz, 2016; p. 81; see also de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013, for similar arguments).

As discussed in Section 5.6.1.3, speed fluency concerns the speed of L2 processing that occurs as L2 speakers plan and execute speech. In L2 research, speed fluency is usually measured by calculating speech rate or number of pruned syllables, including pauses, that L2 speakers produce per second (Ahmadian, Tavakoli & Dastjerdi, 2015; Derwing, Munro, Thomson & Rossiter, 2009; Lambert, Kormos & Minn, 2017; Michel, 2011; Towell, 2002). In contrast, breakdown fluency concerns the ongoing flow of L2 processing during which L2 speakers plan and execute speech. To measure breakdown fluency, researchers usually calculate the frequency of filled and unfilled pauses that occur within clausal boundaries per second. Finally, repair fluency concerns error rates that occur during planning and executing speech. Repair fluency is usually measured as frequency of repairs,

reformulations or replacements that occur during production per second (Skehan, 1996, 1998a, 2009a; Skehan & Foster, 1997, 1999; Skehan & Tavakoli, 2005).

Among the first empical evidence supporting the validity of L2 utterance fluency measures as a measure of L2 automatization was reported in Segalowitz and Freed (2004). These researchers examined whether gains in L2 utterance fluency were related to gains in automaticity in L2 processing. The study was conducted with 40 adult learners of L2 Spanish learning the L2 either in a US or Spanish university for a semester. Gains in L2 utterance fluency were measured by examining changes in the speech rate of the learners' L2 oral production as they completed L2 proficiency interviews at the beginning and end of the study. On the other hand, gains in L2 automaticity were measured by observing changes in the speed (RTs) and stability (CV) of L2 processing as these learners completed speeded semantic classification tests requiring lexical access (e.g., *a girl = animate*; *a chair = inanimate*). Results of Segalowitz and Freed's (2004) study generally indicated that gains in the L2 utterance fluency correlated significantly with gains in automaticity in online L2 processing.

With respect to L2 development, studies by de Jong and associates (e.g., de Jong, 2016; de Jong, & Bosker, 2013; de Jong, Groenhout, Schoonen & Hulstijn, 2015; de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013) further suggest that measures of L2 utterance fluency can also predict the development of learners' L2 knowledge. De Jong, et al. (2013), for instance, analyzed L2 oral task performances of 179 adult learners using a variety of L2-utterance-fluency measures: (1) speech rate (operationalized as mean duration of syllables, or "speaking time divided by total number of syllables" [p. 899]), (2) number of filled and unfilled pauses, (3) mean duration of unfilled pauses, and (4) number of reformulations and repetitions.

To measure the learners' L2 knowledge, these researchers employed three speeded linguistic tests (vocabulary, grammar and pronunciation). Results of de Jong, et al.'s (2013) study indicated that measures of L2 utterance fluency significantly correlated with learners' L2 knowledge, except for mean duration of unfilled pauses.

To this end, the utility of measures of L2 utterance fluency in measuring L2 automatization and L2 development has not only been motivated theoretically, but also has been tested empirically. From both theoretical and empirical standpoint, significant increases in the speech rate *accompanied with* significant decreases in mid-clause pauses, repetitions and repairs during spontaneous L2 production can indicate automatization—in that the underlying cognitive processes not only operate more rapidly but also proceed with a higher degree of stability (Segalowitz, 2010, 2016). In contrast, absence of significant changes in any of these utterance-fluency measures might indicate that automatization does not take place (see also de Bot, 1992; Howell & Au-Yeung, 2002; Kormos, 1999, 2000, 2006; Lambert, Kormos & Minn, 2017; McLaughlin, 1987; Skehan, 2009a, 2014, for similar arguments). Questions remain, however, as to how measures of L2 utterance fluency may be used to examine automatization of learners' spontaneous production of a specific task-relevant L2 form in the current study (i.e., ECSs).

5.6.2.3 Local Measures of L2 Utterance Fluency Employed in the Current Study

In line with the discussion in Section 5.6.1.4, the current study employed *local* rather than global measures of L2 utterance fluency. That is, these measures were specifically used to examine gains in automaticity of learners' spontaneous

prodiction of a *specific* task-relevant L2 form (ECSs) in naturally-occurring L2 communication. Furthermore, based on the discussion in the previous sections, gains in automaticity are operationally defined in the current study as significant increases in speed fluency of the form (indexed by a faster rate of production of ECSs per second) *accompanied with* significant decreases in breakdown and repair fluencies of the form (indexed by low frequency of filled and unfilled pauses, verbatim repetitions and repairs during the production of the form per second). In this respect, significant increases in the speed fluency alone *without* significant decreases in the breakdown and repair disfluencies in learners' producion the form (ECSs) will be regarded as an indicator of a mere speedup of the form. Chapter 6, Section 6.9, outlines the procedures for measuring automaticity the target L2 form in the current study.

5.7 Summary

This chapter has provided an operational definition of uptake and automaticity (dependent variables) as employed in the current study. Based on usage-based accounts of L2 learning, uptake is operationally defined in this study as (1) the number of times a task-relevant L2 form (ECSs) was attempted by learners as they completed given tasks orally, and (2) the proportion of accurate attempts in using the form. Furthermore, using a widely accepted model of L1 and L2 speech production, automaticity is operationally defined as significant increases in the speed fluency (indexed by a faster rate of production) along with significant decreases in the breakdown and repair disfluencies (indexed by low frequency of filled/unfilled pauses, verbatim repetitions and repairs) as learners attempted to produce ECSs in naturally-occurring task-based L2 communication. The specific

ways in which these dependent variables are coded in the current study as a means to estimate the effectiveness of pre-modified input or interactional feedback in task-based L2 instruction (independent variables) are discussed in the next chapter.

Chapter 6

Methodology

With an aim of replicability, this chapter provides a detailed discussion of the objectives, design, participants, materials, target L2 form, procedures, measurements, and statistical analysis used to analyze the data for the study.

6.1 Objectives

This study investigated the effectiveness of pre-modified input and interactional feedback during task-based L2 instruction on the uptake and automaticity of English comparative structures (ECSs). The acquisition of ECSs (Clark, 1970; Clark & Card, 1969; Collins, 1994; Kennedy, 2007) is under-researched in L2 studies. Tasks were adopted for the present study for which previous research incorporating English native-speaker baseline data (Lambert, 2019; Lambert & Nakamura, 2019) had shown that ECSs to be both useful and commonly occurring, though not essential.

In this study, pre-modified input was realized—as discussed in Chapter 4—through model performances in task-based L2 instruction that were prepared to contain exemplars of ECSs in contextualized communication (R. Ellis, 2001; Long, 2019; Shintani, 2016; Yano, Long & Ross, 1994). In contrast, interactional feedback was realized through the provision of ‘recasts’ when learners produced the form inaccurately, and ‘confirmation checks’ when learners continually avoided the form while performing the tasks (Gass & Mackey, 2015; Long, 1996, 2007, 2009, 2015;

Mackey, 1999; Mackey & Oliver, 2002; Mackey, Oliver & Leeman, 2003; Oliver, 1998, 2002).

The research questions for this study were as follows:

1. Do pre-modified input and interactional feedback differ in their *effects* on the uptake and automaticity of ECSs as reflected in differences between intermediate-level adult Indonesian learners' pre-test and post-test oral task performances?
2. How *stable* are gains in the uptake and automaticity of these structures from the two conditions as reflected in differences between the learners' post-test and delayed post-test oral task performances?
3. How *transferable* are gains in the uptake and automaticity of these structures from the two conditions as reflected in differences between the learners' oral performances on the same and parallel tasks given in the post-test and delayed post-test?

Based on the review of the literature presented in the previous chapters, four hypotheses were formulated in line with the research questions:

1. Interactional feedback will be more *effective* than pre-modified input in promoting *uptake* of ECSs. This is because interactional feedback is more likely to induce learners to carry out *cognitive comparisons* between their non-target-like uses of the structures and the target-like use provided by their interlocutor (the researcher), leading them to notice gaps in their L2 resources during communication.
2. Interactional feedback will also be more *effective* than pre-modified input in promoting *automaticity* of ECSs. This is because interactional feedback is more

likely to stimulate *cognitive rehearsing* in the learners' working memory, especially when they attempt to use the structures in their own production after having them recast or modelled in confirmation checks by their interlocutor (the researcher).

3. The gains in the uptake and automaticity of ECSs that interactional feedback promotes will also be more *stable* than those promoted by pre-modified input. This is because learners' production and *re*production of the structures following recasting or modelling in confirmation check by the interlocutor (the researcher) are more likely to stimulate *deep syntactic processing* which facilitates retention of target linguistic information in the learners' long-term memory.
4. The gains in the uptake and automaticity of ECSs that interactional feedback promotes will also be more *transferable* to parallel tasks than those promoted by pre-modified input as a result of the same process of *deep syntactic processing*.

As discussed in Chapter 5, uptake and automaticity of ECSs were assessed through performance testing on tasks comparable to those that learners performed during the instructional treatments (see Section 6.7 for details).

6.2 Design

The present study employed a pre-test, post-test, delayed post-test design with two experimental groups (pre-modified input and interactional feedback) and one comparison group (no input or feedback). These three groups received three task-based instructional sessions based on the same task-based instructional materials

between the task-based pre-test and post-test, but the procedures used to implement these materials differed in each condition (see Sections 6.4 and 6.6 for details).

6.3 Participants

The participants were fifty-one adult Indonesian learners of English as a foreign language ranging in age from 18 to 25 years (*M* = 20.65, *SD* = 1.98). These participants were students in the English language teaching (ELT) program of a university in East Java Province, Indonesia. Most of the participants were female, and none of these participants had travelled or lived in an English-speaking country prior to participating in this study. These participants were recruited for the study based on their ability to complete a recruitment task (see Appendix 1; see also Section 6.6, below).

Prior to commencing their study at the university, the participants had already learned English for six years in high school following a structural syllabus. They were also required to pass national examinations before continuing their study at tertiary level. The national examinations for English consisted of tests of receptive skills (i.e., listening and reading) and grammar. The tests were administered in paper-based format and consisted of multiple-choice items. The curriculum in the ELT program of the university aimed to prepare the participants to become high school English teachers in contexts where they would be expected to implement a structural syllabus. Therefore, an emphasis in the ELT training provided in the program was on developing their explicit knowledge of English grammatical structures along with their listening and reading skills in English.

Nonetheless, there were also four units in the program devoted to helping the participants to develop their oral proficiency in English. These four oral units were

delivered during the first two years in the program and were taught for 90 minutes per week over four 14-week semesters.

The four oral units were graded based on the extent to which they required controlled or free language production practice. In the first oral unit, for instance, learners (participants) were simply asked to practice their oral skills through enacting scripted dialogs with their peers. The dialogs were prepared by the teacher in advance. In this way, learners' oral production was largely controlled. This first oral unit was also aptly named "guided speaking".

In the second oral unit, learners were usually asked to construct their own dialogs with peers before performing them in front of the class. Classroom activities normally involved 'role plays' between learners. The main goal of this unit was to enable learners to carry out interpersonal communication in English in a range of communication contexts (e.g., on the telephone, in a market, bank, etc.). This second oral unit was called "responsive speaking".

The third oral unit was called "productive speaking". As the name implies, learners were expected to use English productively, albeit not spontaneously. Oral practice in this unit comprised storytelling and giving oral presentations in English. Learners were usually given preparation time (usually one or two weeks) to prepare the materials that they would present in front of the class.

Finally, the fourth oral unit was called "argumentative speaking". This last oral unit comprised both productive and spontaneous oral practice. The main activities involved debates aimed at helping learners to express ideas clearly and spontaneously in English.

To this end, the four oral units in the program were not specifically designed in ways that might have helped learners (participants) use ECSs in real time orally.

6.4 Materials

The materials used in the experimental groups (with pre-modified input and interactional feedback, respectively) and in the comparison group (with no input or feedback) consisted of twelve sets of oral description tasks adopted and adapted from Lambert (2019). Each task set contained four pictures of very similar items (see Appendix 2). These tasks were designed in a way that would result in a one-way flow of information. That is, participants in the three conditions were either assigned the role of 'speaker' (information provider) or 'listener' (information receiver) for each task performance (see Figure 6.1, below). Although the flow of information was in one direction (i.e., from the speaker to the listener), the tasks still allowed interaction in the experimental groups for the purpose of requesting clarification or for providing interactional feedback (see Chapter 4 for details; see also R. Ellis, 2001; Erlam, 2019; Erlam & Ellis, 2018; Gass & Varonis, 1994; Long, 1983b, 1983c, 1985, 2015, 2019; Mackey, 1999; Oliver, 1995, 1998, 2002; Shintani, 2011, 2012, 2016; Shintani & R. Ellis, 2010 for similar designs).

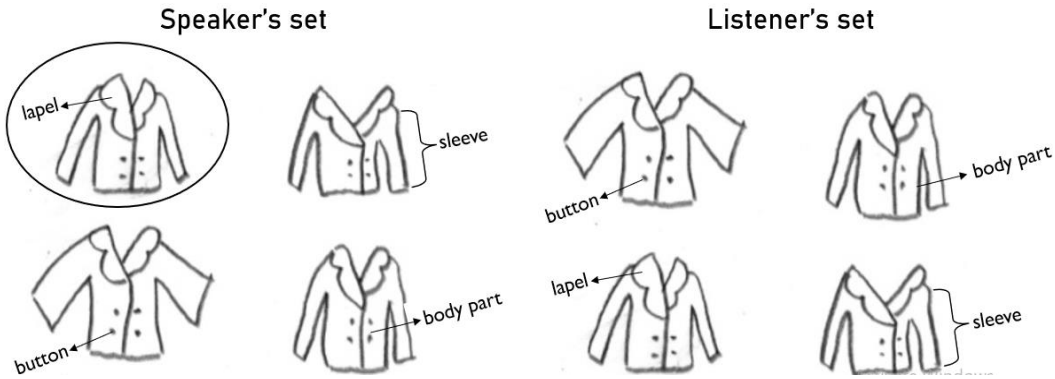


Figure 6.1: Sample materials (from Lambert, 2019)

As illustrated in Figure 6.1, both participants had access to four pictures when performing the tasks. On the speaker's picture set, a target item that needed to be identified was circled. The pictures were ordered differently on the speaker's and listener's picture set so that the target item could not be identified by its position on the page. The tasks thus required the speaker to describe a target item in enough detail that the interlocutor could identify it from very similar items. Furthermore, since the study was concerned with the uptake and automaticity of specific task-relevant L2 morphosyntactic structures, key lexical items that might interfere with the online processing of the structures during performance were provided in the tasks. In this way, the study did not confound lexical and syntactic processing during online L2 use (de Jong, 2016; Skehan, 2009a). Put another way, as the participants had the required lexis, their attention was not absorbed with a lexical search and/or paraphrase (Lambert, 2019). As a result, the observed variation in the participants' oral task performance reflected their capacity to encode meaning morphosyntactically in their L2 (see Chapters 3 and 5 for a detailed discussion of linguistic processing).

Table 6.1. The distribution of the task sets used in the study

Week	**Day**	**PI**	**IF**	**C**	**Tasks employed in each lesson**
1	1	Pre-test	Pre-test	Pre-test	Task 1
	2	Lesson 1	Lesson 1	Lesson 1	Tasks 4-6
	3	Lesson 2	Lesson 2	Lesson 2	Tasks 7-9
	4	Lesson 3	Lesson 3	Lesson 3	Tasks 10-12
	5	Post-test	Post-test	Post-test	Tasks 1 & 2
2-3	6-21	No instruction was given during this period. [1]			
4	22	Delayed post-test	Delayed post-test	Delayed post-test	Tasks 1 & 3

Notes: PI = pre-modified input condition; IF = interactional feedback condition; C = comparison condition.

[1] in an attempt to control for exposure to the target form outside the study, teachers (especially those teaching English grammar in the university) were asked not to focus on teaching English comparative structures during these two weeks.

The twelve description tasks used in the study were implemented according to Table 6.1. That is, Set 1, 2 and 3, which had been used in previous research with native and non-native speakers of English (Lambert, 2019; Lambert & Nakamura, 2019) to establish baseline use of English comparative structures were used in the pre-test, post-test and delayed post-test, while Set 4 to 12 were parallel versions of the tasks and were used only in the treatments.

6.5 Target L2 Form

As English comparative structures (ECSs) had been shown to be both useful and commonly occurring on the tasks selected for the study, the structures were selected as a target form to compare the effectiveness of pre-modified input and interactional feedback during task-based L2 instruction. More specifically, the study focused on the use of ECSs that involves explicit comparative marker *-er* or *more*, as in *this jacket <u>has slimmer sleeves (than</u> the other three jackets)*; *the buttons in the middle <u>are more symmetrical (than</u> the top right button)* (Collins, 1994; Kennedy, 2006, 2007; Kennedy & McNally, 2005; Peters, 1996; Sapir, 1944; Stassen, 1984).

As noted in Section 6.5, previous research by Lambert (2019; Lambert & Nakamura, 2019) demonstrated that ECSs constitute task-relevant structures that both native and non-native speakers of English naturally used when describing items shown in the tasks employed in the current study. Furthermore, a pilot study conducted with a small number of Indonesian learners (N = 9) at the university at which the present study was conducted confirmed the naturalness of these structures with the target population for the study. Furthermore, while ECSs constitute a commonly taught form in ELT practices in Indonesia, these structures

are not commonly practiced in communicative contexts, at least in the instructional context where this study was conducted (see Section 6.3, above). Hence, examining improvements in the uptake (use and accurate use) and automaticity (fluent use) of the structures by the target population was possible.

The present study thus *did not* examine whether pre-modified input and interactional feedback facilitate learning of previously unknown L2 forms, but rather whether these two instructional options facilitated the consolidation of learners' knowledge of a specific task-relevant L2 form so they were able to use the form more accurately and fluently in spontaneous L2 production.

6.6 Procedures

During the recruitment process the researcher went to each class announcing that there was an opportunity for all learners to participate in a study. They were informed that their participation would involve meeting with the researcher six times over a four-week period to complete some speaking or listening tasks in English. These meetings would be audio recorded for research purposes only and would not be used to evaluate or identify them personally in any way. Learners were also made aware that they would receive and sign a consent form if they decided to participate in the study.

Learners were accepted into the study based on their abilities in performing an oral recruitment task in English. The recruitment task constituted a parallel version of the main tasks used in this study (see Appendix 1). The use of the recruitment task aimed to ensure that all learners participating in this study were at a high enough level of proficiency to unambiguously identify the target referent and, thus, complete the tasks. That is, the learners' proficiency was neither too low so that

they would not be able to carry out the tasks, nor that they had already automatized their knowledge of English comparative structures, leaving no room for improvement (Mackey & Gass, 2005, p. 138). The criteria for doing so were two: (1) the time taken by learners to complete the recruitment task, and (2) their ability to use comparative structures when completing the task (see also Chapter 2, Section 2.1.3: Assessment of Task-Based L2 learning).

With regard to (1), learners who were able to complete the task in less than 3 minutes or were not able to complete the task after 10 minutes were excluded from the study. This was done by observing the time taken by learners to complete the task. With regard to (2), learners who were able to use comparative structures accurately (i.e., ≥ 50% accuracy) in at least two different linguistic contexts (e.g., *X* *is bigger than* *Y, but Y* *is wider than* *X*), as well as were able to use the structures fluently (i.e., ≥ 2,93 syllables per second) when attempting to complete the recruitment task were also excluded from the study. This was done by examining transcripts of the learners' oral task performance and by assessing the rate of their production of the structures as they completed the task (see also Section 6.9: Measuring Uptake and Automaticity of the Target L2 Form). It it is worth noting that none of the prospective participants met these criteria.

The use of the recruitment task in this study was also intended to provide prospective participants with background knowledge concerning the nature of the tasks and what was required of them. Hence, as the participants performed a task in the pre-test session, they were already familiar with the task procedure and, thus, were able to focus on using their L2 resources to cope with the task demands. In this way, the participants' task performances in the pre-test session might have

reflected their true capacity to complete the task on the pre-test (Lambert, Aubrey & Leeming, 2020; Norris, 2009; see also Chapter 8).

Following the recruitment task, participants were randomly assigned to one of the three instructional conditions: (1) pre-modified input, (2) interactional feedback, or (3) comparison. Detailed descriptions of the treatments are provided in separate sub-sections below (see Sections 6.6.1 – 6.6.3). The treatments in the three conditions were conducted on a one-and-one basis between each participant and the researcher. The treatments were also conducted with one group before moving to the next group to maintain consistency and avoid switching the implementation procedures each time.

At the outset of every instructional session, participants in all three conditions received instructions in their first language (Indonesian) specifying how they would need to perform the tasks (see Appendix 2 for complete English translations of the instructions given). Furthermore, before doing the tasks, these participants were reminded that what they would do was not a test of their knowledge or skills. Thus, they did not need to worry about being graded. The participants were also encouraged to use English freely as if talking to a friend. They were also told that they needed to pretend that they were completing the tasks on the telephone. Hence, they would not be able to communicate using gestures. A screen was set up between the researcher and the participant to avoid the use of non-verbal communication.

6.6.1 Pre-modified Input Condition

In this condition, participants were assigned the role of ‘listener’ where they needed to identify the target items in three tasks based on descriptions given by the

researcher. For each of the three tasks, the participants were given a set of four pictures of similar items (see Figure 6.1: Listener's set, for an example). The researcher described the target item orally, and the participants had to identify which one of the four was being described. To ensure that all participants in this condition received equal amounts of input, they were asked to listen to the entire description given by the researcher before selecting an item. In other words, participants were *not* allowed to select an item until the researcher had finished his full description of the target item (see Appendix 2).

The descriptions for each version of the task were fixed and given to all participants in the pre-modified input condition. The descriptions were prepared in advance with the number of occurrences of the target structure controlled (see Appendix 3). Furthermore, the total occurrences of comparative structures in this condition were counted and used as a criterion for providing a comparable amount of opportunities for the occurrences of comparative structures in the interactional feedback condition (see Section 6.6.2, below).

When describing the items, the researcher spoke naturally (i.e., neither too fast nor too slow). At the end of each description, participants were then asked to choose an item. Before showing the item to the researcher, participants were asked to listen for a second time to confirm their answer. This repetition was intended to ensure the comparability of the time on tasks between this condition and the other two conditions (i.e., the interactional feedback and comparison conditions).

After the repetition, participants were then asked to show their answer to the researcher. The researcher then told the participants whether their answer was correct or not. Although all participants in this condition were allowed (but not

required) to request a clarification when they failed to comprehend any parts of the descriptions being given (R. Ellis, 2001; Shintani, 2016), none of them did so. This does not mean, however, that the participants did not interact with the researcher when receiving the input. Nonetheless, the interaction that did take place was primarily concerned with confirming comprehension of the input that the participants received (e.g., *yes, ok, there are two [items that look like that]*, see Chapter 4, Excerpt 4.4.1, for an illustration).

6.6.2 Interactional Feedback Condition

The tasks and procedures given to participants in this condition were identical to the pre-modified input condition described above except that the roles were reversed between the researcher and the participants. That is, participants in this condition were assigned the role of 'speaker' where they needed to describe target items clearly and unambiguously to the researcher so that he could identify them from the sets of four very similar items (see Figure 6.1). Following the procedure outlined in Lambert and Nakamura (2019), the criteria for successful task completion applied in this condition was that enough information was supplied by the participants that a third listener, who was unfamiliar with the tasks, would have been able to identify the target referent unambiguously as well (p. 254).

Participants in this condition were encouraged to use English only when describing target items in the tasks. In the case where the description that the participants gave was ambiguous in that it did not clearly discriminate between target items and distracters, the researcher prompted them to provide more detailed information about the target items by saying '*tell me more*' or '*I have several like that*' (following the procedure outlined in Lambert (2019) and Yule (1997)). When this

failed, the researcher prompted key differences between items in the tasks with 'confirmation checks' (e.g., '*Is X slightly bigger than Y?*'). As outlined in Chapter 4, such a prompt usually led to the occurrence of the target L2 form in natural communication (Gass, 1997) and, thus, helped make the exposure to the target L2 form in this condition comparable to that in the pre-modified input condition (Long, 2009; Mackey, 1999). Finally, in the case where participants produced the target L2 form inaccurately when comparing items in the tasks, the researcher provided interactional feedback in the form of 'recasts' (Long, 1996, 2007, 2015; see Chapter 4, Excerpts 4.4.2 and 4.4.3, for illustrations).

6.6.3 Comparison Condition

The tasks and procedures in the comparison condition were identical to the interactional feedback condition described above in that participants were assigned the role of 'speaker'. Unlike the interactional feedback condition, however, participants in this comparison condition received neither recasts nor prompts that contained the target L2 form while performing the tasks orally. That is, when the participants failed to provide clear or specific descriptions about a target item, the researcher only said '*tell me more*' or '*I have several like that*'. In this way, this condition resembled what Long (1991, 1998; Long & Robinson, 1998) refers to as a Focus on Meaning (FonM), and served as a baseline of improvement that participants made from performing the tasks alone (i.e., without interactional input or feedback on the target form) (Lewis-Beck, Bryman & Liao, 2004, pp. 153-155; Rogers & Révész, in press).

6.7 Task-Based Performance Tests

The pre-test, post-test and delayed post-test in this study consisted of task-based L2 performance tests (Long, 2015, 2016). That is, participants were asked to perform versions of the tasks that were not used in the treatment sessions (see Table 6.1, above).

In all test sessions, participants were assigned the role of 'speaker', where they needed to describe one target item to the researcher clearly and unambiguously so that he could identify which of the four pictures it was. A screen was also set up between the researcher and the participant to avoid non-verbal communication and, thus, to require that all essential aspects of the descriptions be verbalized. In this way, the testing conditions were designed in a way that would allow the use of the target L2 form to arise naturally in connection with task completion based on participants' own L2 resources.

As in the treatments, the tests were also carried out on a one-on-one basis between the researcher and each participant in a quiet room to enable clear recordings. In turn this allowed the the participant's L2 oral task performances in tests to be transcribed verbatim. Computer-based analysis of speech production using PRATT (Boersma & Weenink, 2017) was used to transcribe and detect silent pause behaviors manifested in the participant's L2 oral task performance (see Section 6.8, below, for a more detailed description for recording and transcribing data).

Participants in all three conditions performed the tasks in the same way during the testing sessions in that they were asked to describe features of a target item in as much detail as possible to the researcher. No information was provided regarding what to describe or how to describe it, nor was corrective feedback of any type

provided to participants during the testing sessions. These procedures were intended to test the uptake and automaticity of the ECSs from the respective treatments.

As indicated in Table 6.1, the pre-test consisted only of Task 1, while the post-test consisted of Task 1 and Task 2. Finally, the delayed post-test consisted of Task 1 and Task 3. The inclusion of the same task (i.e., Task 1) and new tasks (i.e., Tasks 2 and 3) in the post-tests was intended to identify: (1) the *learning gains* from the provision of pre-modified input or interactional feedback in task-based L2 instruction; (2) the *stability* of the gains from each of these task-based instructional conditions over time; and (3) the *transferability* of the gains to comparable tasks (see Section 6.10: Statistical Analysis, below).

6.8 Data Collection and Transcription

To record the participants' oral task performances, a digital audio recorder was placed near them. Participants, however, could not see the recorder, since it was placed behind the screen. The recording was stored in an MP3 file format.

To transcribe the data, PRAAT software version 6.0.29 designed by Boersma and Weenink (2017) was used. This software enabled the researcher to listen to any part of the recording multiple times and, at the same time, transcribe what was being uttered verbatim (de Jong & Wempe, 2009; Kormos & Préfontaine, 2017). Step-by-step procedures for the transcription process used is provided in Appendix 4.

Prior to carrying out any measurements, pruned and unpruned transcription of the audio recordings containing learners' production of the target L2 form (ECSs) were prepared. These pruned and unpruned transcriptions were needed to measure

different domains associated with the uptake and automaticity of the form. Pruned transcriptions were used to measure uptake, error-free uses, and the rate of learners' production of ECSs, while unpruned transcriptions were used to observe breakdown and repair fluencies that took place during spontaneous production of ECSs.

With regard to pruning, false starts (i.e., utterances containing the use of ECSs but were abandoned before they were completely uttered) were pruned and, thus, were excluded from analysis (e.g., ~~*the left is more*~~ *the jacket is rather small I think*). In the case of repairs, only participants' final production was considered for analysis (e.g., *the jacket* ~~*is is is small*~~ *is smaller than the others*) (Foster, Tonkyn & Wigglesworth, 2000).

6.9 Measuring Uptake and Automaticity of the Target L2 Form

As discussed in Chapter 5, 'uptake' was operationally defined in this study as participants' incorporation of ECSs from the input or feedback into their L2 production. On the other hand, 'automaticity' was operationally defined as participants' abilities in using ECSs rapidly without undue pauses, repetitions and repairs in naturally-occurring task-based L2 communication.

To measure uptake, the *proportion* of AS-Units containing ECSs that participants produced when completing tasks in the pre-test, post-test and delayed post-test was first calculated. This is done by dividing the number of AS-Units containing ECSs with the total number of AS-units that participants produced during task work. Following this, all instances of ECSs were analyzed in terms of their accuracy. To do so, an error analysis was carried out. Following R. Ellis and Barkhuizen (2005, pp. 51-72), each accurate production of ECSs by participants (e.g., *X* ***is {smaller,***

more upright} than *Y*) was given a score of 1, while their inaccurate production (e.g., **X smaller than Y; *X more upright than Y; *X is big than Y*) was marked 0. A research assistant was trained to code for the accuracy of the participant's use of ECSs during task performance. Following this, 10% of the database were independently coded by the researcher and the research assistant. An inter-rater agreement was established, resulting in nearly perfect inter-rater agreement (i.e., 98% similarity). Such agreement might be due to the fact that accurate and inaccurate use of ECSs were clearly different, at least in the context where they were produced by the participants in the current study. Finally, the remainder of database containing the use of ECSs were coded by the researcher, and the *proportion* of participants' accurate uses of ECSs was calculated by dividing the number of accurate uses of ECSs with the total number of ECSs produced (both accurate and inaccurate).

With regard to measuring automaticity, participants' production of ECSs was calculated from the termination of the last syllable of the word preceding the production of ECSs to the onset of the first syllable of the next word following the word '*than*' (e.g., *this jacket*(start) ***is smaller than*** (end)*the other three jackets*) or to the final syllable of a given comparative (e.g., *this jacket*(start) ***is smaller***(end); *the other jackets*(start) ***are more upright***(end)). Next, participants' production of ECSs was examined in terms of three measures of fluency: *speed fluency*, *breakdown fluency* and *repair fluency* (Skehan, 2009a, 2009b; Skehan & Tavakoli, 2005). As discussed in Chapter 5, such measures of fluency are intimately linked to specific dimensions of automaticity in L2 processing: *speed* and *stability* (Segalowitz, 2010, 2016; see Chapter 5, Section 5.6.2.2 for details).

In the present study, **speed fluency** was assessed by calculating the total number of pruned syllables, including filled (e.g., '*um*', '*uh*') and unfilled pauses (.), per second as participants attempted to produce ECSs (e.g., *this jacket* ***is uh (.) smaller than*** *the other jackets*). On the other hand, **breakdown fluency** was assessed by calculating the frequency of filled and unfilled pauses that took place within and/or between structural boundaries per second (e.g., *this jacket is* ***uh*** *smaller than* ***(.)*** *the other three jackets; this jacket* ***uh*** *has more buttons*). The occurrences of filled and unfilled pauses were coded manually by examining unpruned transcriptions. Following de Jong (2016; de Jong & Bosker, 2013), the present study considered silences of 0.25 seconds or longer as unfilled pauses. PRAAT software (Boersma & Weenink, 2017) was used to detect these silences reliably (de Jong & Wempe, 2009; Kormos & Préfontaine, 2017). Furthermore, verbatim repetitions and formulaic fillers such as '*I think*', '*you know*', etc. which occurred within and/or between structural boundaries (e.g., *this jacket* ***is is I think*** *is smaller than the other three jackets), and it* ***has um*** *has* ***you know*** *more buttons*) were also coded manually. These features were also considered the same as filled pauses since, as Skehan (2014) indicates, "they contribute no meaning to the ongoing discourse, and serve mainly to ease the pressure of time" (p. 20). Finally, **repair fluency** was assessed by calculating the frequency of self-corrections, replacements and/or reformulations that occurred as learners attempted to produce ECSs per second (e.g., *this jacket* ***is big than I mean*** *is bigger than the other three jackets*). The occurrences of repair fluency were also coded manually by examining unpruned transcriptions.

6.10 Statistical Analyses

Once the scores from these measures of uptake and automaticity of ECSs were obtained, descriptive statistics were performed for all three groups on all tests. Following this, preliminary data screening was performed to test the normality of distribution as well as the homogeneity of variance of the test scores. As will be discussed in the next chapter, the analysis was conducted by running separate non-parametric statistics because the assumption of normality of score distributions for all of the score distributions for both uptake and automaticity was not met. Nor could these distributions be corrected through transformation.

To test the null hypothesis for between-subjects comparisons, Kruskal-Wallis H tests followed by Mann-Whitney U tests and Bonferroni corrections were carried out (Kraska-Miller, 2014). To test the null hypothesis for within-subjects comparisons, Friedman tests followed by Wilcoxon signed-rank tests and Bonferroni corrections were performed. Finally, effect sizes for non-parametric analyses (r) were calculated to determine the magnitude of difference for between- and within-subject measures.

6.11 Summary

This chapter has outlined the objectives, research questions, hypotheses and the design of the study. It has also provided information about the participants, materials used and specific procedures employed to transcribe, code and analyze the data in the study. The next chapter will present the results of statistical analyses that were used to test the research hypotheses outlined at the beginning of this chapter.

Chapter 7

Results

This chapter begins by presenting a series of tests performed to screen the data to determine which statistical analyses were appropriate. Next, the results of the statistical analyses are presented in separate sections. Finally, the findings are reported in conjunction with the research questions posed in Chapter 6, Section 6.1.

7.1 Preliminary Data Screening

A series of tests were performed using SPSS23 for Window to examine the normality (Shapiro-Wilk, $p > .05$) and the homogeneity of variance (Levene's test, $p > .05$) of the three (pre-modified input, interactional feedback and comparison) groups' pre-test, post-test and delayed post-test scores on each of the different dependent variables. As stated in Chapter 6, Section 6.9, these dependent variables include: (i) the proportion of AS-units containing English comparative structures (ECSs); (ii) the proportion of error-free uses of ECSs; (iii) the total number of pruned syllables uttered per second in the use of ECSs; (iv) the frequency of combined filled and unfilled pauses, verbatim repetitions and formulaic expressions that took place within and between ECS structural boundaries per second; and, finally (v) the frequency of combined self-corrections, replacements and reformulations per second as learners attempted to use ECSs in spontaneous L2 production.

The preliminary data screening showed that the assumption of normality of the three groups' pre-test scores on each of the five dependent variables was violated

(Shapiro-Wilk, $p < .05$). Furthermore, from the visual inspection of histograms it was clear that the three groups' pre-test scores were positively skewed. This is due to the fact that many participants in the three groups produced no or very few ECSs as they completed the pre-test task (see also Section 7.2, below). Square root transformation was performed to correct the data, but this did not solve the problem.

With regard to the three groups' scores in the post-test and delayed post-test on both the same and parallel tasks, the screening revealed that the scores of the first four of the five dependent variables (i.e., (i) – (iv)) were normally distributed for the two experimental (i.e., pre-modified input and interactional feedback) groups (Shapiro-Wilk, $p > .05$), but not for the comparison group (Shapiro-Wilk, $p < .05$). That is, the assumption of normality of the comparison group's post-test and delayed post-test scores on all tests consisting of the same and parallel tasks was violated. The visual inspection of histograms also revealed that the comparison group's post-test and delayed post-test scores were positively skewed. Square root transformation was again conducted to correct the normality of these scores, but this did not solve the problem. Therefore, it was decided that the data would be analyzed as they were (i.e., without transformation) using a series of non-parametric statistics.

A number of non-parametric analyses, however, may not produce accurate results if the assumption of homogeneity of variance are not met (Kraska-Miller, 2014). Therefore, a series of Levene's tests were performed on the three groups' pre-test, post-test and delayed post-test scores of each of the five dependent variables to confirm their homogeneity of variance. The results showed that the assumption of homogeneity of variance of the tests was met for each of the five dependent

variables (Levene's test, $p > .05$). This suggests, therefore, that non-parametric analyses were appropriate to use for analyzing the data in this study.

In what follows, descriptive statistics for the three groups' scores on each of the five dependent variables are first reported. As regards the results of non-parametric analyses of both within- and between-group measures, only the scores on the first four of the five dependent variables are reported: (i) – (iv). With respect to the fifth dependent variable, that is, (v) the frequency of combined self-corrections, replacements and reformulations per second as learners attempted to use ECSs, only descriptive statistics of the three groups' scores are presented. This is due to a very low number of occurrences of this dependent variable in the participants' production of ECSs in the three test sessions. That is, the participants in the three groups did not engage in enough repair (or monitoring) on any of the tests to analyze. Chapter 8 discusses why this might be the case.

For all tests of statistical significance, the alpha level was set at .05. To test the null hypothesis for *within*-subject measures, Friedman tests with post-hoc Wilcoxon signed-rank tests followed by Bonferroni corrections for a number of comparisons were performed. To test the null hypothesis for *between*-subject measures, Kruskal-Wallis H tests with post-hoc Mann-Whitney U tests, also followed by Bonferroni corrections for a number of comparisons, were carried out. In addition to these non-parametric analyses, effect sizes (r) were also calculated to identify the magnitude of difference for both within- and between-group measures. Following Plonsky and Oswald (2014), the effect sizes are interpreted as either small ($r = .25$), medium ($r = .40$) or large ($r = .60$) (cf., Cohen, 1988).

Table 7.1: Descriptive statistics for the production of ECSs*

Session	Task Set	Descriptive Statistics	Group		
			PI (N = 17)	IF (N = 17)	C (N = 17)
Pre-test	Task Set 1	Mean	0.15	0.14	0.06
		95% CI (lower)	0.07	0.08	-0.01
		95% CI (upper)	0.22	0.20	0.14
		Minimum	0.00	0.00	0.00
		Maximum	0.50	0.33	0.60
		Median	0.11	0.11	0.00
		SD	0.14	0.11	0.15
		Range	0.50	0.33	0.60
Post-test	Task Set 1	Mean	0.34	0.42	0.09
		95% CI (lower)	0.24	0.32	0.03
		95% CI (upper)	0.44	0.53	0.15
		Minimum	0.00	0.07	0.00
		Maximum	0.67	0.75	0.33
		Median	0.29	0.40	0.00
		SD	0.20	0.20	0.12
		Range	0.67	0.68	0.33
	Task Set 2	Mean	0.42	0.41	0.11
		95% CI (lower)	0.31	0.28	0.05
		95% CI (upper)	0.52	0.53	0.18
		Minimum	0.00	0.00	0.00
		Maximum	0.75	1.00	0.38
		Median	0.43	0.43	0.11
		SD	0.20	0.24	0.13
		Range	0.75	1.00	0.38
Delayed Post-test	Task Set 1	Mean	0.29	0.36	0.09
		95% CI (lower)	0.17	0.24	0.03
		95% CI (upper)	0.40	0.47	0.15
		Minimum	0.00	0.00	0.00
		Maximum	0.75	0.75	0.33
		Median	0.25	0.33	0.00
		SD	0.23	0.23	0.12
		Range	0.75	0.75	0.33
	Task Set 3	Mean	0.27	0.22	0.05
		95% CI (lower)	0.20	0.13	0.02
		95% CI (upper)	0.33	0.30	0.08
		Minimum	0.08	0.00	0.00
		Maximum	0.50	0.67	0.18
		Median	0.25	0.21	0.00
		SD	0.13	0.16	0.06
		Range	0.42	0.67	0.18

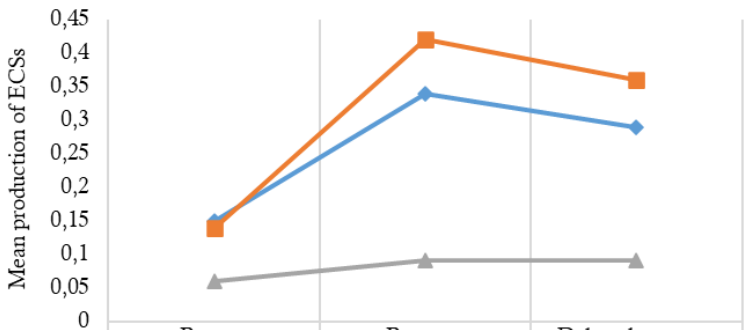
Notes. *Values in the table represents the proportion of AS-units containing the use of ECS.
PI = pre-modified input; IF = interactional feedback; C = comparison

7.2 Proportion of ECSs Produced

Table 7.1, above, shows the descriptive statistics for the proportion of ECSs produced by each of the three groups. The maximum possible score for this test

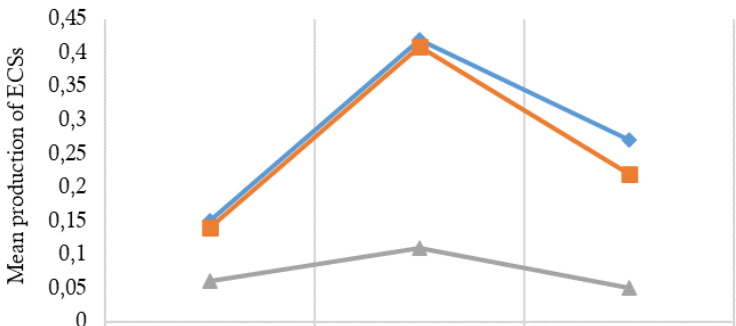
was 1. As discussed in Chapter 5, Section 5.5, this measure represents the ability to access or recall knowledge of the structures to address functional demands of the tasks. Values in Table 7.1 thus reflect the accessibility of the knowledge of the structures before and after the three groups received respective treatments.

As shown in Table 7.1, the scores of the experimental (i.e., PI and IF) groups on the use of ECSs improved from the pre-test to post-test on both the same and parallel tasks. Nonetheless, their scores decreased in the delayed post-test. Despite this, the groups' delayed post-test scores remained high in comparison to those in the pre-test. In contrast, the comparison group performed similarly across the tests. That is, the comparison group's scores on the use of ECSs from the pre-test to delayed post-test were comparable. Figures 7.1 and 7.2 present graphs of the mean scores for each group's use of ECSs on the same task (i.e., Task Set 1-1-1) and on the parallel version of the task (i.e., Task Set 1-2-3), respectively, in the pre-test, post-test and delayed post-test.



	Pre-test	Post-test	Delayed post-test
Pre-modified Input	0,15	0,34	0,29
Interaction Feedback	0,14	0,42	0,36
Comparison	0,06	0,09	0,09

Figure 7.1: Proportion of the production of ECSs on the same task.



	Pre-test	Post-test	Delayed post-test
Pre-modified Input	0,15	0,42	0,27
Interaction Feedback	0,14	0,41	0,22
Comparison	0,06	0,11	0,05

Figure 7.2: Proportion of the production of ECSs on parallel tasks.

7.2.1 Within-Group Differences for the Production of ECSs

For *the same* tasks (see Figure 7.1), Friedman tests revealed that there were statistically significant differences for the two experimental (PI and IF) groups' scores of the use of ECSs between the three repeated tests: for the PI group, $\chi^2(2) = 12.862$, $p = .002$; for the IF group, $\chi^2(2) = 16.092$, $p = .000$. In contrast, no significant difference was found between the three repeated tests for the comparison group, $\chi^2(2) = 1.163$, $p = .559$ (see Table 7.2, Task Set 1-1-1).

Table 7.2: Friedman tests for the production of ECSs

Task Set	**Group**					
	PI (N = 17)		**IF (N = 17)**		**comparison (N = 17)**	
	Chi-square (χ^2)	**Asymp. Sig. (*p*)**	**Chi-square (χ^2)**	**Asymp. Sig. (*p*)**	**Chi-square (χ^2)**	**Asymp. Sig. (*p*)**
1 – 1 – 1	12.862	.002*	16.092	.000*	1.163	.559
1 – 2 – 3	17.853	.000*	15.647	.000*	3.912	.141

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

Post-hoc Wilcoxon Signed Rank Tests followed by Bonferroni corrections for the number of comparisons were then performed to further examine the experimental

groups' scores on the same task (see Table 7.3, Task Set 1 – 1 – 1). These post-hoc tests revealed that the pre-test and post-test scores of the experimental groups were significantly different on this task: for the PI group, z = -3.352, p = .001; for the IF group, z = -3.352, p = .001. Furthermore, the magnitude of the differences was large in both groups: the PI group, r = .609, the IF group, r = .589. This suggests, therefore, that the improvement in the use of ECSs by these two experimental groups on the same task from the pre-test to post-test was both significant and meaningful.

Statistical comparisons of the experimental groups' post-test and delayed post-test scores further revealed that the improvement was stable, as indicated by no significant differences between the groups' post-test and delayed post-test scores: for the PI group, z = -1.256, p = .209; for the IF group, z = -0.982, p = .326. Indeed, the magnitude of the differences was either very small or close to null in these groups: for the PI group, r = .215; for the IF group, r = .168. Likewise, statistical comparisons of these groups' pre-test and delayed post-test scores revealed significant differences: for the PI group, z = -2.668, p = .008; for the IF group, z = -3.432, p = .001. Nonetheless, the magnitude of the difference between the pre-test and delayed post-test scores decreased in the PI group, but not in IF group. That is, while the magnitude of the difference of the IF group's pre-test and delayed post-test scores remained large, r = .577, the magnitude of the difference of the PI group's pre-test and delayed post-test scores turned into medium, r = .458. This suggests that although the significant improvement in the production of ECSs made by the PI and IF group on the same task was stable over time, the improvement made by the IF group appeared to be more stable than that made by the PI group.

Table 7.3: Wilcoxon Signed Ranks tests for the production of ECSs

Group	Task Set	Post-hoc Wilcoxon Signed Ranks Tests			Effect size (*r*)
		Session	Z score	Asymp. Sig. (*p*)	
PI (N = 17)	1-1-1	Pre- to Post-test	-3.352	.001*	.609
		Pre- to Delayed Post-test	-2.668	.008*	.458
		Post- to Delayed Post-test	-1.256	.209	.215
	1-2-3	Pre- to Post-test	-3.551	.001*	.575
		Pre- to Delayed Post-test	-1.979	.048	.339
		Post- to Delayed Post-test	-2.486	.013	.426
IF (N = 17)	1-1-1	Pre- to Post-test	-3.352	.001*	.589
		Pre- to Delayed Post-test	-3.432	.001*	.577
		Post- to Delayed Post-test	-0.982	.326	.168
	1-2-3	Pre- to Post-test	-3.352	.001*	.575
		Pre- to Delayed Post-test	-1.773	.083	.297
		Post- to Delayed Post-test	-2.471	.013	.424
comparison (N = 17)	1-1-1	Pre- to Post-test	-0.534	.594	.092
		Pre- to Delayed Post-test	-0.942	.346	.162
		Post- to Delayed Post-test	-0.051	.959	.009
	1-2-3	Pre- to Post-test	-1.377	.169	.236
		Pre- to Delayed Post-test	-0.507	.612	.087
		Post- to Delayed Post-test	-2.002	.045	.343

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

Post-hoc analyses with Wilcoxon Signed Rank Tests followed by Bonferroni corrections for the number of comparisons were also performed to examine the comparison group's scores on the same task (see Table 7.3, Task Set 1 – 1 – 1). The analyses confirmed that the comparison group did not make any significant improvement in the production of ECSs throughout the three repeated tests: from the pre-test to post-test, z = -0.534, *p* = .594; from the post-test to delayed post-test, z = -0.051, *p* = .959; and from the pre-test to delayed post-test, z = -0.942, *p* = .346.

Furthermore, the calculations for the effect sizes revealed that the magnitude of the differences was either very small or null: between the pre-test and post-test, r = .092; between the post-test and delayed post-test, r = .009, and between the pre-test and delayed post-test, r = .162. This, therefore, suggests that task performance with a focus on meaning (FonM) that the comparison group received during the treatments appeared to be ineffective to promote uptake (increased production) of specific task-relevant L2 forms.

As regards *the parallel version* of the tasks (see Figure 7.2, above), Friedman tests revealed that there were statistically significant differences for the two experimental (PI and IF) groups between the three repeated tests: for the PI group, $\chi^2(2) = 17.853$, $p = .000$; for the IF group, $\chi^2(2) = 15.647$, $p = .000$. In contrast, no significant difference was found for the comparison group between the three repeated tests, $\chi^2(2) = 3.912$, $p = .141$ (see Table 7.2, Task Set 1-2-3).

Post-hoc analyses with Wilcoxon Signed Rank Tests followed by Bonferroni corrections (see Table 7.3, Task Set 1 – 2 – 3) for the experimental groups revealed that the pre-test and post-test scores of these groups were significantly different: for the PI group, z = -3.551, $p = .001$; for the IF group, z = -3.352, $p = .001$, with the magnitude of the differences that was large in these two esperimental groups: for the PI group, r = .575; for the IF group, r = .575. This suggests, therefore, that the improvement in the use of ECSs made by these two experimental groups on the parallel version of the task in the post-test was both significant and meaningful.

Notwithstanding, the gains in the use of ECSs made by these two experimental groups decreased significantly in the delayed post-test, as reflected in no significant differences in the use of ECSs in the pre-test and delayed post-test: for the PI

group, z = -1.979, *p* = .048, for the IF group, z = -1.773, *p* = .083. The magnitude of difference in the use of ECSs in the pre-test and delayed post-test was also small: for the PI group, *r* = .339; for the IF group *r* = .297. These significant decreases were also confirmed by significant differences in these two groups' post-test and delayed post-test scores: for the PI group, z = -2.486, *p* = .013; for the IF group, z = -2.471, *p* = .013, with the magnitude of the differences that was medium in sizes: for the PI group, *r* = .426; for the IF group *r* = .424. This, therefore, suggests that although gains in the use of ECSs appeared to transfer to parallel tasks for these two experimental (PI and IF) groups, the transfer appeared to be less stable over time (in the sense that these experimental groups were less able to use the structures with the same level of frequency to address the functional demands of the parallel task in the delayed post-test, relative to that in the post-test).

As regards the comparison group's scores of the use of ECSs on the parallel version of the tasks, Friedman tests indicated that there were no statistically significant differences in the three repeated tests, $\chi^2(2) = 1.163$, *p* = .559. Post-hoc analyses with Wilcoxon Signed Rank Tests followed by Bonferroni corrections for the number of comparisons (see Table 7.3, Task Set 1 – 2 – 3) also indicated that the comparison group did not make any significant improvement in their use of ECSs from the pre-test to post-test, z = -1.377, *p* = .169; from the post-test to delayed post-test, z = -2.002, *p* = .045; and from the pre-test to delayed post-test, z = -0.507, *p* = .612. The magnitude of the differences was also either small or null in sizes: between the pre-test and post-test, *r* = .236; between the post-test and delayed post-test, *r* = .343; and between the pre-test and delayed post-test, *r* = .087. This suggests that task performance with a focus on meaning (FonM) was, again,

ineffective to promote uptake of specific task-relevant L2 forms on parallel versions of a task.

7.2.2 Between-Group Differences for the Production of ECSs

Between-group differences for the production of ECSs were investigated by conducting Kruskal-Wallis H tests (see Table 7.4, below). The analyses showed that there was already a significant difference between groups in the pre-test, $\chi^2(2) = 8.400$, $p = .015$, and that the difference remained constant in the post-test on both the same task, $\chi^2(2) = 21.932$, $p = .001$, and the parallel version of the task, $\chi^2(2) = 19.438$, $p = .001$. The difference was also noted in the delayed post-test on both the same task, $\chi^2(2)$ 13.056, $p = .001$, and the parallel version of the task, $\chi^2(2) = 22.556$, $p = .001$.

Post-hoc Mann-Whitney U tests followed by Bonferroni corrections were performed to further examine the groups' *pre-test* scores (see Table 7.5). The results revealed that the experimental (PI and IF) groups started off using ECSs more than the comparison group on the task in the pre-test: PI vs. comparison, $U = 96.500$, $p = .013$; IF vs. comparison, $U = 70.000$, $p = .007$. The calculations for the effect sizes further revealed that these differences were medium in magnitude: PI vs. comparison, $r = .410$; IF vs. comparison, $r = .465$. No significant difference was found between the experimental groups in terms of the proportion of ECSs that they produced in the pre-test: PI vs. IF, $U = 141.000$, $p = .903$. Indeed, the magnitude of the difference was very close to null: PI vs. IF, $r = .021$.

Table 7.4: Kruskal-Wallis H tests for the production of ECSs

Session	Task Set	Group	Mean Rank	Kruskal-Wallis H Tests	
				Chi-square (χ^2)	Asymp. Sig. (*p*)
Pre-test	Task Set 1	PI (N = 17)	29.618	8.400	.015
		IF (N = 17)	30.588		
		C (N = 17)	17.794		
Post-test	Task Set 1	PI (N = 17)	30.000	21.932	.001*
		IF (N = 17)	35.353		
		C (N = 17)	12.647		
	Task Set 2	PI (N = 17)	33.118	19.438	.001*
		IF (N = 17)	31.765		
		C (N = 17)	13.118		
Delayed Post-test	Task Set 1	PI (N = 17)	28.794	13.056	.001*
		IF (N = 17)	33.382		
		C (N = 17)	15.824		
	Task Set 3	PI (N = 17)	35.618	22.556	.001*
		IF (N = 17)	29.824		
		C (N = 17)	12.559		

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

In the *post-test* session, the analyses showed that the scores of the experimental (PI and IF) groups for the use of ECSs remained significantly different from those of the comparison group on both the same task: PI vs. comparison, $U = 43.000$, $p = .001$; IF vs. comparison, $U = 19.000$, $p = .001$, and the parallel version of the task: PI vs. comparison, $U = 30.500$, $p = .001$; IF vs. comparison, $U = 39.500$, $p = .001$. Nonetheless, while the differences between the experimental and comparison groups' scores in the pre-test were medium in magnitude, the differences in the post-test turned into large in magnitude on both the same task: PI vs. comparison, $r = .611$; IF vs. comparison, $r = .749$, and the parallel version of the task: PI vs.

comparison, r = .680; IF vs. comparison, r = .629. No significant difference was found between the two experimental groups on the same task, U = 144.500, p = .981, as well as on the parallel version of the task, U = 137.500, p = .809. The magnitude of the differences of these groups (PI and IF) also remained very close to null on the same task, r = .002, and on the parallel version of the task, r = .041. This, therefore, suggests that the provision of pre-modified input or interactional feedback in task-based L2 instruction is effective to promore uptake (more frequent use) and the transferability of uptake of task-relevant L2 forms. In contrast, task performance with a focus on meaning (FonM) appeared to be ineffective to promote gains in the uptake of task-relevant L2 forms.

Finally, in the *delayed post-test* the analyses generated similar results. That is, the gains of the use of ECSs by the two experimental groups were significantly different from those of the comparison group on both the same task: PI vs. comparison, U = 73.500, p = .012; IF vs. comparison, U = 42.500, p = .001, and the parallel version of the task: PI vs. comparison, U = 16.500, p = .001; IF vs. comparison, U = 44.000, p = .001. The calculations for the effect sizes also showed that the magnitude of the differences between these experimental and comparison groups' scores remained large on the same task: PI vs. comparison, r = .612; IF vs. comparison, r = .611, and on the parallel version of the task, PI vs. comparison, r = .612; IF vs. comparison, r = .612. No significant differences were found between the experimental groups' scores on both the same task, PI vs. IF, U = 121.000, p = .417, and the parallel version of the task, PI vs. IF, U = 109.000, p = .220. The calculations for the effect sizes also revealed that the magnitude of the differences remained close to null on the same task, r = .139, and was very small on the parallel version of the task, r = .210.

Table 7.5: Mann-Whitney U tests for the production of ECSs

Session	Task Set	Post-hoc Mann-Whitney U Tests			Effect size (*r*)
		Group	*U* statistic	Asymp. Sig. (*p*)	
Pre-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	141.000	.903	.021
		PI (N = 17) vs. C (N = 17)	96.500	.013	.410
		IF (N = 17) vs. C (N = 17)	70.000	.007*	.465
Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	144.500	.981	.002
		PI (N = 17) vs. C (N = 17)	43.000	.001*	.611
		IF (N = 17) vs. C (N = 17)	19.000	.001*	.749
	Task Set 2	PI (N = 17) vs. IF (N = 17)	137.500	.809	.041
		PI (N = 17) vs. C (N = 17)	30.500	.001*	.680
		IF (N = 17) vs. C (N = 17)	39.500	.001*	.629
Delayed Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	121.000	.417	.139
		PI (N = 17) vs. C (N = 17)	73.500	.012	.622
		IF (N = 17) vs. C (N = 17)	42.500	.001*	.611
	Task Set 3	PI (N = 17) vs. IF (N = 17)	109.000	.220	.210
		PI (N = 17) vs. C (N = 17)	16.500	.001*	.612
		IF (N = 17) vs. C (N = 17)	44.000	.001*	.612

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

In summary, the results indicate that task-based L2 instruction which incorporated FonF—operationalized through pre-modified input (PI) or interactional feedback (IF)—was effective in promoting uptake and the transferability of uptake of specific task-relevant L2 forms (ECSs). In contrast, task-based L2 instruction which solely focused on meaning (FonM) was ineffective in promoting both uptake and the transferability of uptake of specific task-relevant L2 forms. Furthermore, there was evidence that gains in the uptake and the transferability of uptake of task-relevant L2 forms that the PI and IF conditions promoted were stable. Nonetheless,

it appears that the gains promoted by the IF condition were slightly better than those promoted by the PI condition.

Table 7.6: Descriptive statistics for the proportion of error-free uses of ECSs*

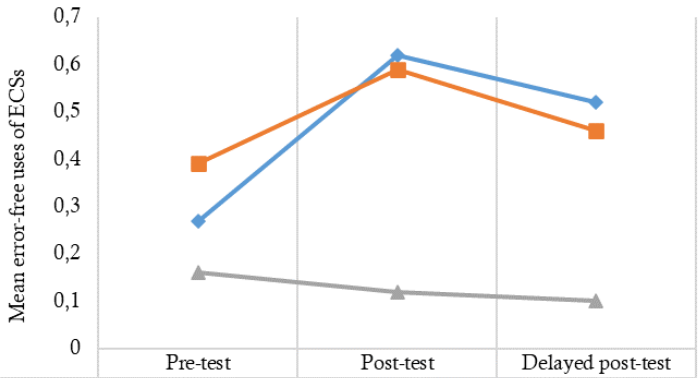
Session	Task Set	Descriptive Statistics	Group		
			PI (N = 17)	IF (N = 17)	C (N = 17)
Pre-test	Task Set 1	Mean	0.27	0.39	0.16
		95% CI (lower)	0.04	0.16	-0.03
		95% CI (upper)	0.49	0.63	0.34
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.00	0.00	0.00
		SD	0.44	0.46	0.36
		Range	1.00	1.00	1.00
Post-test	Task Set 1	Mean	0.62	0.59	0.12
		95% CI (lower)	0.44	0.39	-0.05
		95% CI (upper)	0.80	0.79	0.29
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.67	0.67	0.00
		SD	0.35	0.39	0.33
		Range	1.00	1.00	1.00
	Task Set 2	Mean	0.59	0.61	0.26
		95% CI (lower)	0.40	0.42	0.01
		95% CI (upper)	0.79	0.80	0.46
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.67	0.67	0.00
		SD	0.38	0.36	0.44
		Range	1.00	1.00	1.00
Delayed Post-test	Task Set 1	Mean	0.52	0.46	0.10
		95% CI (lower)	0.31	0.24	-0.07
		95% CI (upper)	0.72	0.68	0.18
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.67	0.50	0.00
		SD	0.39	0.43	0.24
		Range	1.00	1.00	1.00
	Task Set 3	Mean	0.62	0.53	0.18
		95% CI (lower)	0.43	0.32	-0.03
		95% CI (upper)	0.81	0.74	0.38
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.60	0.50	0.00
		SD	0.37	0.42	0.39
		Range	1.00	1.00	1.00

Notes. *Values in the table represents the proportion of AS-Units produced by learners containing accurate use of ECSs.
PI = pre-modified input, IF = interactional feedback, C = comparison

7.3 Proportion of Error-Free Uses of ECSs

The descriptive statistics for the proportion of error-free uses of ECSs which participants produced when performing the same and parallel version of the tasks is provided in Table 7.6, above. The maximum possible score for this measure was 1. As discussed in Chapter 5, Section 5.5, this measure (error-free uses of ECSs) represents the participants' knowledge of the structures. Thus, values in Table 7.6 reflect the knowledge of the structures before and after the participants in each of the three groups received respective treatments.

As Table 7.6 indicates, the scores of error-free uses of ECSs that the two experimental (PI and IF) groups made increased quite substantially in the post-test on both the same and parallel version of the tasks. Furthermore, these gains remained stable on the delayed post-test. By contrast, the scores of error-free uses of ECSs that the comparison group made were comparable across the three repeated tests on both the same and parallel version of the tasks. Figures 7.1 and 7.2 present graphs of the mean scores of error-free uses of ECSs that each of the three groups made on the same task (i.e., Task Set 1-1-1) and on parallel tasks (i.e., Task Set 1-2-3), respectively, in the pre-test, post-test and delayed post-test.



	Pre-test	Post-test	Delayed post-test
Pre-modified Input	0,27	0,62	0,52
Interaction Feedback	0,39	0,59	0,46
Comparison	0,16	0,12	0,1

Figure 7.3: Proportion of error-free uses of ECSs on the same task

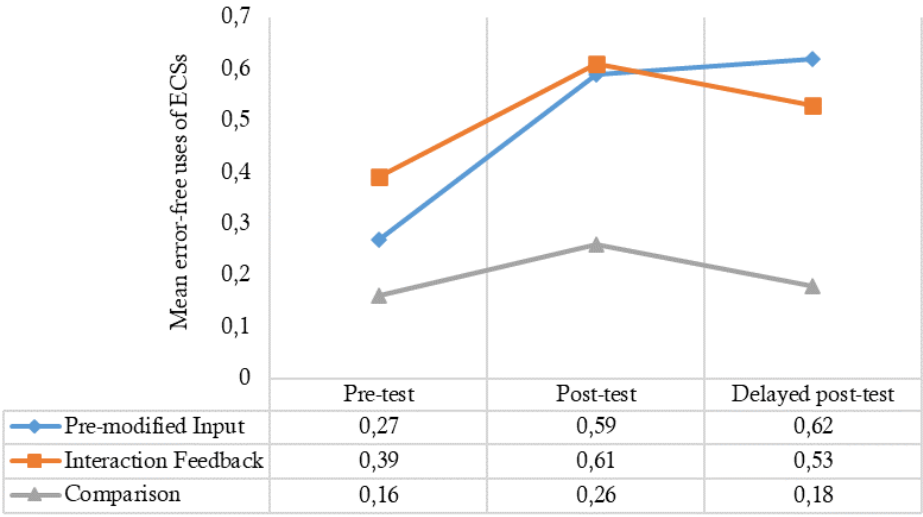


Figure 7.4: Proportion of error-free uses of ECSs on parallel tasks

7.3.1 Within-Group Differences for Error-Free Uses of ECSs

Although the descriptive statistics show substantial increases in the mean scores of error-free uses of ECSs by the two experimental (PI and IF) groups between the pre-test and the post-tests on both the same and parallel version of the tasks (see Table 7.6; see also Figures 7.3 and 7.4), Friedman tests carried out to examine the three (PI, IF, and comparison) groups' pre-test, post-test and delayed post-test scores revealed no significant differences on the same and parallel version of the tasks (see Table 7.7).

Table 7.7: Friedman tests for the proportion of error-free uses of ECSs

Task Set	Group					
	PI (N = 17)		IF (N = 17)		C (N = 17)	
	Chi-square (χ^2)	Asymp. Sig. (*p*)	Chi-square (χ^2)	Asymp. Sig. (*p*)	Chi-square (χ^2)	Asymp. Sig. (*p*)
1-1-1	5.061	.080	1.529	.465	1.077	.584
1-2-3	3.206	.201	3.261	.196	0.118	.983

Notes. PI = pre-modified input, IF = interactional feedback, C = comparison.

Table 7.8: Wilcoxon Signed Ranks tests for the proportion of error-free uses of ECSs

Group	Task Set	Post-hoc Wilcoxon Signed Ranks Tests			Effect size (*r*)
		Session	Z score	Asymp. Sig. (*p*)	
PI (N = 17)	1-1-1	Pre- to Post-test	-2.667	.008*	.457
		Pre- to Delayed Post-test	-1.578	.115	.271
		Post- to Delayed Post-test	-1.304	.192	.224
	1-2-3	Pre- to Post-test	-2.446	.014	.419
		Pre- to Delayed Post-test	-2.453	.014	.421
		Post- to Delayed Post-test	-0.175	.861	.030
IF (N = 17)	1-1-1	Pre- to Post-test	-1.854	.064	.318
		Pre- to Delayed Post-test	-1.078	.281	.185
		Post- to Delayed Post-test	-0.874	.382	.150
	1-2-3	Pre- to Post-test	-2.446	.014*	.419
		Pre- to Delayed Post-test	-2.453	.014*	.421
		Post- to Delayed Post-test	-0.175	.861	.030
C (N = 17)	1-1-1	Pre- to Post-test	-0.447	.655	.077
		Pre- to Delayed Post-test	-0.756	.450	.130
		Post- to Delayed Post-test	-0.577	.564	.099
	1-2-3	Pre- to Post-test	-1378	.168	.236
		Pre- to Delayed Post-test	-0.893	.372	.153
		Post- to Delayed Post-test	-1.454	.146	.249

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

Nonetheless, post-hoc analyses with Wilcoxon Signed Rank Tests followed by Bonferroni corrections (see Table 7.8) revealed that there were statistically significant differences in the PI groups' pre-test and post-test accuracy scores on the same task, z = -2.667, p = .008, as well as on the parallel version of the task, z = -2.446, p = .014. The calculations for the effect sizes also revealed that the differences were medium in magnitude on both the same task, r = .457, and on the parallel version of the task, r = .415. However, further analyses of the PI group's

pre-test and delayed post-test scores on the same task revealed no significant difference, z = -1.578, *p* = .115, and that the magnitude of the difference turned into small, *r* = .271. Despite this, the group's scores on the parallel version of the task in the delayed post-test remained significantly different from those in the pre-test, z = -2.453, *p* = .014, and that the magnitude of the difference was medium in size, *r* = .421. This, therefore, suggests that the effects of the PI condition on learner error-free uses of the target form were largely transferable and, at least, partially stable.

Similar results were found in the IF group, where the post-hoc analyses with Wilcoxon Signed Rank Tests followed by Bonferroni corrections revealed that there was a significant difference in this group's pre-test and post-test scores on the parallel version of the task, z = -2.446, *p* = .014, but not on the same task, z = -1.854, *p* = .064. The calculations for the effect sizes also revealed that the magnitude of the differences of the IF group's pre-test and post-test scores was medium on the parallel task, *r* = .419, but was small on the same task, *r* = .318. The same results were found with respect to the IF group's scores in the delayed post-test, in that while the IF group's delayed post-test scores of error-free uses of ECSs on the parallel version of the task were significantly different from their pre-test scores, z = -2.453, *p* = .014, no significant difference was found with respect to their pre-test and delayed post-test scores on the same task, z = -1.078, *p* = .281. The calculations for the effect sizes also revealed that this group's pre-test and delayed post-test scores were medium on the parallel task, *r* = .421, but were close to null on the same task, *r* = .185. This suggests that the effects of the interactional feedback on error-free uses of ECSs were transferable only to parallel tasks.

Post-hoc analyses with Wilcoxon Signed Rank Tests followed by Bonferroni corrections were also performed on the comparison group's pre-test, post-test,

delayed post-test scores on both the same and parallel version of the tasks. The results confirmed that there were no statistically significant differences for error-free uses of ECSs in this group's scores throughout the three repeated tests. The calculations for the effect sizes also revealed that the magnitude of the differences of this group's scores of error-free uses between (i) the pre-test and post-test; (ii) post-test and delayed post-test; and (iii) pre-test and delayed post-test on both the same and parallel version of the tasks were either small or very close to null (see Table 7.8). This suggests that task-based L2 instruction which solely focused on meaning (FonM) that the participants in the comparison group received was ineffective in promoting gains in error-free uses of the target L2 form (ECSs).

7.3.2 Between-Group Differences for Error-Free Uses of ECSs

With respect to between-group measures, the Kruskal-Wallis H tests (see Table 7.9, below) revealed that there was no a statistically significant difference in the error-free uses of ECSs between groups in the pre-test, $\chi^2(2) = 2.986$, $p = .225$. However, there was in the post-test on the same task, $\chi^2(2) = 15.300$, $p = .000$, but not on the parallel version of the task, $\chi^2(2) = 7.806$, $p = .020$. In the delayed post-test, a significant difference occurred on both the same task, $\chi^2(2) = 12.898$, $p = .002$, and the parallel version of the task, $\chi^2(2) = 10.326$, $p = .006$.

Post-hoc analyses with Mann-Whitney U tests followed by Bonferroni corrections performed to further examine the groups' *pre-test* scores (see Table 7.10) confirmed that none of the differences between the experimental (PI and IF) and comparison groups reached statistical significance: PI vs. comparison, $U = 127.000$, $p = .415$; IF vs. comparison, $U = 103.500$, $p = .087$; PI vs. IF, $U = 122.000$, $p = .370$. The calculations for the effect sizes of the groups' pre-test

scores also revealed that the magnitude of the differences was either close to null or small: PI vs. comparison, r = .140; PI vs. IF, r = .154; IF vs. comparison, r = .293.

Table 7.9: Kruskal-Wallis H tests for the proportion of error-free uses of ECSs

Session	Task Set	Group	Mean Rank	Kruskal-Wallis H Tests	
				Chi-square (χ^2)	Asymp. Sig. (p)
Pre-test	Task Set 1	PI (N = 17)	25.706	2.986	.225
		IF (N = 17)	29.735		
		C (N = 17)	22.559		
Post-test	Task Set 1	PI (N = 17)	31.618	15.300	.000*
		IF (N = 17)	31.382		
		C (N = 17)	15.000		
	Task Set 2	PI (N = 17)	29.294	7.806	.020
		IF (N = 17)	30.588		
		C (N = 17)	18.118		
Delayed Post-test	Task Set 1	PI (N = 17)	31.794	12.898	.002*
		IF (N = 17)	29.765		
		C (N = 17)	16.441		
	Task Set 3	PI (N = 17)	31.882	10.326	.006*
		IF (N = 17)	28.882		
		C (N = 17)	17.235		

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

In the *post-test* session, the analyses showed that the gains made by the experimental groups in comparison to those made by the comparison group differed. That is, the PI group outperformed the comparison group only in terms of their error-free uses of ECSs on the same task, U = 51.500, p = .000, with the magnitude of the difference that was large, r = .598, whereas their scores on the parallel version of the task did not reach statistical significance, U = 84.000, p = .024. Thus, the PI group did not outperform the comparison group for error-free

uses of ECSs on the parallel task. Indeed, the magnitude of the difference between these groups on this dependent measure remained small, $r = .386$. In contrast, the IF groups outperformed the comparison group on both the same task, $U = 50.500$, $p = .000$, and the parallel version of the task, $U = 71.500$, $p = .007$, with the magnitude of the differences that was large for the same task, $r = .606$, and medium for the parallel version of the task, $r = .459$. No significant difference was found between the experimental (PI and IF) groups' scores of error-free uses of ECSs on either the same task, $U = 142.000$, $p = .930$, or on the parallel version of the task, $U = 140.000$, $p = .875$. The calculations for the effect sizes also revealed that the differences of these experimental groups' scores on these tasks were close to null in magnitude: for both the same task, $r = .015$; for the parallel version of the task, $r = .027$.

Finally, in the *delayed post-test* the two experimental groups outperformed the comparison group for error-free uses of ECSs on the same task: PI vs. comparison, $U = 55.500$, $p = .001$; IF vs. comparison, $U = 71.000$, $p = .002$, as well as on the parallel version of the task: PI vs. comparison, $U = 63.000$, $p = .002$; IF vs. comparison, $U = 77.000$, $p = .010$. The calculations for the effect sizes also showed that the differences between the PI and comparison groups were large on the same task, $r = .602$, and medium on the parallel version of the task, $r = .520$. At the same time, the differences between the IF and comparison groups were medium on the same task, $r = .524$, and on the parallel version of the task, $r = .459$. No significant difference was found between the experimental (PI and IF) groups' scores of error-free uses of ECSs on the same task, $U = 135.000$, $p = .337$, and on the parallel version of the task, $U = 126.000$, $p = .510$. Indeed, the magnitude of the differences

remained close to null on these tasks: for the same task, r = .058; for the parallel version of the task, r = .113.

Table 7.10: Mann-Whitney U tests for the proportion of error-free uses of ECSs

Session	Task Set	Post-hoc Mann-Whitney U Tests			Effect size (r)
		Group	*U* statistic	Asymp. Sig. (p)	
Pre-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	122.000	.370	.154
		PI (N = 17) vs. C (N = 17)	127.000	.415	.140
		IF (N = 17) vs. C (N = 17)	103.500	.087	.293
Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	142.000	.930	.015
		PI (N = 17) vs. C (N = 17)	51.500	.000*	.598
		IF (N = 17) vs. C (N = 17)	50.500	.000*	.606
	Task Set 2	PI (N = 17) vs. IF (N = 17)	140.000	.875	.027
		PI (N = 17) vs. C (N = 17)	84.000	.024	.386
		IF (N = 17) vs. C (N = 17)	71.500	.007*	.459
Delayed Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	135.000	.337	.058
		PI (N = 17) vs. C (N = 17)	55.500	.001*	.602
		IF (N = 17) vs. C (N = 17)	71.000	.002*	.524
	Task Set 3	PI (N = 17) vs. IF (N = 17)	126.000	.510	.113
		PI (N = 17) vs. C (N = 17)	63.000	.002*	.520
		IF (N = 17) vs. C (N = 17)	77.000	.010	.459

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

In sum, the statistical analyses for between-group comparisons revealed that the PI group outperformed the comparison group in the error-free uses of ECSs on the same and parallel version of the task only in the delayed post-test, whereas the IF group outperformed the comparison group on this measure in both the post-test and delayed post-test. That is, gains in the accurate use of ECSs appeared to be slightly

better in the IF group than the PI in the immediate post-test, but not in the delayed post-test.

Table 7.11: Descriptive statistics for the speech rate of ECSs*

Session	Task Set	Descriptive Statistics	Group		
			PI (N = 17)	IF (N = 17)	Control (N = 17)
Pre-test	1	Mean	0.83	1.10	0.25
		95% CI (lower)	0.44	0.64	0.04
		95% CI (upper)	1.22	1.55	0.47
		Minimum	0.00	0.00	0.00
		Maximum	2.41	2.53	1.03
		Median	0.68	1.10	0.00
		SD	0.76	0.88	0.41
		Range	2.41	2.53	1.03
Post-test	1	Mean	1.38	1.39	0.49
		95% CI (lower)	1.00	1.42	0.16
		95% CI (upper)	1.76	1.63	0.82
		Minimum	0.00	0.57	0.00
		Maximum	2.77	2.45	1.89
		Median	1.35	1.37	0.00
		SD	0.73	0.48	0.65
		Range	2.77	1.88	1.89
	2	Mean	1.16	1.29	0.43
		95% CI (lower)	0.90	0.84	0.18
		95% CI (upper)	1.43	1.72	0.69
		Minimum	0.00	0.00	0.00
		Maximum	1.91	2.83	1.63
		Median	1.24	1.13	0.39
		SD	0.52	0.86	0.49
		Range	1.91	2.83	1.63
Delayed Post-test	1	Mean	1.40	1.28	0.61
		95% CI (lower)	0.86	0.93	0.13
		95% CI (upper)	1.93	1.62	1.09
		Minimum	0.00	0.00	0.00
		Maximum	2.96	2.77	3.36
		Median	1.49	1.32	0.00
		SD	1.04	0.67	0.93
		Range	2.96	2.77	3.36
	3	Mean	1.35	1.29	0.41
		95% CI (lower)	1.04	0.62	0.09
		95% CI (upper)	1.65	1.96	0.73
		Minimum	0.54	0.00	0.00
		Maximum	2.90	5.58	1.77
		Median	1.19	1.21	0.00
		SD	0.59	1.30	0.63
		Range	2.36	5.58	1.77

Notes. *Values in the table represents the rate of production of ECS per second.
PI = pre-modified input; IF = interactional feedback; C = comparison

7.4 Syllables per Second during the Production of ECSs

Table 7.11, above, provides the descriptive statistics for the total number of pruned syllables uttered per second as learners in the three groups produced ECSs. As discussed in Chapter 5, Section 5.6.1.3, this measure reflects processing speed of the structures during spontaneous L2 production. There were no maximum scores for this dependent measure. Rather, values in Table 7.11 represent that the higher the scores that learners the three groups achieved, the faster their rate of production of ECSs was.

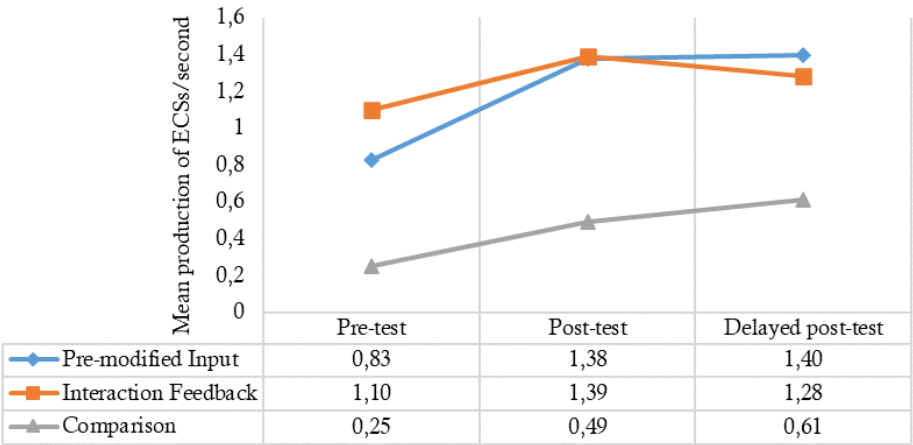


Figure 7.5: Syllables per second during the production of ECSs on the same task.

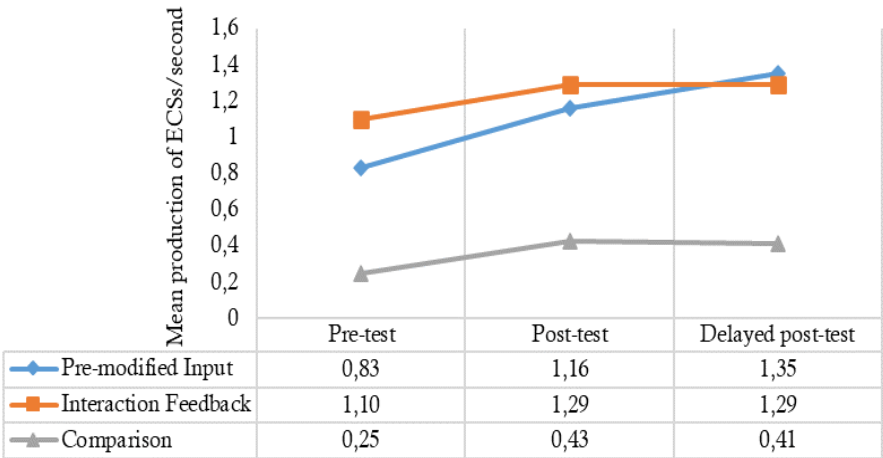


Figure 7.6: Syllables per second during the production of ECSs on parallel tasks.

As indicated in Table 7.11, the three groups increased in the rate of production of ECSs from the pre-test to the post-test on both the same and parallel version of the tasks. Furthermore, such increases were maintained until the delayed post-test. Figures 7.5 and 7.6 present graphs of the mean scores for each of the three groups' rate of production of ECSs on the same task (i.e., Task Set 1-1-1) and on parallel tasks (i.e., Task Set 1-2-3), respectively, in the pre-test, post-test and delayed post-test.

7.4.1 Within-Group Differences for the Rate of Production of ECSs

Although the descriptive statistics show that the three groups' scores of the rate of production of ECSs increased from the pre-test to delayed post-test on both the same and parallel version of the tasks (see Table 7.11; see also Figures 7.5 and 7.6), Friedman tests carried out on these scores revealed that only the PI group made significant increases in the rate of production of ECSs on both the same task, $\chi^2(2) = 12.133$, $p = .002$, and the parallel version of the task, $\chi^2(2) = 6.118$, $p = .047$, whereas the IF and comparison groups did not (see Table 7.12).

Table 7.12: Friedman tests for the speech rate of ECSs

Task Set	**Group**					
	PI (N = 17)		**IF (N = 17)**		**C (N = 17)**	
	Chi-square (χ^2)	**Asymp. Sig. (*p*)**	**Chi-square (χ^2)**	**Asymp. Sig. (*p*)**	**Chi-square (χ^2)**	**Asymp. Sig. (*p*)**
1 – 1 – 1	12.133	.002*	0.353	.838	1.409	.494
1 – 2 – 3	6.118	.047	0.352	.832	0.735	.692

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

Table 7.13: Wilcoxon Signed Ranks tests for the speech rate of ECSs

Group	Task Set	Post-hoc Wilcoxon Signed Ranks Tests			Effect size (*r*)
		Session	Z score	Asymp. Sig. (*p*)	
PI (N = 17)	1-1-1	Pre to Post-test	-2.726	.006*	.575
		Pre to Delayed Post-test	-2.215	.027	.380
		Post to Delayed Post-test	-0.189	.850	.032
	1-2-3	Pre to Post-test	-2.415	.016	.414
		Pre to Delayed Post-test	-1.633	.102	.280
		Post to Delayed Post-test	-0.402	.687	.069
IF (N = 17)	1-1-1	Pre to Post-test	-1.491	.136	.256
		Pre to Delayed Post-test	-1.112	.266	.191
		Post to Delayed Post-test	-0.639	.523	.110
	1-2-3	Pre to Post-test	-0.568	.570	.097
		Pre to Delayed Post-test	-0.398	.691	.068
		Post to Delayed Post-test	-0.852	.394	.146
comparison (N = 17)	1-1-1	Pre to Post-test	-1.718	.086	.295
		Pre to Delayed Post-test	-1.177	.239	.202
		Post to Delayed Post-test	-0.000	1.000	0.000
	1-2-3	Pre to Post-test	-1.376	.169	.236
		Pre to Delayed Post-test	-1.014	.310	.174
		Post to Delayed Post-test	-0.533	.594	.091

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

Post-hoc analyses with Wilcoxon Signed Ranks tests followed by Bonferroni corrections (see Table 7.13, above) performed on the PI groups' test scores further showed that the significant increases were notable only in the group's pre-test and post-test scores on both the same task, z = -2.726, p = .006, and the parallel version of the task, z = -2.415, p = .016, with the magnitude of the differences was large for the same task, r = .575, but was medium for the parallel version of the task, r = .414. Nonetheless, no significant difference was found in the PI group's pre-test

and delayed post-test scores on the same task, z = -2.215, p = .027, and on the parallel version of the task, z = -1.633, p = .102. Indeed, the magnitude of the difference was small for both the same task, r = .380, and the parallel version of the task, r = .280. This, therefore, suggests that the gains and the transferability of the gains in the rate of production of ECSs made by the PI group were unstable.

Post-hoc analyses with Wilcoxon Signed Ranks tests followed by Bonferroni corrections were also performed on the IF and comparison groups' pre-test, post-test and delayed post-test scores. The analyses confirmed that these two groups did not make significant increases in their rate of production of ECSs from the pre-test to delayed post-test on both the same task and the parallel version of the task. The calculations for the effect sizes of these groups' scores also revealed that the magnitude of the differences between their scores in the (i) pre-test and post-test; (ii) post-test and delayed post-test; and (iii) pre-test and delayed post-test on both the same and parallel version of the tasks were small (see Table 7.13, above). Unlike the comparison group, however, absence of significant gains by the IF group in the post-test and delayed post-test may be due to a ceiling effect, since the IF group had started off producing ECSs faster than the other two (PI and comparison) groups in the pre-test (see Table 7.11).

7.4.2 Between-Group Differences for the Rate of Production of ECSs

Kruskal-Wallis H tests (see Table 7.14, below) revealed that there was already a significant difference between the three groups in their rate of production of ECSs in the pre-test, $\chi^2(2) = 10.079$, p = .006, and that such a difference continued in the post-test on both the same task, $\chi^2(2) = 14.803$, p = .001, and the parallel version of the task, $\chi^2(2) = 14.851$, p = .001. The difference also occured in the delayed post-

test, too, on both the same task, $\chi^2(2)$ 8.849, p = .014, and the parallel version of the task, $\chi^2(2)$ = 14.102, p = .001.

Table 7.14: Kruskal-Wallis H tests for the speech rate of ECSs

Session	Task Set	Group	Mean Rank	Kruskal-Wallis H Tests	
				Chi-square (χ^2)	**Asymp. Sig. (*p*)**
Pre-test	Task Set 1	PI (N = 17)	28.441	10.079	.006*
		IF (N = 17)	32.294		
		C (N = 17)	17.265		
Post-test	Task Set 1	PI (N = 17)	31.353	14.803	.001*
		IF (N = 17)	31.912		
		C (N = 17)	14.735		
	Task Set 2	PI (N = 17)	32.235	14.851	.001*
		IF (N = 17)	31.029		
		C (N = 17)	14.735		
Delayed Post-test	Task Set 1	PI (N = 17)	30.706	8.849	.014
		IF (N = 17)	29.765		
		C (N = 17)	17.529		
	Task Set 3	PI (N = 17)	33.588	14.102	.001*
		IF (N = 17)	29.059		
		C (N = 17)	15.353		

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

Post-hoc analyses with Mann-Whitney U tests followed by Bonferroni corrections (see Table 7.15) performed to examine these groups' *pre-test* scores, however, revealed that a significant difference in the pre-test was identified only between the IF and comparison group, U = 61.000, p = .002, with the magnitude of difference that was medium, r = .476. No significant difference was found between the PI and comparison group, U = 79.500, p = .017, or between the IF and PI group, U = 121.000, p = .414. However, the effect size calculated for the PI and comparison

group indicated that the magnitude of the difference was medium, r = .392, whereas the difference between the IF and PI group was null, r = .140.

Table 7.15: Mann-Whitney U tests for the speech rate of ECSs

Session	Task Set	Post-hoc Mann-Whitney U Tests			Effect size (r)
		Group	*U* statistic	Asymp. Sig. (p)	
Pre-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	121.000	.414	.140
		PI (N = 17) vs. C (N = 17)	79.500	.017	.392
		IF (N = 17) vs. C (N = 17)	61.000	.002*	.476
Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	144.000	.986	.003
		PI (N = 17) vs. C (N = 17)	54.000	.002*	.544
		IF (N = 17) vs. C (N = 17)	43.000	.000*	.602
	Task Set 2	PI (N = 17) vs. IF (N = 17)	140.000	.877	.027
		PI (N = 17) vs. C (N = 17)	43.000	.001*	.605
		IF (N = 17) vs. C (N = 17)	54.500	.002*	.539
Delayed Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	129.000	.593	.092
		PI (N = 17) vs. C (N = 17)	80.000	.022	.410
		IF (N = 17) vs. C (N = 17)	65.000	.006*	.521
	Task Set 3	PI (N = 17) vs. IF (N = 17)	119.000	.380	.151
		PI (N = 17) vs. C (N = 17)	41.000	.001*	.619
		IF (N = 17) vs. C (N = 17)	67.000	.006*	.471

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

In the *post-test* session, the analyses showed that both the IF and PI groups outperformed the comparison group on the same task: PI vs. comparison, U = 54.000, p = .002; IF vs. comparison, U = 43.000, p = .000, and on the parallel version of the task, PI vs. comparison, U = 43.000, p = .001; IF vs. comparison, U = 54.500, p = .002. Unlike in the pre-test, however, the magnitude of the differences increased substantially on the same task: PI vs. comparison, r = .544; IF vs. comparison, r = .602, as well as on the parallel version of the task: PI vs.

comparison, r = .609; IF vs. comparison, r = .539. No significant difference was found between the PI and IF group on the same task, U = 144.000, p = .986, and on the parallel version of the task, U = 140.000, p = .877. Furthermore, the magnitude of the differences remained null between the PI and IF groups for both the same task, r = .003, and the parallel version of the task, r = .027. This suggests that the gains in the rate of production of ECSs that the IF and PI group made in the post-test session were more substantial than those made by the comparison group.

In the *delayed post-test*, however, the difference between the experimental (IF and PI) and comparison groups decreased, especially in terms of their performances on the same task. This resulted in fairly similar results in the statistical comparisons as those reported in the pre-test: PI vs. comparison, U = 80.000, p = .022, r = .410; IF vs. comparison, U = 65.000, p = .006, r = .521. Such decreases may be due to the fact that the comparison group also improved quite substantially in their rate of production of ECSs when completing the same task in the delayed post-test (see Table 7.11). Notwithstanding, the difference between the experimental and comparison groups remained significantly different with respect to their performance on the parallel task: PI vs. comparison, U = 41.000, p = .001; IF vs. comparison; U = 67.000, p = .006. The calculations for the effect sizes further revealed that the magnitude of the difference was large between the PI and comparison group, r = .619, and was medium between the IF and comparison group, r = .471. Finally, no significant difference was found between the experimental (PI and IF) groups' scores of the rate of production of ECSs in the delayed post-test on both the same task, U = 129.000, p = .593, and the parallel version of the task, U = 119.000, p = .380. Furthermore, the magnitude of the

differences remained null for these tasks: for the same task, r = .092, and for the parallel version of the task, r = .151.

These results thus indicate that conditions that allow learners exposure to specific task-relevant L2 forms recurrently while performing certain L2 communication tasks *aurally* or *orally* (such is realized through pre-modified input or interactional feedback) are more effective in facilitating increased speed of processing of the forms than those which only allow learners to perform tasks alone without the provision of input or feedback. Despite this, it appears that the PI condition is slightly better in promoting speed of processing of task-relevant L2 forms than the IF condition.

7.5 Frequency of Breakdowns during the Production of ECSs

Table 7.16, below, provides the descriptive statistics for the frequency of breakdowns per second as participants in the three groups produced ECSs when completing the same and parallel tasks. As discussed in Chapter 5, Section 5.6.2.2, this measure reflects one domain of processing stability in the use of the structures in spontaneous L2 production. There are no maximum scores for this measure. Rather, values in Table 16 reflect that the lower the frequency of breakdowns that occurred per second as the participants used ECSs in real-time production, the more stable their processing of the form was.

As shown in Table 7.16, the frequency of breakdowns during the production of ECSs per second decreased in the three groups in the post-test on both the same and parallel tasks, and that such decreases remained constant until the delayed post-test. Figures 7.7 and 7.8 present graphs of the mean scores of the frequency of breakdowns in the production of ECSs per second by each of the three groups on

the same task (i.e., Task Set 1-1-1) and on the parallel version of the task (i.e., Task Set 1-2-3), respectively, in the pre-test, post-test and delayed post-test.

Table 7.16: Descriptive statistics for the frequency of breakdowns per second during the production of ECSs

Session	Task Set	Descriptive Statistics	Group		
			PI (N = 17)	IF (N = 17)	Control (N = 17)
Pre-test	Task Set 1	Mean	1.03	0.82	1.60
		95% CI (lower)	0.68	0.45	1.27
		95% CI (upper)	1.39	1.19	1.93
		Minimum	0.14	0.00	0.43
		Maximum	2.00	2.00	2.00
		Median	0.61	0.55	2.00
		SD	1.86	0.72	0.65
		Range	0.69	2.00	1.57
Post-test	Task Set 1	Mean	0.71	0.54	1.36
		95% CI (lower)	0.45	0.48	0.99
		95% CI (upper)	0.97	0.59	1.74
		Minimum	0.33	0.27	0.09
		Maximum	2.00	0.72	2.00
		Median	0.50	0.55	2.00
		SD	0.51	0.12	1.91
		Range	1.67	0.45	0.73
	Task Set 2	Mean	0.62	0.63	1.20
		95% CI (lower)	0.42	0.35	0.79
		95% CI (upper)	0.81	0.90	1.61
		Minimum	0.26	0.24	0.24
		Maximum	2.00	2.00	2.00
		Median	0.56	0.46	0.82
		SD	0.38	0.53	0.79
		Range	1.74	1.76	1.72
Delayed Post-test	Task Set 1	Mean	0.86	0.54	1.34
		95% CI (lower)	0.51	0.33	0.96
		95% CI (upper)	1.20	0.74	1.73
		Minimum	0.17	0.28	0.21
		Maximum	2.00	2.00	2.00
		Median	0.58	0.43	2.00
		SD	0.67	0.40	1.79
		Range	1.83	1.72	0.74
	Task Set 3	Mean	0.52	0.73	1.42
		95% CI (lower)	0.43	0.41	1.04
		95% CI (upper)	0.62	1.05	1.79
		Minimum	0.15	0.00	0.43
		Maximum	0.86	2.00	2.00
		Median	0.54	0.50	2.00
		SD	0.18	0.63	0.73
		Range	0.71	2.00	1.57

Notes. *Values in the table represents the frequency of breakdowns in the use of ECS per second. PI = pre-modified input; IF = interactional feedback; C = comparison

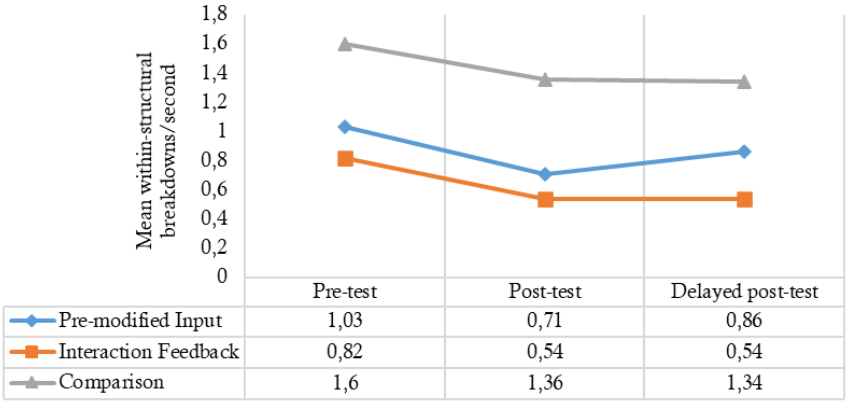


Figure 7.7. Frequency of breakdowns per second during the production of ECSs on the same task

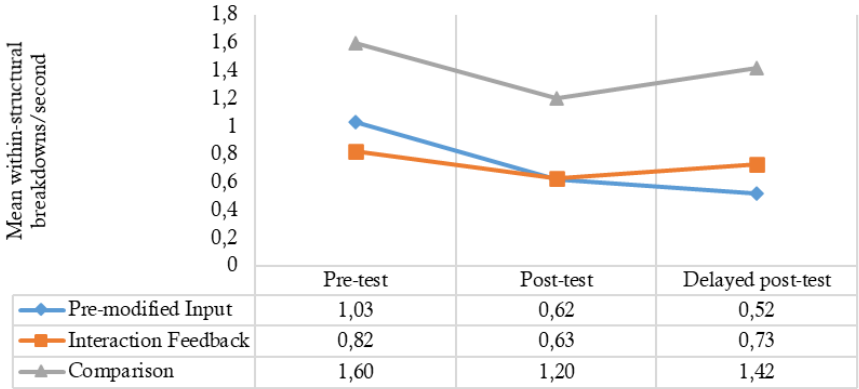


Figure 7.8. Frequency of breakdowns per second during the production of ECSs on parallel tasks

7.5.1 Within-Group Differences for the Frequency of Breakdowns during the Production of ECSs

Although the descriptive statistics show quite noticeable decreases in the frequency of breakdowns during the production of ECSs by the three groups from the pre-test to the delayed post-test (see Table 7.16; see also Figures 7.7 and 7.8), Friedman tests carried out on the three repeated tests revealed that none of the decreases by the three groups reached statistical significance (see Table 7.17).

Table 7.17: Friedman tests for the frequency of breakdowns per second during the production of ECSs

Task Set	Group					
	PI (N = 17)		IF (N = 17)		C (N = 17)	
	Chi-square (χ^2)	Asymp. Sig. (*p*)	Chi-square (χ^2)	Asymp. Sig. (*p*)	Chi-square (χ^2)	Asymp. Sig. (*p*)
1 – 1 – 1	1.882	.390	.824	.662	.735	.692
1 – 2 – 3	3.029	.220	.824	.662	2.735	.255

Notes. PI = pre-modified input, IF = interactional feedback, C = comparison.

Table 7.18: Wilcoxon Signed Ranks Tests for the frequency of breakdowns per second during the production of ECSs

Group	Task Set	Post-hoc Wilcoxon Signed Ranks Tests			Effect size (*r*)
		Session	Z score	Asymp. Sig. (*p*)	
PI (N = 17)	1-1-1	Pre- to Post-test	-1.676	.094	.287
		Pre- to Delayed Post-test	-1.022	.307	.175
		Post- to Delayed Post-test	-0.909	.363	.156
	1-2-3	Pre- to Post-test	-1.758	.079	.302
		Pre- to Delayed Post-test	-2.059	.039	.353
		Post- to Delayed Post-test	-0.166	.868	.028
IF (N = 17)	1-1-1	Pre- to Post-test	-0.734	.463	.126
		Pre- to Delayed Post-test	-1.018	.309	.175
		Post- to Delayed Post-test	-1.108	.268	.190
	1-2-3	Pre- to Post-test	-0.881	.379	.151
		Pre- to Delayed Post-test	-0.341	.733	.058
		Post- to Delayed Post-test	-1.108	.268	.190
C (N = 17)	1-1-1	Pre- to Post-test	-0.949	.343	.163
		Pre- to Delayed Post-test	-0.903	.367	.155
		Post- to Delayed Post-test	-0.178	.859	.030
	1-2-3	Pre- to Post-test	-1.784	.074	.306
		Pre- to Delayed Post-test	-1.183	.237	.203
		Post- to Delayed Post-test	-1.156	.248	.198

Notes. PI = pre-modified input, IF = interactional feedback, C = comparison.

Post-hoc analyses with Wilcoxon Signed Rank Tests followed by Bonferroni corrections performed on the three groups' pre-test, post-test, delayed post-test scores on both the same and parallel version of the tasks confirmed that no significant decreases took place in the frequency of breakdowns during the production of ECSs. The calculations for the effect sizes also revealed that the magnitude of the differences in the three groups' scores between (i) the pre-test and post-test; (ii) post-test and delayed post-test; and (iii) pre-test and delayed post-test were either small or null. Table 7.18 summarizes the results of these within-group comparisons.

7.5.2 Between-Group Differences for the Frequency of Breakdowns during the Production of ECSs

With respect to between-group measures, the Kruskal-Wallis H tests (see Table 7.19) revealed that there was already a statistically significant difference between groups in the pre-test, $\chi^2(2) = 10.213$, $p = .006$, and that such a difference remained constant in the post-test, especially on the same task, $\chi^2(2) = 13.229$, $p = .001$, but not on the parallel version of the task, $\chi^2(2) = 5.094$, $p = .078$. In the delayed post-test, however, significant differences of the groups' scores on this test occurred on both the same task, $\chi^2(2)$ 12.504, $p = .002$, and the parallel version of the task, $\chi^2(2) = 12.352$, $p = .002$.

Post-hoc analyses with Mann-Whitney U tests followed by Bonferroni corrections performed to examine the groups' *pre-test* scores (see Table 7.20) further revealed that a significant difference in the pre-test was identified only between the IF and comparison group, $U = 61.000$, $p = .002$, with the magnitude of difference that was medium to large, $r = .521$. No significant difference was found between the PI and

comparison group, $U = 83.500$, $p = .025$, or between the IF and PI group, $U = 109.000$, $p = .217$. However, the effect size calculated for the PI and comparison group indicated that the magnitude of the difference was small to medium, $r = .385$, whereas the difference between the IF and PI group was null to small in magnitude, $r = .212$.

Table 7.19: Kruskal-Wallis H tests for the frequency of breakdowns per second during the production of ECSs

Session	Task Set	Group	Mean Rank	Kruskal-Wallis H Tests	
				Chi-square (χ^2)	Asymp. Sig. (*p*)
Pre-test	Task Set 1	PI (N = 17)	24.500	10.213	.006*
		IF (N = 17)	19.000		
		C (N = 17)	34.500		
Post-test	Task Set 1	PI (N = 17)	21.706	13.229	.001*
		IF (N = 17)	19.706		
		C (N = 17)	36.588		
	Task Set 2	PI (N = 17)	25.029	5.094	.078
		IF (N = 17)	20.824		
		C (N = 17)	32.147		
Delayed Post-test	Task Set 1	PI (N = 17)	26.353	12.504	.002*
		IF (N = 17)	16.912		
		C (N = 17)	34.735		
	Task Set 3	PI (N = 17)	20.000	12.352	.002*
		IF (N = 17)	21.794		
		C (N = 17)	36.206		

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

In the *post-test* session, the analyses showed that both PI and IF groups outperformed the comparison group on the same task: PI vs. comparison, $U = 63.500$, $p = .005$; IF vs. comparison, $U = 45.500$, $p = .001$, but not on the parallel version of the task: PI vs. comparison, $U = 95.500$, $p = .088$; IF vs. comparison, U

= 89.000, p = .053. The calculation for effect sizes also revealed that the magnitude of the differences between the two experimental groups and the comparison group was medium to large on the same task: PI vs. comparison, r = .487; IF vs. comparison, r = .590, and was small on the parallel version of the task: PI vs. comparison, r = .292; IF vs. comparison, r = .332. No significant difference was found between the PI and IF group on both the same task, U = 136.500, p = .783, and the parallel version of the task, U = 112.000, p = .262. Furthermore, the magnitude of the differences was null for these tasks: for the same task, r = .047; for the parallel version of the task, r = .192. This, therefore, suggests that decreases in the frequency of breakdowns during production of ECSs made by the IF and, particularly, PI group in the post-test session were more substantial than those made by the comparison group.

In the *delayed post-test*, only the IF group outperformed the comparison group on the same task, U = 49.500, p = .001, with the magnitude of difference that was large, r = .569, while there was no statistically significant difference between the PI and comparison group, U = 91.000, p = .058, or between the PI and IF group, U = 85.000, p = .040. Furthermore, the magnitude of the differences of these latter group comparisons was also small: PI vs. comparison, r = .325, PI vs. IF, r = .352. As indicated in Table 7.16, these results may be due to the fact that all groups (PI, IF, and comparison) produced fewer breakdowns when producing ECSs while completing the same task in the delayed post-test. Despite this, the IF group produced much fewer breakdowns than the other two (PI and comparison) groups.

With regard to the groups' performance on the parallel task, both the IF and PI groups produced substantially fewer breakdowns during the production of ECSs than the comparison group, resulting in significant differences between the IF and

comparison group, $U = 67.000$, $p = .006$, as well as the PI and comparison group, $U = 48.500$, $p = .001$, with the magnitude of the differences that was medium between the IF and C group, $r = .471$, and medium to large between the PI and comparison group, $r = .574$. No significant difference was found between the IF and PI group, $U = 138.500$, $p = .836$, with the magnitude of the difference that was null in size, $r = .035$.

Table 7.20: Mann-Whitney U tests for the frequency of breakdowns per second during the production of ECSs

Session	**Task Set**	**Post-hoc Mann-Whitney U Tests**			**Effect size (*r*)**
		Group	***U* statistic**	**Asymp. Sig. (*p*)**	
Pre-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	109.000	.217	.212
		PI (N = 17) vs. C (N = 17)	83.500	.025	.385
		IF (N = 17) vs. C (N = 17)	61.000	.002*	.521
Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	136.500	.783	.047
		PI (N = 17) vs. C (N = 17)	63.500	.005*	.487
		IF (N = 17) vs. C (N = 17)	45.500	.001*	.590
	Task Set 2	PI (N = 17) vs. IF (N = 17)	112.000	.262	.192
		PI (N = 17) vs. C (N = 17)	95.500	.088	.292
		IF (N = 17) vs. C (N = 17)	89.000	.053	.332
Delayed Post-test	Task Set 1	PI (N = 17) vs. IF (N = 17)	85.000	.040	.352
		PI (N = 17) vs. C (N = 17)	91.000	.058	.325
		IF (N = 17) vs. C (N = 17)	49.500	.001*	.569
	Task Set 3	PI (N = 17) vs. IF (N = 17)	138.500	.836	.035
		PI (N = 17) vs. C (N = 17)	48.500	.001*	.574
		IF (N = 17) vs. C (N = 17)	67.000	.006*	.471

Notes. *mean difference is significant at the .01 level (2-tailed).
PI = pre-modified input, IF = interactional feedback, C = comparison.

In sum, the statistical analyses on this measure indicate that although the three groups produced fewer breakdowns during the production of ECSs in the post-test

and delayed post-tests than those in the pre-test, the PI and IF groups appeared to produce fewer breakdowns than the comparison group. Nonetheless, it appears that the PI group performed slightly better than the IF group in the immediate post-test on the same task, but not on the parallel task. In the delayed post-test, on the other hand, the IF group seemed to perform slightly better than the PI group on both the same and parallel version of the tasks.

7.6 Frequency of Repairs in the Production of ECSs per Second

Table 7.21, below, presents descriptive statistics of the frequency of repairs during the production of ECSs by the participants in the three groups. As stated in Chapter 5, Section 5.6.2.2, measures of frequency of repairs similarly relate to processing stability as indexed by measures of frequency of breakdowns reported in the previous section. Thus, values in Table 7.21 also indicate that the lower the frequency of repairs that occurred as the participants produced ECSs in real time, the more stable their processing of the structures was.

As stated in the beginning of this chapter, however, only descriptive statistics are reported in this section. This is because the frequency of repairs (self-corrections, replacements and/or reformulations) which occurred as the participants attempted to produce ECSs in real time was very low, making statistical analyses meaningless to carry out.

As shown in Table 7.21, differences between the three groups' self-repairs in the production of ECSs were nominal.

Table 7.21: Descriptive statistics for the frequency of repairs during the production of ECSs*

Session	Task Set	Descriptive Statistics	Group		
			PI (N = 17)	IF (N = 17)	Control (N = 17)
Pre-test	Task Set 1	Mean	0.01	0.01	0.01
		95% CI (lower)	-0.00	-0.01	-0.01
		95% CI (upper)	0.02	0.03	0.04
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.00	0.00	0.00
		SD	0.02	0.04	0.05
		Range	0.08	0.17	0.19
Post-test	Task Set 1	Mean	0.02	0.02	0.02
		95% CI (lower)	-0.00	-0.00	-0.00
		95% CI (upper)	0.04	0.04	0.05
		Minimum	0.00	0.00	0.00
		Maximum	1.00	0.14	1.00
		Median	0.00	0.00	0.00
		SD	0.04	0.04	0.49
		Range	0.14	0.14	0.13
	Task Set 2	Mean	0.00	0.00	0.01
		95% CI (lower)	-0.00	-0.00	-0.01
		95% CI (upper)	0.01	0.01	0.02
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.00	0.00	0.00
		SD	0.01	0.01	0.02
		Range	0.04	0.05	0.09
Delayed Post-test	Task Set 1	Mean	0.01	0.02	0.01
		95% CI (lower)	-0.01	-0.02	-0.01
		95% CI (upper)	0.03	0.07	0.03
		Minimum	0.00	0.00	0.00
		Maximum	1.00	1.00	1.00
		Median	0.00	0.00	0.00
		SD	0.04	0.09	0.04
		Range	0.17	0.36	0.16
	Task Set 3	Mean	0.03	0.08	0.02
		95% CI (lower)	0.00	-0.05	-0.00
		95% CI (upper)	0.06	0.20	0.05
		Minimum	0.00	0.00	0.00
		Maximum	0.18	1.00	0.00
		Median	0.00	0.00	0.00
		SD	0.06	0.24	0.05
		Range	0.18	1.00	0.14

Notes. *Values in the table represents the proportion of AS-units containing the use of ECS.
PI = pre-modified input; IF = interactional feedback; C = comparison

7.7 Summary of the Overall Findings of the Study

In summary, the results of statistical analyses presented above indicated that PI and IF conditions were effective in promoting uptake and some domains of automaticity of a specific task-relevant L2 form (ECSs). Such effectiveness was indicated by no significant differences between the PI and IF groups' pre-test and post-test scores on the two dependent measures representing uptake (i.e., the proportion of the uses of ECSs, and the proportion of error-free uses of ECSs), as well as on other two dependent measures representing specific domains of automaticity (i.e., the total number of pruned syllables uttered per second, and the frequency of breakdowns occurred per second as participants produced ECSs). Despite this, it appears that the IF condition is slightly better than the PI condition in promoting uptake and specific domains of automaticity, as reflected in the gains made by the IF group on the four dependent measures of uptake and automaticity in comparison to those made by the comparison group in the pos-test and delayed post-test.

Gains in the uptake and some domains of automaticity of the task-relevant L2 form (ECSs) that the PI and IF conditions promoted also appeared to be stable. The stability of the gains was indicated by no significant differences in the post-test and delayed post-test scores made by participants in the PI and IF groups. That is, participants in these two groups similarly achieved comparable level of gains that made the magnitude of differences between them in terms of the four dependent measures from the post-test to delayed post-test were small throughout. As compared with the gains made by the comparison group, however, those made by the IF group appeared to be slightly better than those made by the PI group.

Finally, the results of statistical analyses of the PI and IF groups' scores on parallel task performances further revealed that gains in the uptake and in some domains of automaticity of ECSs from PI and IF conditions were transferable to parallel tasks. The transferability of the gains, however, seems slightly better in the IF group than those in the PI group. Chapter 8 will discuss these findings in light of theories of and research on task-based L2 learning and presented in previous chapters.

Chapter 8

Discussion

This chapter discusses the overall results in light of relevant theoretical accounts associated with task-based L2 research and pedagogy. To begin, each research question and hypothesis outlined in the beginning of Chapter 6 are discussed in relation to the results reported in Chapter 7 as well as to those established in the earlier work. Following this, theoretical and methodological implications of the study to the field of task-based L2 learning are presented, and the applicability of the procedures applied in the study to task-based L2 instruction is discussed. The chapter concludes with outlining limitations of the study, and how they may be addressed in future L2 research.

8.1 Overview of the Study

As described in Chapters 2 and 4, the current study was motivated by the claim that the provision of interactional feedback (which occurs following learners' non-target-like or less comprehensible L2 output) might be more effective than the provision of pre-modified input (which affords only positive evidence) in facilitating, or driving forward, L2 development (Gass & Mackey, 2015; Long, 1996, 2007, 2015; Mackey, 1999). Drawing on usage-based accounts of L2 development, the abstract notion of L2 development was conceptualized in this study as gradual improvements in the use of task-relevant L2 forms by learners over time, especially as they produced the forms under real-operating conditions (see Chapter 3). More specifically, such gradual improvements were

conceptualized in relation to ‘uptake’ and ‘automaticity’ of English comparative structures (ECSs), i.e., morphosyntactic structures which commonly occurred on the tasks used in the study as determined by English native-speaker baseline data (Lambert, 2019; Lambert & Nakamura, 2019). In this study, the term ‘uptake’ was thus operationally defined as learners’ incorporation of ECSs from the input or feedback into his or her L2 production. Furthermore, such incorporation was specifically indicated by *increased frequency* and/or *increased accuracy of the structures* as learners attempted to complete relevant L2 communication tasks orally in the post-test and/or delayed post-test, relative to those in the pre-test. On the other hand, the term ‘automaticity’ was operationally defined as learners’ abilities in using ECSs *rapidly without undue pauses*, *repetitions and repairs* in naturally-occurring task-based L2 communication. Hence, gains in automaticity were indicated by improvements in aspects of cognitive fluency that are associated with automaticity underlying learners’ spontaneous production of the structures. These include (1) *speed fluency* (operationalized as the number of pruned syllables, including pauses, in the production of the structures per second); (2) *breakdown fluency* (operationalized as the frequency of filled/unfilled pauses and verbatim repetitions that occurred during the production of the structures per second); and (3) *repair fluency* (operationalized as the frequency of overt self-repairs that occurred during the production of the structures per second) (see Chapter 5 for the rationale behind the use of these measures of uptake and automaticity).

It is worth noting that usage-based accounts maintain that L2 development is not necessarily indicated only by increased frequency, accuracy and/or fluency in learners’ use of L2 forms (which constitute a function of *stabilization* in learners’ L2 resources). Rather, such development may also be indicated by decreases in

these three domains of L2 use (e.g., as a result of *destabilization* in learners' L2 resources and/or *emergence* of knowledge of form-function-meaning mappings—see Chapter 3 for details). Despite this, few (if any) would deny that improvements in the frequency, accuracy and/or fluency in learners' use of L2 forms within the context of naturally-ocurring task-based L2 production can demonstrate that L2 development has taken place (Crossley, Salsbury, McNamara & Jarvis, 2011; de Bot, Lowie & Verspoor, 2007; DeKeyser, 2017; N. Ellis, 2008, 2009a, 2019; R. Ellis, 1990; Gries & N. Ellis, 215; Herdina & Jessner, 2002; Horst & Collins, 2006; Larsen-Freeman, 2002, 2013a; Norris & Ortega, 2003, 2013; Segalowitz, 2010, 2016; van Geert, 1991).

8.1.1 Effects of the Conditions on the Uptake and Automaticity of ECSs

The first research question asked whether pre-modified input and interactional feedback differed in their effects on the uptake and automaticity of English comparative structures (ECSs) as reflected in differences in the use of these structures by Indonesian adult learners of L2 English as they completed tasks orally in the pre-test and post-test (Chapter 6, Section 6.1). Based on the review of the literature presented in Chapters 2, 3 and 4, two research hypotheses were proposed in line with this first research question. First, it was hypothesized that the provision of interactional feedback would be more effective than the provision of pre-modified input in promoting *uptake* of the ECSs. This is because interactional feedback was more likely to induce learners to carry out *cognitive comparisons* between their non-target-like use and the target-like use provided by their interlocutor (the researcher), leading them to notice gaps in their L2 resources during communication (Gass, 1997; Gass & Mackey, 2015; Long, 1996, 2007,

2015). Second, the provision of interactional feedback would be more effective than the provision of pre-modified input in promoting *automaticity* of ECSs. This is because interactional feedback was more likely to stimulate *cognitive rehearsing* in learners' working memory, especially when they attempted to use the structures in their own production after having them recast or modelled in confirmation checks by their interlocutor (the researcher) (N. Ellis & Sinclair, 1996; Robinson, 2001, 2003b, 2010, 2011a, 2015; Skehan, 1998a, 2007, 2009b, 2018; Swain, 1995).

Both the first and the second research hypothesis were not supported, since the two conditions did not differ as predicted. That is, results of the study indicated that the provision of pre-modified input or interactional feedback similarly led to immediate gains in the uptake and in the two aspects of cognitive fluency associated with automaticity of ECSs: *speed* and *breakdown fluencies*. Put another way, participants in the pre-modified input group and those in interactional feedback group similarly produced these structures more frequently, more accurately and more rapidly along with a lower frequency of filled/unfilled pauses and verbatim repetitions as they attempted to complete the same task orally in the post-test, relative to those in the pre-test (see Chapter 7, Sections 7.2 - 7.5). In contrast, those in the comparison group, who only received communication practice (FonM), did not show noticeable gains in any measures of uptake and automaticity of ECSs. Thus, FonM appeared to be ineffective in facilitating improvements in the frequency, accuracy and fluency of these task-relevant L2 structures (see Chapter 7, Sections 7.2 - 7.6).

These findings thus echo, and extend, those reported on in previous L2 studies (e.g., Ayoun, 2001; Gass & Torres, 2005; Lyddon, 2011; Shintani, 2011; Shintani & R. Ellis, 2010). That is, the present study indicated that the provision of pre-

modified input or interactional feedback was more effective that FonM, and that these two forms of feedback did not differ significantly in promoting improvements in the frequency and/or accuracy of specific task-relevant L2 forms. Previous studies which showed that the provision of interactional feedback was more effective than the provision of pre-modified input in promoting improvements in the frequency and/or accuracy of specific task-relevant L2 forms also revealed that the provision of pre-modified input in task-based L2 instruction was effective in promoting improvements in the frequency and/or accuracy of specific task-relevant L2 forms (e.g., de la Fuente, 2002; Long, Inagaki & Ortega, 1998; Mackey, 1999; Iwashita, 2003). This suggests, therefore, that although the provision of interactional feedback in task-based L2 instruction may be more effective than the provision of pre-modified input in facilitating uptake of specific task-relevant L2 forms, these two instructional options (*under certain circumstances*) can be equally effective in promoting uptake of the forms (see also Section 8.2, below).

With regard to automaticity, previous studies on the efficacy of pre-modified input and interactional feedback have rarely investigated whether these two instructional options facilitate automaticity of specific task-relevant L2 forms. The findings of the current study thus extend earlier work in that both options appear to be effective in facilitating uptake of specific task-relevant L2 forms and promoting automaticity of the forms.

Some L2 researchers, however, may argue that gains in the speed and breakdown fluencies of ECSs that the participants in the two experimental groups made may be insufficient to provide evidence of gains in automaticity of these structures. This is because, as discussed in Chapter 5, Section 5.2.2.2, gains in automaticity also entail significant decreases in the repair fluency. Thus, a criticism that may be

levelled against the discussion presented in this section is that the gains in the processing of ECSs that participants in the pre-modified input and interactional feedback groups made may be better conceived as mere speed-ups, rather than increases in automaticity, of these structures (see Chapter 5, Section 5.6.2, for a distinction between a speed-up and automaticity). Although such a criticism is reasonable on methodological grounds (de Jong, Steinel, Florijn, Schoonen & Hulstijn, 2013; Segalowitz, 2003; Suzuki, 2018), it is not clear whether gains in the processing of ECSs that the two experimental groups made may constitute evidence of mere speed-ups. This is because the data in the current study did not allow statistical tests to be performed on this dependent measure of repair fluency of ECSs (i.e., due to inadequate number of instances—see Chapter 7, Section 7.6). Given that 'absence of evidence' is *not* necessarily the same as 'evidence of absence' (Godfroid & Spino, 2015, Plonsky, 2017), gains in the online processing of ECSs that the participants in the pre-modified input and interactional feedback groups made will be conceptualized in relation to gains in automaticity, rather than mere speed-ups, of the structures.

Another plausible criticism concerning such a low number of occurrences of repairs might be that the participants of this study had already possessed relatively automatized knowledge of the ECSs before the treatment. Such criticism, however, may not hold given the operational definition of automaticity in the present study. That is, as noted in Section 5.6, the term of automaticity is operationally defined in the current study as an ability in using the target L2 form (ECSs) *rapidly without undue pauses, repetitions and repairs* in naturally-occurring task-based L2 communication. Given that the participants often experienced breakdowns (indexed by the frequency of filled/unfilled pauses) as they attempted to use the form while

completing a given task in the pre-test, a plausible interpretation of such low occurrences of repairs is that the participants rarely monitored their production of the structures during task-based performance (see Ch. 5, for details).

8.1.2 Stability of the Effects

The second research question asked whether gains in the uptake and automaticity of ECSs from pre-modified input or interactional feedback were stable over time as reflected in differences in the use of these structures by Indonesian adult learners of L2 English while completing relevant L2 communication tasks orally in the post-test and the delayed post-test (Chapter 6, Section 6.1). Based on the review of the literature in Chapters 2, 3 and 4, it was hypothesized that gains in the uptake and automaticity of ECSs through interactional feedback would be more stable than those thorough pre-modified input. This is because learners' production and *re*production of the structures following recasting and/or modelling in confirmation check by the interlocutor (the researcher) were argued to be more likely to stimulate *deep syntactic processing*. As discussed in Chapter 4, Sections 4.5 and 4.6, deep syntactic processing plays an important role in facilitating retention of target linguistic information in learners' long-term memory (Gass, 1997; Gass & Mackey, 2015; Long, 1996, 2007, 2015).

Results of the current study did not support the hypothesis. The results indicated that gains in the uptake (i.e., more frequent and/or more accurate use) of ECSs and in the two aspects of cognitive fluency (i.e., speed and breakdown fluencies) associated with automaticity of ECSs that the two experimental groups made did not differ significantly between the post-test and the the delayed post-test. In contrast, those in the comparison group did not make substantial gains in any

measures of uptake and automaticity of these structures as reflected in no significant differences in their use of ECSs in the pre-test, post-test and delayed post-test (see Chapter 7, Sections 7.2 – 7.6).

These findings thus echo, and further extend, those established in previous L2 studies (e.g., Gass & Torres, 2005; Lyddon, 2011; Shintani, 2011; Shintani & R. Ellis, 2010), where they showed that gains in the uptake of specific task-relevant L2 forms from pre-modified input or interactional feedback were equally stable (cf., Ayoun, 2001; de la Fuente, 2002; Iwashita, 2003; Long, Inagaki & Ortega, 1998; Mackey, 1999; Strapp, Helmick, Tonkovich, & Bleakney, 2011). Again, however, this line of L2 studies did not examine whether these two instructional options facilitated automaticity of task-relevant L2 forms, and whether gains in the automaticity of the forms were stable over time (see Section 8.1.1 above). Hence, findings of the current study extend those established in the earlier work in suggesting that the stability of gains from pre-modified input and interactional feedback were not only evident in the uptake of task-relevant L2 forms but also in the increased automaticity of the forms.

8.1.3 Transferability of the Effects

The third research question asked whether gains in the uptake and automaticity of ECSs from the provision of pre-modified input or interactional feedback were *transferable* as reflected in differences in the use of these structures by Indonesian adult learners of L2 English while completing the same versus parallel tasks on the post-test and on the delayed post-test (Chapter 6, Section 6.1). Based on the review of the literature presented in Chapters 2, 3 and 4, it was hypothesized that gains in the uptake and automaticity of ECSs that interactional feedback promoted would be

more transferable to parallel tasks than those promoted by pre-modified input as a result of the same process of deep syntactic processing. That is, deep syntactic processing which facilitates retention of critical linguistic information in learners' long-term memory (see Section 8.1.2, above) also plays a pivotal role in the retrieval of the information as learners attempt to satisfy similar or related functional demands of tasks like those they already successfully performed (N. Ellis, 2001, 2019; Fenk-Oczlon, 2001; Lightbown, 2008; MacWhinney, 1999, 2001a; Robinson, 2003a, 2003b, 2005b, 2010, 2011a, 2015; Segalowitz, 2010, 2016; Segalowitz & Lightbown, 1999).

This research hypothesis was also not supported. Results of the study indicate that gains in both uptake (i.e., more frequent and/or more accurate use) and in the two domains of cognitive fluency (i.e., speed and breakdown fluencies) associated with automaticity of ECSs that the participants in the pre-modified input and interactional feedback groups made did not differ significantly in their transferability to parallel tasks in the post-test and delayed post-test. In contrast, those in the comparison group neither showed noticeable nor transferable gains in any measures of uptake and automaticity of ECSs as reflected in a lack of statistically significant differences in their use of the structures while completing the tasks given in the pre-test, post-test and delayed post-test (see Chapter 7, Sections 7.2 - 7.6).

Given these findings, this study lends support to, and extends, those reported on by Shintani (2011) and Shintani and R. Ellis (2010). That is, these researchers show that gains resulting from the provision of pre-modified input or interactional feedback in task-based L2 instruction did not differ in their transferability to parallel tasks. Unlike in the present study, however, the transferability of gains in

these earlier L2 studies by Shintani (2011) and Shintani and R. Ellis (2010) was specifically measured in terms of frequency and accuracy in their participants' use of task-relevant L2 forms. In this way, findings of the present study extend those established in this earlier L2 work in suggesting that the transferability of gains resulting from the provision of pre-modified input or interactional feedback was not only evident in the uptake of task-relevant L2 forms, but also in the automaticity of L2 forms.

8.2 Possible Explanations for the Findings

The previous three subsections have established the findings of the study in relation to those established in the earlier work. This section focuses on three primary factors that may have mediated the effectiveness of pre-modified input and interactional feedback in the study. These include (1) the learners' *prior knowledge* of the target L2 form (ECSs); (2) the *relative complexity* of the tasks used in the study; and (3) the *frequency* of which the target L2 form was presented by way of pre-modified input or interactional feedback.

8.2.1 Role of Prior Knowledge of the Target L2 Form

As described in Chapter 6, Section 6.3, the participants in this study were Indonesian adult learners of L2 English who had at least six years of experience learning English by way of a structural syllabus. Hence, it was quite likely that these participants had already had prior knowledge of the target L2 form (ECSs) prior to participating in this study. Indeed, the observation of their pre-test task performances indicated that many of them were able to use the form—albeit less accurately and/or less fluently—to address the functional demand of the task given

in the pre-test (See Chapter 7, Section 7.2). Given such a prior learning profile of the participants, this study *did not* examine whether the provision of pre-modified input or interactional feedback facilitated the development of new knowledge of form-function-meaning mappings, but rather whether these two instructional options facilitated the consolidation of partially-learned knowledge of specific task-relevant L2 forms so it could be accessed more accurately and fluently in spontaneous L2 production (see Chapter 6, Section 6.5).

From a theoretical standpoint, the fact that the participants had already had prior knowledge of ECSs might explain the lack of difference in their gains resulting from the provision of pre-modified input or interactional feedback. That is, such knowledge may have helped them notice the use of the structures in the input or feedback, resulting in no significant differences in their uptake and automaticity of the structures (N. Ellis, 2005, 2011; see Chapter 3 for details). Although the findings in this study did not allow a conclusion to be drawn about such a potential role of the participants' prior knowledge of ECSs in facilitating their noticing of the strucrures in the input or feedback, studies by Park (2011) and Park and Han (2008) provide relevant empirical support for such a conclusion. Park (2011) and Park and Han (2008) examined, *inter alia*, whether L2 learners with and without prior L2 knowledge differed in their abilities to notice specific L2 forms during meaning-focused instruction. They found that learners with prior L2 knowledge were able to notice more input items than those without prior L2 knowledge.

8.2.2 Role of Task Complexity

In addition to prior knowledge of the structures, the effectiveness of pre-modified input and interactional feedback in this study may also have been enhanced by the

complexity of the tasks used (see Robinson, 2007; Robinson & Gilabert, 2007). As noted in Chapter 6, the tasks were relatively complex in their conceptual and functional demands in that they required the participants to describe a picture very clearly and unambiguously to the researcher so that he could identify it from very similar pictures, or, alternatively, to comprehend a series of descriptions of a picture given by the researcher in order that they were able to identify it from very similar pictures. Furthermore, since all items included in the pictures differed only in specific features (e.g., "their relative size, shape, position and orientation"), the use of comparative structures constitutes *useful* linguistic features for these tasks (Lambert, 2019, p. 116, emphasis added; see also Lambert & Nakamura, 2019, for empirical support of the complexity of the tasks used in the current study). In other words, the use of comparative structures could help the participants describe or comprehend specific features of target items very clearly or unambiguously (e.g., *the jacket that I want you to choose <u>looks slightly smaller than</u> the other three jackets*).

As discussed in Chapter 2, Section 2.2.2, the role of task complexity in inducing selective attention to specific task-relevant L2 forms during task work is encapsulated in Robinson's Cognition Hypothesis (Robinson, 2003a, 2003b, 2005b, 2010, 2011a, 2015). This hypothesis suggests that the use of cognitively complex tasks in L2 instruction—like those used in the study—can prompt learners to direct their attentional resources to specific L2 forms that are made available during task work (either through the provision of pre-modified input or interactional feedback), thereby facilitating incidental learning of the forms (for empirical support, see e.g., Robinson, 2007; Révész, 2009; 2011). Note, however, that the Cognition Hypothesis also stipulates a close relationship between *task*

complexity (i.e., how certain elements embedded in tasks are likely to impose certain level of cognitive demands on the part of learners) and *task difficulty* (i.e., how learners actually perceive cognitive demands of given tasks). That is, although certain L2 communication tasks may be designed in ways that manifest certain level of cognitive demands on the part of learners, the learners may perceive such demands differently (i.e., as relatively simple or complex). For this reason, one criticism that may be levelled against the argument presented in this section is that the complexity of the tasks used may not have exerted the same influence on the part of the participants, especially with regard to how these participants were likely to allocate their attentional resources to the target L2 form during task work (see e.g., Foster & Skehan, 1996; Skehan, 2014, Ch. 1).

As noted in Chapter 6, Section 6.3, however, participants in this study were recruited based on their abilities to complete a recruitment task (see Appendix 1) which resembled those used in the testing and treatment sessions (see Appendix 2). That is, only those who were able to complete the recruitment task with effort and, at the same time, were still less capable of using ECSs accurately and fluently were recruited into the study. Given this specific recruitment process, it may thus be assumed that the participants may have perceived the complexity of the tasks in relatively the same way (i.e., as neither too complex as to discourage them, nor too simple as to bore them). Such a perception, *ceteris paribus*, may have provided the same platform for incidental learning of the task-relevant L2 forms (ECSs) as they occurred in the input or feedback (Robinson, 2003a, 2003b, 2005b, 2007, 2010, 2011a, 2015; Robinson & Gilabert, 2007). The complexity of the tasks used in the current study may thus explain a lack of difference in the uptake and automaticity

of ECSs resulting from the provision of pre-modified input or interactional feedback.

8.2.3 Role of Input Frequency

Notwithstanding, incidental learning of ECSs from pre-modified input or interactional feedback as triggered by the complexity of the tasks and/or by learners' prior knowledge of the structures (see Sections 8.2.1 and 8.2.2, above) may be transitory "if there is no repeated experience to build it up further" (Sharwood-Smith, 2008, p. 10). For this reason, *frequency* of exposure to the structures as realized through pre-modified input or interactional feedback may also have mediated the efficacy of these two instructional options. That is, the fact that both pre-modified input and interactional feedback operationalized in this study afforded relatively high frequency of exposure to ECSs during task-based L2 communication, such exposure may have facilitated long-term retention of the structures in the participants' long-term memory (Baddeley, 2003; N. Ellis, 2001; Williams, 2012), making them capable of accessing the structures for use in future contexts (see Chapter 3 for details).

The role of *frequency* in facilitating L2 learning and processing has also been put forward by L2 researchers from distinct theoretical perspectives. Lightbown (2008), for instance, maintains that input frequency facilitates 'transfer-appropriate processing' where L2 forms that recurrently occur in the input or feedback may be retrieved (more) easily in future contexts that approximate those of learning conditions. She also argues that such facilitation is *not* solely determined by the type of processing that occurs on tasks (e.g., receptive vs. productive; aural vs. oral), but rather by the functional demands that the tasks impose on learners (e.g.,

using ECSs to clearly convey differences of similar pictures—see also Larsen-Freeman, 2013b; Lightbown & Segalowitz, 1999, for similar arguments). In the same vein, N. Ellis (2005, 2008, 2019) argues that when particular L2 forms are frequently used in processing meaning in contextualized communication, their association with contextualized use will be strengthened. Hence, since the tasks used to measure gains in the uptake and automaticity of ECSs in the study were similar to those used for providing treatments, it is quite reasonable that such similarities might have fostered the participants' ability to use the structures based on what they have learned from receiving pre-modified input or interactional feedback.

To sum up, the combination of these learner-internal and -external factors (i.e., prior knowledge of ECSs, the complexity of the tasks used, and the frequency of exposure to ECSs in the input or feedback) may explain a lack of differences in the uptake and automaticity of ECSs from the provision of pre-modified input and interactional feedback.

8.3 Implications for Task-Based L2 Learning Research

This section discusses implications of the study for task-based L2 learning research. The focus will be on measurement, namely, how the approach used to measuring uptake and automaticity in the study might contribute to the task-based L2 research paradigm in terms of (i) improving assessments of knowledge of task-relevant L2 forms, and (ii) improving assessment of the automaticity of task-relevant forms.

As noted in Chapter 5, point (i) was measured in terms of uptake (frequency and accuracy) of ECSs, whereas point (ii) was measured in terms of utterance fluency

(speech rate, frequency of filled/unfilled pauses and verbatim repetitions, and frequency of overt self-repairs) during the production of ECSs. The utility of these two sets of measures for capturing development in TBLT performance research is briefly reviewed in the next two subsections.

8.3.1 Developmental Measures of Uptake

As discussed in Chapters 3 and 5, learners' uptake of task-relevant L2 forms within contextualized L2 communication largely reflect improvements in their developing L2 resources. To date, however, research on instructed L2 development have typically employed only specific measures of accuracy to gauge the effects of instructional conditions on L2 development (Norris & Ortega, 2000; Shintani, Li, & R. Ellis, 2003). As discussed in Chapter 5, Section 5.6.1.4, using accuracy data as the only dependent measure in instructed L2 development research cannot provide adequate information as to whether instructional conditions under investigation are truly effective in facilitating L2 development. This is because, as Sanz and Grey (2015, p. 303) note, specific measures of accuracy alone cannot provide adequate evidence as to whether learners are able to access their developing knowledge of L2 forms in relevant L2 communicative contexts (see Chapter 5 for details). Given this limitation of employing only specific measures of accuracy in instructed L2 development research, this study employed specific measures of uptake (frequency and accuracy) to detect improvements in learners' developing knowledge of the target L2 form (ECSs).

Based on the results reported in Chapter 7, Sections 7.2 and 7.3, employing specific measures of uptake has, indeed, proved useful to detect differences in the level of accessibility of the participants' knowledge of the target L2 form (ECSs). For

instance, the results revealed that although the participants in the two experimental and comparison groups *did not* differ in their accuracy of using ECSs in the pre-test, they *did* differ in their frequency of using these structures, suggesting that the level of accessibility of their knowledge of the structures differed even before they received any treatments. Hence, if this study had only employed specific measures of accuracy, such a difference in the level of accessibility of the knowledge of the target L2 form might have remained undetected, leading to the interpretation that the particpants' knowledge of the target L2 form was roughly equal.

8.3.2 Utterance Fluency as a Measure of Automaticity of Task-Relevant L2 Forms

Another important methodological innovation that this study contributes to the field of task-based L2 learning research is that it employed specific measures of L2 utterance fluency as a means to measure gains in automaticity underlying the participants' production of the target L2 form *within the context of naturally-occurring task-based L2 communication*. That is, as discussed in Chapter 5, Section 5.6, although L2 researchers generally agree that the processing features such as 'speed' and 'stability' constitute necessary conditions that distinguish automaticity from mere speed-ups, relatively little is known as to how to measure these two features manifested in learners' production of specific L2 forms during naturally-occurring L2 communication (see Chapter 5 for details). In this study, three measures of L2 utterance fluency which have been commonly used in task-based L2 speech processing research were adapted to measure both the speed and the stability of learners' production of the target L2 form (ECSs) in naturally-occurring task-based L2 communication. These three measures of L2 utterance fluency

were—as noted earlier in this section—speech rate, frequency of filled/unfilled pauses and verbatim repetitions, and frequency of overt self-repairs that occurred during the production of the form. The validity of these three measures of L2 utterance fluency as a measure of L2 automaticity is discussed in detail in Chapter 5 (see also Segalowitz, 2010, 2016; Segalowitz & Freed, 2004).

8.4 Implication for Task-Based L2 Pedagogy

This section discusses pedagogic implications of the study to task-based L2 instruction so that L2 teachers may find the outcomes of the study useful and practical.

As the principles of increasing descriptive demands of tasks by way of increasing referent similarity (like those used in this study) have been discussed in detail in previous work (see e.g., Lambert, 2019; Yule, 1997), the discussion to be presented in the next three subsections focuses on the principles of designing and using pre-modified input and of providing interactional feedback during task work so as to optimize incidental learning of task-relevant L2 forms.

8.4.1 Designing Pre-modified Input Materials

As noted in Chapter 4, the current study employed *elaborated* pre-modified input (i.e., model performances in task-based L2 instruction that have been prepared [written] to contain redundancy, regularity and/or explicitness of thematic structure as a means to improving comprehensibility on the part of L2 learners). Following Long (2015, 2019), the principal motivation for employing elaborated pre-modified input is that such input can not only minimize communication breakdowns due to comprehension difficulty on the part of L2 learners, but also can help learners learn

task-relevant L2 forms incidentally while attempting to comprehend meaning in the input (see Chapter 4 for details). In what follows, guidelines for elaborating input gained from this study are outlined.

1. Input elaboration should preferably be done based on samples of NSs' oral and/or written performances on target tasks. This is because such samples can provide learners with a valuable source of information as to how specific L2 forms are naturally used by NSs to realize meaning and function in target discourse (Long, 2015, 2019). In this study, for instance, input elaboration was done based on samples of English NSs' oral task performances as they communicated with an English NS to complete the same tasks as those used in the study (Lambert, 2019).
2. To elaborate input, teachers can focus on features that are likely to create comprehension difficulty on the part of L2 learners, or those that potentially become the major source of (new) knowledge for the learners. These may include complex sentence structures, low-frequency words, and/or forms that NSs typically use when completing target tasks in target discourse.
3. To elaborate complex sentence structures, teachers can add redundancy, regularity and/or explicitness of thematic structure to improve comprehensibility of the structures and, thus, to help learners learn the underlying construction of the structures incidentally (see Chapter 4, especially Section 4.4, for an illustration as to how this can be achieved).
4. To elaborate low-frequency words, teachers can add synonyms, paraphrases and/or informal (non-dictionary-type) definitions (Long, 2019, p. 5). Alternatively, they can employ arrow-word definitions to provide a visual meaning of target words, as illustrated in Figure 8.1 below (see also Appendix

2). Indeed, this study indicates that this latter option proves useful for helping learners infer meaning of target words efficiently (i.e., without spending too much time questioning the meaning of the words).

5. Finally, specific forms that commonly occur in NSs' task performances may be elaborated through increasing their frequency in the input. As noted in previous chapters, commonly-occurring forms in NSs' task performances usually indicate that such forms constitute useful linguistic features in target discourse. Hence, increasing the frequency of commonly-occurring forms in the input, such as by way of pre-modified input or interactional feedback provided in conjunction with parallel tasks (i.e., tasks that manifest related functional demands—see e.g., Appendix 2 for illustrations), may help learners learn or, at least, grasp the utility of the forms and, thus, potentially lead them to use the forms when attempting to communicate in similar communicative contexts.

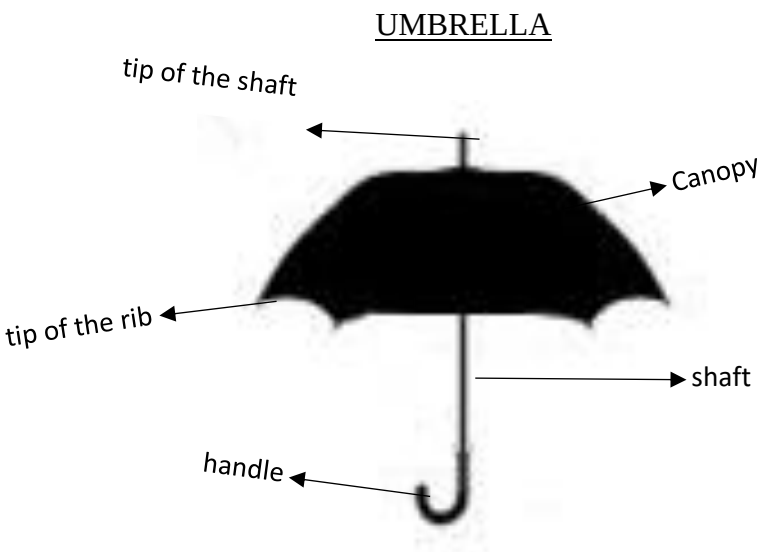


Figure 8.1: An umbrella and its various parts.

8.4.2 Using Pre-modified Input Materials in L2 Classroom

Although the principles for elaborating input—as outlined in the preceding section—can improve comprehensibility on the part of L2 learners, the utility of such input in facilitating incidental learning of form-function-meaning mappings is largely determined by how it is implemented in L2 classroom. In what follows, three methodological principles, originally proposed in Long (2009, 2015, 2019), will be adopted to inform how elaborated pre-modified input may be implemented in L2 classroom so as to foster incidental learning of form-function-meaning mappings. These include (1) presenting the input materials in the form of '*task*', rather than 'text', (2) providing opportunities for *negotiation for meaning*, and (3) providing *multiple contexts* for processing the input.

In line with the definition of 'tasks' outlined in Chapter 1, presenting pre-modified input in the form of task can encourage learners to focus on meaning, rather than on forms (R. Ellis, 2009). This is because tasks allow learners to attend to the input within the context of meaningful L2 communication (Long, 2009, 2015, 2019). In this study, for instance, the pre-modified input materials were presented within the context of information-gap tasks which created a necessity for learners to listen to and comprehend the input in order that they were capable of completing given tasks successfully (i.e., identifying a picture from a set of similar pictures based on descriptions given by the researcher).

In the case where learners fail to comprehend certain parts of the input due to, for instance, a distraction that occurs during listening or due to their limited knowledge of L2 forms that occur in the input, providing them with opportunities to 'negotiate for meaning' might prove beneficial to help them comprehend the input better and,

at the same time, promote incidental learning of form-function-meaning mappings (Long, 2019; see also Shintani, 2011, 2012, 2016, for empirical support).

Next, to help learners consolidate their currently emerging knowledge of form-function-meaning mappings, teachers can provide multiple contexts that allow learners to retrieve the knowledge recurrently within contextualized communication. This may be done by designing parallel tasks and/or creating parallel versions of pre-modified input materials (see Appendix 3, for illustrations). Following this, teachers can ask the learners to complete parallel versions of the tasks orally with their peers. From a pedagogic perspective, getting learners to listen to model task performances (i.e., pre-modified input) and then having them perform relevant L2 communication tasks orally might make instructional activities motivating. Lambert (2019) notes, "learners in L2 classroom are too often motivated only by the fact that it is their turn to speak rather than by the types of physical, social or emotional needs for understanding that motivate authentic language use" (p. 2).

Finally, when learners produce non-target-like or less comprehensible L2 output while completing given tasks orally, the teacher can provide interactional feedback.

8.4.3 Providing Interactional Feedback during Task Work

There are multiple forms of interactional feedback that teachers can use to respond to learners' non-target-like or less comprehensible L2 use. In what follows, guidelines for providing interactional feedback gained from this study are outlined in order to help learners notice gaps in their L2 resources and, thus, encourage uptake.

1. The provision of interactional feedback may be more effective when targeting specific L2 forms, rather than a wide range of L2 forms. That is, interactional feedback which targets specific L2 forms might afford recurrent opportunities for learners to compare their non-target-like use of the forms and the target-like use produced by the teacher.
2. The effectiveness of interactional feedback may be enhanced by having learners perform complex tasks, rather than simple tasks. This is because complex tasks normally impose certain level of cognitive demands that can induce sensitivity to the feedback provided during task work.

8.5 Limitations of the Study and Recommendations for Future Research

There are (at least) three main limitations in this study that need addressing in future research. The first concerns sample size. The relatively small sample size in this study (i.e., 17 in each condition) limits the generalizability of the findings. Furthermore, unequal abilities between the participants in the experimental groups and those in the comparison group made statistical comparisons between these groups inappropriate. Future research may address these issues by pooling a larger sample size.

The second limitation concerns the internal validity. As noted in Chapter 2, Section 2.2, this study was motivated by both theoretical and pedagogic concerns regarding the utility of pre-modified input and interactional feedback in task-based L2 instruction. Given such motivation, the study faced two options: either maintaining the internal validity at the expense of the ecological validity, or vice versa. Maintaining the internal validity typically results in research outcomes that are highly valuable for theoretical advancement, but are usually less informative for

practical purposes. In contrast, maintaining the ecological validity can, in principle, result in research outcomes that are highly valuable for practical purposes, but are usually less valuable for advancing theories (Plonsky, 2017; Revez, 2019). Although the study attempted to achieve both, a greater emphasis was placed on the ecological validity. That is, although the treatments were designed in ways that would potentially allow learners in the experimental conditions to receive comparable amount of exposure to the target L2 form during treatments, it did not strictly control variation in the type/token frequency in the input and feedback. From theoretical standpoint, variation in type/token frequency can affect learning and processing of the target form differently (see Chapter 3). Hence, future L2 research may need to address this issue. One way to do so may be to design pre-modified input materials based on the number of types and tokens which occur in the provision of interactional feedback (see e.g., Denhovska, Serratrice & Payne, 2016; McDonough & Kim, 2009).

Finally, although the procedures of implementing pre-modified input and interactional feedback in this study were described in enough detail in Chapter 6, this study did not document the process features during which the input and feedback were provided during the treatment sessions. Further studies may also need to address this issue. One way to do so may be to design a study like those conducted by Shintani (2011) and Shintani and R. Ellis (2010, 2014).

8.6 Summary

This chapter has provided an overview of the present study and discussed the results of the study in relation to those established in the earlier work. Implications of the study for task-based L2 learning research and pedagogy have also been

outlined. The chapter ends with outlining the limitations of the study and how they may be addressed in future research. Despite the limitations, it is hoped that the results of this study contribute to the advancement of knowledge and practice of task-based L2 instruction.

Chapter 9

Conclusion

The goal of the study was to examine whether the provision of pre-modified input and interactional feedback in task-based L2 instruction are effective in promoting uptake (more frequent and accurate use) as well as automaticity (fluent use) of a specific task-relevant L2 form (ECSs). In doing so, the study employed a pre-test, post-test, delayed post-test design with two experimental groups (pre-modified input and interactional feedback) and one comparison group. Findings of the study indicate that pre-modified input and interactional feedback had comparable effects on promoting uptake and automaticity of ECSs.

Despite these positive results, the study raises important questions that need addressing in future research. Firstly, the comparable findings on pre-modified input and interactional feedback in promoting uptake and automaticity of the target L2 form (ECSs) were established in a laboratory-type experiment. Hence, it remains to be seen whether these findings also hold in real classroom settings. It has been suggested that laboratory-type experiments differ from classroom settings in many respects (Foster, 1998; Foster & Ohta, 2005; cf., Gass, Mackey & Ross-Feldman, 2011). Thus, investigating the efficacy of pre-modified input and interactional feedback in real classroom settings may prove beneficial to determine the generalizability of the current findings into real L2 instructional contexts.

Furthermore, although the present study considers learner-internal and -external factors (e.g., prior knowledge, task complexity, input or feedback frequency) in explaining the lack of difference in learning gains resulting from pre-modified

input and interactional feedback, the account provided was mainly speculative or theoretical. Thus, further studies are needed to investigate whether these learner-internal and -external factors do mediate the efficacy of pre-modified input or interactional feedback in task-based L2 instruction.

Despite these shortcomings, it is hoped that the findings established in the current study contribute to the accumulation of knowledge of task-based L2 research and pedagogy and, at the same time, spark interest among L2 researchers to conduct a related line of investigation.

References

Achard, M., & Niemeier, S. (Eds.). (2008). *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching.* Berlin: Walter de Gruyter.

Adams, R. (2003). L2 Output, Reformulation and Noticing: Implications for IL Development. *Language Teaching Research*, *7*(3), 347-376.

Ahmadian, M. (2011). The Effect of 'Massed' Task Repetitions on Complexity, Accuracy and Fluency: Does It Transfer to a New Task?. *The Language Learning Journal*, *39*(3), 269-280.

Ahmadian, M., & Tavakoli, M. (2011). The Effects of Simultaneous Use of Careful Online Planning and Task Repetition on Accuracy, Complexity, and Fluency in EFL Learners' Oral Production. *Language Teaching Research*, *15*(1), 35-59.

Ahmadian, M., Tavakoli, M., & Dastjerdi, H. (2015). The Combined Effects of Online Planning and Task Structure on Complexity, Accuracy and Fluency of L2 Speech. *The Language Learning Journal*, *43*(1), 41-56.

Allwood, J. (2003). Meaning Potentials and Context: Some Consequences for the Analysis of Variation in Meaning. In H. Cuyckens, R. Dirven, & J. Taylor (Eds.), *Cognitive Approaches to Lexical Semantics* (pp. 29–65). Berlin: Mouton de Gruyter.

Ammar, A., & Hassan, R. (2018). Talking It through: Collaborative Dialogue and Second Language Learning. *Language Learning*, *68*(1), 46-82.

Ammar, A., & Spada, N. (2006). One Size Fits All?: Recasts, Prompts, and L2 Learning. *Studies in Second Language Acquisition*, *28*(4), 543-574.

Anderson A., & Lynch, T. (1988). *Listening*. Oxford: Oxford University Press.

Anderson, J. (1992). Automaticity and the ACT* Theory. *American Journal of Psychology*, *105*(2), 165-180.

Anderson, J. (1993). *Rules of the Mind*. Hillsdale, NJ: Erlbaum.

Anderson, J., Fincham, J., & Douglass, S. (1997). The Role of Examples and Rules in the Acquisition of a Cognitive Skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 932-945.

Arnon, I., & Snider, N. (2010). More than Words: Frequency Effects for Multi-Word Phrases. *Journal of Memory and Language*, *62*(1), 67-82.

Ayoun, D. (2001). The Role of Negative and Positive Feedback in the Second Language Acquisition of the Passé Composé and Imparfait. *The Modern Language Journal*, *85*(2), 226-243.

Azkari, A., & Mayo, M. (2015). Task-Modality and L1 Use in EFL Oral Interaction. *Language Teaching Research*, *19*(5), 550-571.

Baker, N., & Nelson, K. (1984). Recasting and Related Conversational Techniques for Triggering Syntactic Advances by Young Children. *First Language*, *5*(13), 3-21.

Bao, M., Egi, T., & Han, Y. (2011). Classroom Study on Noticing and Recast Features: Capturing Learner Noticing with Uptake and Stimulated Recall. *System*, *39*(2), 215-228.

Baddeley, A. (2003). Working Memory and Language: An Overview. *Journal of Communication Disorders*, *36*(3), 189-208.

Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-Based Models of Language*. Stanford, CA: CSLI Publications.

Barsalou, L. (2003). Situated Simulation in the Human Conceptual System. *Language and Cognitive Processes*, *18*(5-6), 513-562.

Barsalou, L. (2008). Grounded Cognition. *Annu. Rev. Psychol.*, *59*, 617-645.

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney, B. & E. Bates (Eds.). *The Crosslinguistic Study of Sentence Processing* (pp. 3-76). New York: Cambridge University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M.H., Croft, W., Ellis, N.C., Holland, J., Ke, J., Larsen-Freeman, D. and Schoenemann, T. (2009), Language Is a Complex Adaptive System: Position Paper. *Language Learning*, *59*(s1).1-26.

Behrens, H. (2009). Usage-Based and Emergentist Approaches to Language Acquisition. *Linguistics*, *47*(2), 383-411.

Bialystok, E. (2011). How Analysis and Control Lead to Advantages and Disadvantages in Bilingual Processing. In C. Sanz, & R. Leow (Eds.), *Implicit and Explicit Language Learning: Conditions, Processes, and Knowledge In SLA and Bilingualism* (pp. 49-58). Washington, DC: Gerorgetown University Press.

Bird, S. (2012). Expert Knowledge, Distinctiveness, and Levels of Processing in Language Learning. *Applied Psycholinguistics*, *33*(4), 665-689.

Boersma, P., & Weenik, D. (2017). *Praat: Doing phonetics by computer, 6.0.29*. Amsterdam: University of Amsterdam.

Bohannon, J., & Stanowicz, L. (1988). The Issue of Negative Evidence: Adult Responses to Children's Language Errors. *Developmental Psychology*, *24*(5), 684.

Bowden, H. (2016). Assessing Second-Language Oral Proficiency for Research: The Spanish Elicited Imitation Task. *Studies in Second Language Acquisition*, *38*(4), 647-675.

Bowles, M. (2011). Measuring Implicit and Explicit Linguistic Knowledge: What can Heritage Language Learners Contribute?. *Studies in Second Language Acquisition*, *33*(2), 247-271.

Boyd, J., & Goldberg, A. (2009). Input Effects within a Constructionist Framework. *The Modern Language Journal*, *93*(3), 418-429.

Branigan, H., Pickering, M., & Cleland, A. (2000). Syntactic Co-ordination in Dialogue. *Cognition*, *75*(2), 13-25.

Brooks, P., Braine, M., Catalano, L., Brody, R., & Sudhalter, V. (1993). Acquisition of Gender-Like Noun Subclasses in an Artificial Language: The Contribution of Phonological Markers to Learning. *Journal of Memory and Language*, *32*(1), 76-95.

Bui, G., & Skehan, P. (2018). Complexity, Fluency and Accuracy. In J. Liontas (Ed.), *the TESOL Encyclopedia of English Language Teaching* (pp. 1-7). Hoboken, NJ: Wiley-Blackwell.

Butler, Y. (2011). The Implementation of Communicative and Task-Based Language Teaching in the Asia-Pacific Region. *Annual Review of Applied Linguistics*, *31*, 36-57.

Butler, Y., & Zeng, W. (2014). Young Foreign Language Learners' Interactions during Task-Based Paired Assessments. *Language Assessment Quarterly*, *11*(1), 45-75.

Bybee, J. (2003). Mechanisms of Change in Grammaticization: The Role of Frequency. In B. Joseph & R. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 602–23). Oxford: Blackwell.

Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, *82*(4), 711-733.

Bybee, J. (2008). Usage-Based Grammar and Second Language Acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 216–236). New York: Routledge.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam: Benjamins.

Bygate, M. (1999). Task as Context for the Framing, Reframing and Unframing of Language. *System*, *27*(1), 33-48.

Bygate, M. (2001). Effects of Task Repetition on the Structure and Control of Oral Language. In M. Bygate, P. Skehan, & M. Swain (Eds.), Researching Pedagogic Tasks (pp. 23–48). Harlow: Longman.

Bygate, M. (2016). Sources, Developments and Directions of Task-Based Language Teaching. *The Language Learning Journal*, *44*(4), 381-400.

Bygate, M., Skehan, P., & Swain, M. (2001). *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. London: Routledge.

Cadierno, T., & Eskildsen, S. (Eds.). (2015). *Usage-Based Perspectives on Second Language Learning*. Berlin: de Gruyter Mouton.

Cameron, L., & Larsen-Freeman, D. (2007). Complex Systems and Applied Linguistics. *International Journal of Applied Linguistics*, *17*(2), 226-239.

Carpenter, H., Jeon, S., Macgregor, D., & Mackey, A. (2006). Learners' Interpretations of Recasts. *Studies in Second Language Acquisition, 28(*2), 209–236.

Chater, N., & Christiansen, M. H. (2018). Language Acquisition as Skill Learning. *Current Opinion in Behavioral Sciences*, *21*, 205-208.

Christiansen, M., & Arnon, I. (2017). More than Words: The Role of Multiword Sequences in Language Learning and Use. *Topics in Cognitive Science*, *9*(3), 542-551.

Christiansen, M., & Chater, N. (1994). Generalization and Connectionist Language Learning. *Mind and Language*, *9*(3), 273–287.

Christiansen, M., & Chater, N. (1999). Toward a Connectionist Model of Recursion in Human Linguistic Performance. *Cognitive Science*, *23*(2), 157-205.

Christiansen, M., & Chater, N. (2016). The Now-or-Never Bottleneck: A Fundamental Constraint on Language. *Behavioral and Brain Sciences*, *39*, E62.

Clark, H. (1969). Linguistic Processes in Deductive Reasoning. *Psychological Review*, *76*(4), 387-404.

Clark, H. (1970). Comprehending Comparatives. In G. Flores d'Arcais & W. Levelt (Eds.), *Advances in Psycholinguistics* (pp. 294-321). Amsterdam: North Holland Press.

Clark, H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H., & Card, S. (1969). Role of Semantics in Remembering Comparative Sentences. *Journal of Experimental Psychology*, *82*(3), 545-553.

Clark, H., & Clark, E. (1977). *Psychology and Language: An Introduction to Psycholinguistics.* New York: Harcourt Brace Jovanovich.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Mahwah, NJ: Lawrence Erlbaum.

Collins, P. (1994). The Structure of English Comparative Clauses, *English Studies*, *75*(2), 157-165.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some Input on the Easy/Difficult Grammar Question: An Empirical Study. *The Modern Language Journal*, *93*(3), 336-353.

Coughlan, P., & Duff, P. (1994). Same Task, Different Activities: Analysis of an SLA Task from an Activity Theory Perspective. In J. Lantolf, & G. Appel (Eds.), *Vygotskian Approaches to Second Language Research* (pp. 157-172). Norwood, NJ: Ablex.

Croft, W. & Cruise, A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Crossley, S., Salsbury, T., McNamara, D., & Jarvis, S. (2011). Predicting Lexical Proficiency in Language Learner Texts Using Computational Indices. *Language Testing*, *28*(4), 561-580.

Crossley, S. A., Skalicky, S., Kyle, K., & Monteiro, K. (2019). Absolute Frequency Effects in Second Language Lexical Acquisition. *Studies in Second Language Acquisition*, *41*(4), 721-744.

Crossley, S., Subtirelu, N., & Salsbury, T. (2013). Frequency Effects or Context Effects in Second Language Word Learning: What Predicts Early Lexical Production?. *Studies in Second Language Acquisition*, *35*(4), 727-755.

Culicover, P., Jackendoff, R., & Audring, J. (2017). Multiword Constructions in the Grammar. *Topics in Cognitive Science*, *9*(3), 552-568.

Davies, A. (2003). *The Native Speaker: Myth and Reality*. Multilingual Matters.

de Bot, K. (1992). A Bilingual Production Model: Levelt's' Speaking' Model Adapted. *Applied Linguistics, 13*, 1–24.

de Bot, K., Lowie, W., & Verspoor, M. (2005). *Second Language Acquisition: An Advanced Resource Book*. London: Routledge.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory Approach to Second Language Acquisition. *Bilingualism: Language and Cognition*, *10*(1), 7–21.

de Graaff, R., & Housen, A. (2009). Investigating the Effects and Effectiveness of L2 Instruction. In M. H. Long & C. J. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 726 – 755). Oxford: Blackwell.

de Jong, N. (2016). Predicting Pauses in L1 and L2 Speech: The Effects of Utterance Boundaries and Word Frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 113-132.

de Jong, N., & Bosker, H. (2013). Choosing a Threshold for Silent Pauses to Measure Second Language Fluency. In R. Eklund (Ed.), *Proceedings of Disfluency in Spontaneous Speech* (pp. 17–20). DiSS 2013.

de Jong, N., & Wempe, T. (2009). PRAAT Script to Detect Syllable Nuclei and Measure Speech Rate Automatically. *Behavior Research Methods*, *41*(2), 385-390.

de Jong, N., Groenhout, R., Schoonen, R., & Hulstijn, J. (2015). Second Language Fluency: Speaking Style or Proficiency? Correcting Measures of Second Language Fluency for First Language Behavior. *Applied Psycholinguistics*, *36*(2), 223-243.

de Jong, N., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2013). Linguistic Skills and Speaking Fluency in a Second Language. *Applied Psycholinguistics*, *34*(5), 893-916.

de la Fuente, M. (2002). Negotiation and Oral Acquisition of L2 Vocabulary: The Roles of Input and Output in the Receptive and Productive Acquisition of Words. *Studies in Second Language Acquisition*, *24*(1), 81-112.

Denhovska, N., Serratrice, L., & Payne, J. (2016). Acquisition of Second Language Grammar under Incidental Learning Conditions: The role of Frequency and Working Memory. *Language Learning*, *66*(1), 159-190.

DeKeyser, R. (1996). Exploring Automatization Processes. *TESOL Quarterly*, *30*(2), 349-357.

DeKeyser, R. (1997). Beyond Explicit Rule Learning: Automatizing Second Language Morphosyntax. *Studies in Second Language Acquisition*, *19*(2), 195-221.

DeKeyser, R. (1998) Beyond Focus on Form: Cognitive Perspectives on Learning and Practicing Second Language Grammar. In C. Doughty and J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 42–63). Cambridge: Cambridge University Press.

DeKeyser, R. (2001). Automaticity and Automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 125–151). New York: Cambridge University Press.

DeKeyser, R. (2003). Implicit and Explicit Learning. In C. J. Doughty & M. H. Long (Eds). *The Handbook of Second Language Acquisition* (pp. 313-348). Oxford: Blackwell.

DeKeyser, R. M. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Language Learning*, *55*(S1), 1-25.

Dekeyser, R. (2007). Situating the Concept of Practice. In R. Dekeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 1–18). Cambridge: Cambridge University Press.

DeKeyser, R. (2015). Skill Acquisition Theory. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd Ed.), (pp. 94–112). New York, NY: Routledge.

DeKeyser, R. (2017). Knowledge and Skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 15–32). New York: Routledge.

DeKeyser, R., & Botana, G. (2015). The Effectiveness of Processing Instruction in L2 Grammar Acquisition: A Narrative Review. *Applied Linguistics*, *36*(3), 290-305.

DeKeyser, R., & Sokalski, K. (1996). The Differential Role of Comprehension and Production Practice. *Language Learning*, *46*(4), 613-642.

DeKeyser, R., Salaberry, R., Robinson, P., & Harrington, M. (2002). What Gets Processed in Processing Instruction? A Commentary on Bill Vanpatten's "Processing Instruction: An Update". *Language Learning*, *52*(4), 805-823.

Derwing, T., Munro, M., Thomson, R., & Rossiter, M. (2009). The Relationship between L1 Fluency and L2 Fluency Development. *Studies in Second Language Acquisition*, *31*(4), 533-557.

Dingemanse, M., Blasi, D., Lupyan, G., Christiansen, M., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603-615.

Doughty, C. (1991). Second Language Instruction Does Make a Difference: Evidence from an Empirical Study of SL Relativization. *Studies in Second Language Acquisition*, *13*(4), 431-469.

Doughty, C. (1994) Fine-Tuning of Feedback by Competent Speakers to Language Learners. In J. Alatis (Ed.) *Georgetown University Round Table on Language and Linguistics*. Washington, DC: Georgetown University Press.

Doughty, C. (2001). Cognitive Underpinnings of Focus on Form. In P. Robinson (Ed.), *Cognition and Second Language Instruction*, (pp. 206-257). New York: Cambridge University Press.

Doughty, C. J., & Varela, E. (1998). Communicative Focus on Form. In C. Doughty & J. Williams (Eds.), *Focus on Form in Classroom Second Language Acquisition* (pp. 114–138). New York: Cambridge University Press.

Doughty, C., & Williams, J. (Eds.) (1998). *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.

Edwards, C., & Willis, J. (Eds.). (2005). *Teachers Exploring Tasks in English Language Teaching*. Basingstoke: Palgrave Macmillan.

Egi, T. (2007). Interpreting Recasts as Linguistic Evidence: The Roles of Linguistic Target, Length, and Degree of Change. *Studies in Second Language Acquisition*, *29*(4), 511-537.

Egi, T. (2010). Uptake, Modified Output, and Learner Perceptions of Recasts: Learner Responses as Language Awareness. *The Modern Language Journal*, *94*(1), 1-21.

Ellis, N. (1993). Rules and Instances in Foreign Language Learning: Interactions of Explicit and Implicit Knowledge. *European Journal of Cognitive Psychology*, *5*(3), 289-318.

Ellis, N. (2001). Memory for Language. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 33–68). New York: Cambridge University Press.

Ellis, N. (2002). Frequency Effects in Language Processing: A Review with Implications for Theories of Implicit and Explicit Language Acquisition. *Studies in Second Language Acquisition*, *24*(2), 143-188.

Ellis, N. (2003). Constructions, Chunking, and Connectionism: The Emergence of Second Language Structure. In C. Doughty & M. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 63–103). Oxford: Blackwell.

Ellis, N. (2005). At the Interface: Dynamic Interactions of Explicit and Implicit Language Knowledge. *Studies in Second Language Acquisition*, *27*(2), 305-352.

Ellis, N. (2006). Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning. *Applied Linguistics*, *27*(2), 164-194.

Ellis, N. (2008). Usage-Based and Form-Focused Language Acquisition: The Associative Learning of Constructions, Learned-Attention, and The Limited L2 Endstate. In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 372–405). London: Routledge.

Ellis, N. (2009a). Optimizing the Input: Frequency and Sampling in Usage-Based and Form-Focussed Learning. In M. Long & C. Doughty (Eds.), *Handbook of Second and Foreign Language Teaching* (pp. 139-158). Oxford: Blackwell.

Ellis, N. (2009b). The Psycholinguistics of the Interaction Approach. In A. Mackey & C. Polio (Eds.), *Multiple Perspectives on Interaction: Second Language Research in Honor of Susan M. Gass* (11–40). New York: Routledge.

Ellis, N. (2011). Implicit and Explicit SLA and Their Interface. In C. Sanz & R. Leow (Eds.), *Implicit and Explicit Language Learning: Conditions, Processes, and Knowledge in SLA and Bilingualism* (pp. 35–47). Washington, DC: Georgetown University Press.

Ellis, N. (2013). Second Language Acquisition. In G. Trousdale & T. Hoffmann (Eds.), *Oxford Handbook of Construction Grammar* (pp. 365–378). Oxford, UK: Oxford University Press.

Ellis, N. (2015). Cognitive and Social Aspects of Learning from Usage. In T. Cadierno & S. Eskildsen (Eds.), *Usage-Based Perspectives on Second Language Learning* (pp. 49–73). Berlin, Germany: Degruyter Mouton.

Ellis, N. (2019). Essentials of a Theory of Language Cognition. *The Modern Language Journal*, *103*(s1), 39-60.

Ellis, N., & Collins, L. (2009). Input and Second Language Acquisition: The Roles of Frequency, Form, and Function—Introduction to the Special Issue. *The Modern Language Journal*, *93*(3), 329-335.

Ellis, N., & Larsen-Freeman, D. (2006). Language Emergence: Implications for Applied Linguistics—Introduction to the Special Issue. *Applied Linguistics*, *27*(4), 558-589.

Ellis, N., & Larsen-Freeman, D. (2009). Constructing a Second Language: Analyses and Computational Simulations of the Emergence of Linguistic Constructions from Usage. *Language Learning*, *59*, 90-125.

Ellis, N., & Ogden, D. (2017). Thinking about Multiword Constructions: Usage-Based Approaches to Acquisition and Processing. *Topics in Cognitive Science*, *9*(3), 604-620.

Ellis, N., & Sagarra, N. (2010). Learned Attention Effects in L2 Temporal Reference: The First Hour and the Next Eight Semesters. *Language Learning*, *60*(s2), 85-108.

Ellis, N., & Sagarra, N. (2011). Learned Attention in Adult Language Acquisition: A Replication and Generalization Study and Meta-Analysis. *Studies in Second Language Acquisition*, *33*(4), 589-624.

Ellis, N., & Schmidt, R. (1998). Rules or Associations in the Acquisition of Morphology? The Frequency by Regularity Interaction in Human and PDP Learning of Morphosyntax. *Language and Cognitive Processes*, *13*(2-3), 307-336.

Ellis, N., & Sinclair, S. (1996). Working Memory in the Acquisition of Vocabulary and Syntax: Putting Language in Good Order. *Quarterly Journal of Experimental Psychology, 49a(1)*, 234–250.

Ellis, N., & Wulff, S. (2015). Usage-Based Approaches in Second Language Acquisition. In B. Vanpatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd Ed.), (pp. 75–93). New York, NY: Routledge.

Ellis, N., O'Donnell, M., & Römer, U. (2013). Usage-Based Language: Investigating the Latent Structures that Underpin Acquisition. *Language Learning*, *63*, 25-51.

Ellis, N., O'Donnell, J., & Romer, U. (2015). Usage-Based Language Learning. In B. MacWhinney & W. O'Grady (Eds.), *The Handbook of Language Emergence* (pp. 163–180). New York: Wiley.

Ellis, R. (1989). Are Classroom and Naturalistic Acquisition the Same?: A Study of the Classroom Acquisition of German Word Order Rules. *Studies in Second Language Acquisition*, *11*(3), 305-328.

Ellis, R. (1990). A Response to Gregg. *Applied Linguistics*, *11*(4), 384-91.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R. (1995). Modified Oral Input and the Acquisition of Word Meanings. *Applied Linguistics*, *16*(4), 409-441.

Ellis, R. (1997). *SLA Research and Language Teaching*. Oxford: Oxford University Press.

Ellis, R. (2000). Task-Based Research and Language Pedagogy. *Language Teaching Research*, *4*(3), 193-220.

Ellis, R. (2001). Non-Reciprocal Tasks, Comprehension and Second Language Acquisition. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching Pedagogic Tasks* (pp. 49–74). Harlow: Pearson Education.

Ellis, R. (2002). Does Form-Focused Instruction Affect the Acquisition of Implicit Knowledge?: A Review of the Research. *Studies in Second Language Acquisition*, *24*(2), 223-236.

Ellis, R. (2003). *Task-Based Language Learning and Teaching*. Oxford: Oxford University Press.

Ellis, R. (2005). Measuring Implicit and Explicit Knowledge of a Second Language: A Psychometric Study. *Studies in Second Language Acquisition*, *27*(2), 141-172.

Ellis, R. (2006). Modelling Learning Difficulty and Second Language Proficiency: The Differential Contributions of Implicit and Explicit Knowledge. *Applied Linguistics*, *27*(3), 431-463.

Ellis, R. (2009). Task-Based Language Teaching: Sorting out the Misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221-246.

Ellis, R. (2016). Focus on Form: A Critical Review. *Language Teaching Research*, *20*(3), 405-428.

Ellis, R. (2017). Position Paper: Moving Task-Based Language Teaching Forward. *Language Teaching*, *50*(4), 507-526.

Ellis, R. (2018a). Meta-Analysis in Second Language Acquisition Research. *Journal of Second Language Studies*, *1*(2), 231-253.

Ellis, R. (2018b). *Reflections on Task-Based Language Teaching*. Bristol: Multilingual Matters.

Ellis, R. (2018c). Towards a Modular Language Curriculum for Using Tasks. *Language Teaching Research*, *Online First*, 1-22.

Ellis, R., & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press.

Ellis, R., & He, X. (1999). The Roles of Modified Input and Output in the Incidental Acquisition of Word Meanings. *Studies in Second Language Acquisition*, *21*(2), 285-301.

Ellis, R., & Roever, C. (2018). The Measurement of Implicit and Explicit Knowledge. *The Language Learning Journal*, 1-16.

Ellis, R., Basturkmen, H., & Loewen, S. (2001). Learner Uptake in Communicative ESL Lessons. *Language Learning*, *51*(2), 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*(2), 339-368.

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-Based Language Teaching: Theory and Practice*. Cambridge: Cambridge University Press.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom Interaction, Comprehension, and the Acquisition of L2 Word Meanings. *Language Learning*, *44*(3), 449-491.

Elman, J. (1999). The Emergence of Language: A Conspiracy Theory. Macwhinney, B. (Ed.), *The Emergence of Language* (pp. 1-27). London: Lawrence Erlbaum.

Erlam, R. (2006). Elicited Imitation as a Measure of L2 Implicit Knowledge: An Empirical Validation Study. *Applied Linguistics*, *27*(3), 464-491.

Erlam, R. (2019). Input-Based Tasks in the French Language Classroom. In Z. Wen & M. Ahmadian (Eds.), *Researching L2 Task Performance and Pedagogy: In Honour of Peter Skehan* (pp. 229-251). Amsterdam/Philadelphia: John Benjamins.

Erlam, R., & Ellis, R. (2018). Input-Based Tasks for Beginner-Level Learners: An Approximate Replication and Extension of Erlam & Ellis (2018). *Language Teaching, First View,* 1-22.

Erlam, R., & Loewen, S. (2010). Implicit and Explicit Recasts in L2 Oral French Interaction. *Canadian Modern Language Review*, *66*(6), 877-905.

Eskildsen, S. (2008). Constructing another Language—Usage-Based Linguistics in Second Language Acquisition. *Applied Linguistics*, *30*(3), 335-357.

Eskildsen, S. (2015). What Counts as a Developmental Sequence? Exemplar-Based L2 Learning of English Questions. *Language Learning*, *65*(1), 33-62.

Ewoldsen, D. (2017). Measurement of Cognition. In J. Matthes, C. Davis & R. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1-10). Hoboken, NJ: Blackwell Publishing.

Farrar, M. (1990). Discourse and the Acquisition of Grammatical Morphemes. *Journal of Child Language*, *17*(3), 607-624.

Fenk-Oczlon, G. (2001). Familiarity, Information Flow, and Linguistic Form. In In J. Bybee & P. Hopper (Eds)., *Frequency and the Emergence of Linguistic Structure* (pp. 431–448). Amsterdam: John Benjamins.

Ferreira, V., & Bock, K. (2006). The Functions of Structural Priming. *Language and Cognitive Processes*, *21*(7-8), 1011-1029.

Foley, J. (1991). A Phycholinguistic Framework for Task-Based Approaches to Language Teaching. *Applied Linguistics*, *12*(1), 62-75.

Fortune, A., & Thorp, D. (2001). Knotted and Entangled: New Light on the Identification, Classification and Value of Language Related Episodes in Collaborative Output Tasks. *Language Awareness*, *10*(2-3), 143-160.

Foster, P. (1998). A Classroom Perspective on the Negotiation of Meaning. *Applied Linguistics*, *19*(1), 1-23.

Foster, P., & Ohta, A. (2005). Negotiation for Meaning and Peer Assistance in Second Language Classrooms. *Applied Linguistics*, *26*(3), 402-430.

Foster, P., & Skehan, P. (1996). The Influence of Planning and Task Type on Second Language Performance. *Studies in Second Language Acquisition*, *18*(3), 299-323.

Foster, P., & Skehan, P. (1999). The Influence of Source of Planning and Focus of Planning on Task-Based Performance. *Language Teaching Research*, *3*(3), 215-247.

Foster, P., & Wigglesworth, G. (2016). Capturing Accuracy in Second Language Performance: The Case for a Weighted Clause Ratio. *Annual Review of Applied Linguistics*, *36*, 98-116.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring Spoken Language: A Unit for All Reasons. *Applied Linguistics*, *21*(3), 354-375.

Fotos, S., & Ellis, R. (1991). Communicating about Grammar: A Task-Based Approach. *TESOL Quarterly*, *25*(4), 605-628.

Gass, S. (1989). How do Learners Resolve Linguistic Conflicts? In S. Gass & J. Schachter (Eds.) *Linguistic Perspectives on Second Language Acquisition* (pp. 183-200). Cambridge: Cambridge University Press.

Gass, S. (1997). *Input, Interaction and the Second Language Learner*. Mahwah, NJ: Lawrence Erlbaum.

Gass, S. (2003). Input and Interaction. In C. Doughty & M. Long (Eds.), *the Handbook of Second Language Acquisition* (pp. 224-256). Malden: Blackwell.

Gass, S., & Mackey, A. (2002). Frequency Effects and Second Language Acquisition: A Complex Picture?. *Studies in Second Language Acquisition*, *24*(2), 249-260.

Gass, S., & Mackey, A. (2015). Input, Interaction and Output in Second Language Acquisition. In J. Williams & B. VanPatten (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd Ed.), (pp. 180–206). New York, NY: Routledge.

Gass, S., & Torres, M. (2005). Attention When?: An Investigation of the Ordering Effect of Input and Interaction. *Studies in Second Language Acquisition*, *27*(1), 1-31.

Gass, S., & Varonis, E. M. (1994). Input, Interaction, and Second Language Production. *Studies in Second Language Acquisition*, *16*(3), 283-302.

Gass, S., Mackey, A., & Pica, T. (1998). The Role of Input and Interaction in Second Language Acquisition: Introduction to the Special Issue. *The Modern Language Journal*, *82*(3), 299-307.

Gass, S., Mackey, A., & Ross-Feldman, L. (2011). Task-Based Interactions in Classroom and Laboratory Settings. *Language Learning*, *61*(s1), 189-220.

Gathercole, S., Willis, C., Emslie, H., & Baddeley, A. (1991). The Influences of Number of Syllables and Wordlikeness on Children's Repetition of Nonwords. *Applied Psycholinguistics*, *12*(3), 349-367.

Givón, T. (1985). Function, Structure, and Language Acquisition. In D. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (pp. 1008–1025). Hillsdale, NJ: Erlbaum.

Givón, T. (1989). *Mind, Code and Context: Essays In Pragmatics.* Hillsdale, NJ: Erlbaum.

Givón, T. (1995). *Functionalism and Grammar*, Amsterdam: John Benjamins.

Godfroid, A. (2016). The Effects of Implicit Instruction on Implicit and Explicit Knowledge Development. *Studies in Second Language Acquisition*, *38*(2), 177-215.

Godfroid, A., & Spino, L. (2015). Reconceptualizing Reactivity of Think-Alouds and Eye Tracking: Absence of Evidence Is not Evidence of Absence. *Language Learning*, *65*(4), 896-928.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: University of Chicago Press.

Goldberg, A. (2003). Constructions: A New Theoretical Approach to Language. *Trends in Cognitive Sciences*, *7*(5), 219-224.

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, A., & Casenhiser, D. (2008). Construction Learning and Second Language Acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 197–215). New York: Routledge.

Goldberg, A., Casenhiser, D., & Sethuraman, N. (2004). Learning Argument Structure Generalizations. *Cognitive Linguistics*, *15*(3), 289-316.

Goldschneider, J., & Dekeyser, R. (2001). Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-Analysis of Multiple Determinants. *Language Learning*, *51*(1), 1-50.

Goldschneider, J., & Dekeyser, R. (2005). Explaining the "Natural Order of L2 Morpheme Acquisition" In English: A Meta-Analysis of Multiple Determinants. *Language Learning*, *55*(S1), 27-77.

Goo, J., & Mackey, A. (2013). The Case against the Case against Recasts. *Studies in Second Language Acquisition*, *35*(1), 127-165.

Gooch, R., Saito, K., & Lyster, R. (2016). Effects of Recasts and Prompts on L2 Pronunciation Development: Teaching English /ɹ/ to Korean Adult EFL Learners. *System*, *60*, 117-127.

Granena, G., & Long, M. (2013). Age of Onset, Length of Residence, Language Aptitude, and Ultimate L2 Attainment in Three Linguistic Domains. *Second Language Research*, *29*(3), 311-343.

Gregg, K. (2001). Learnability and Second Language Acquisition Theory. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 152–180). Cambridge: Cambridge University Press.

Gregg, K. R. (2003). The State of Emergentism in Second Language Acquisition. *Second Language Research*, *19*(2), 95-128.

Grey, S., Sanz, C., Morgan-Short, K., & Ullman, M. (2018). Bilingual and Monolingual Adults Learning an Additional Language: ERPs Reveal Differences in Syntactic Processing. *Bilingualism: Language and Cognition*, *21*(5), 970-994.

Gries, S., & Ellis, N. (2015). Statistical Measures for Usage-Based Linguistics. *Language Learning*, *65*(S1), 228-255.

Haastrup, K., & Henriksen, B. (2000). Vocabulary Acquisition: Acquiring Depth of Knowledge through Network Building. *International Journal of Applied Linguistics*, *10*(2), 221-240.

Han, Z. (2002). A Study of the Impact of Recasts on Tense Consistency in L2 Output. *TESOL Quarterly*, *36*(4), 543-572.

Han, Z., & Kang, E. (2018). Revisiting the Cognition Hypothesis: Bridging a Gap Between the Conceptual and the Empirical. *International Journal of Applied Linguistics*, *28*(3), 391-405.

Harley, B. (1992). Patterns of Second Language Development in French Immersion. *Journal of French Language Studies*, *2*(2), 159-183.

Harley, B., & Swain, M. (1984). The Interlanguage of Immersion Students and Its Implications for Second Language Teaching. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 291–311). Edinburgh: Edinburgh University Press.

Heift, T. (2004). Corrective Feedback and Learner Uptake in CALL. *ReCALL*, *16*(2), 416-431.

Herdina, P., & Jessner, U. (2002). *A Dynamic Model of Multilingualism: Perspectives of Change in Psycholinguistics*. Clevedon, UK: Multilingual Matters.

Hopper, P. (1998). Emergent Grammar. In M. Tomassello (Ed.), *The New Psychology of Language* (pp. 155–175). Mahwah, NJ: Lawrence Erlbaum.

Horst, M., & Collins, L. (2006). From *Faible* to Strong: How does Their Vocabulary Grow?. *Canadian Modern Language Review*, *63*(1), 83-106.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a Clockwork Orange: Acquiring Second Language Vocabulary through Reading. *Reading in a Foreign Language*, *11*(2), 207-223.

Howell, P., & Au-Yeung, J. (2002). The EXPLAN Theory of Fluency Control Applied to the Diagnosis of Stuttering. In E. Fava (Ed.), *Current Issues in*

Linguistic Theory Series: Pathology and Therapy of Speech Disorders (pp. 75–94). Amsterdam: John Benjamins.

Hui, B., & Godfroid, A. (2020). Testing the Role of Processing Speed and Automaticity in Second Language Listening. *Applied Psycholinguistics, FirstView*, 1-27.

Hulstijn, J. (1989). Implicit and Incidental Second Language Learning: Experiments in the Processing of Natural and Partly Artificial Input. In H. W. Dechert & M. Raupach (Eds.), *Interlingual Processing* (pp. 49–73). Tübingen: Gunter Narr Verlag.

Hulstijn, J. (2002). Towards a Unified Account of the Representation, Processing and Acquisition of Second Language Knowledge. *Second Language Research*, *18*(3), 193-223.

Hulstijn, J. (2003). Incidental and Intentional Learning. In C. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 349–81). London, England: Blackwell.

Hulstijn, J., Van Gelderen, A., & Schoonen, R. (2009). Automatization in Second Language Acquisition: What does the Coefficient of Variation Tell Us?. *Applied Psycholinguistics*, *30*(4), 555-582.

Iwashita, N. (2003). Negative Feedback and Positive Evidence in Task-Based Interaction: Differential Effects on L2 Development. *Studies in Second Language Acquisition*, *25*(1), 1-36.

Izumi, S. (2002). Output, Input Enhancement, and the Noticing Hypothesis: An Experimental Study on ESL Relativization. *Studies in Second Language Acquisition*, *24*(4), 541-577.

Izumi, S. (2003). Comprehension and Production Processes in Second Language Learning: In Search of the Psycholinguistic Rationale of the Output Hypothesis. *Applied Linguistics*, *24*(2), 168-196.

Izumi, S., & Bigelow, M. (2000). Does Output Promote Noticing and Second Language Acquisition?. *TESOL Quarterly*, *34*(2), 239-278.

Jackson, D. O., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A Synthesis and Meta-Analysis of Research on Second Language Task Complexity. *Language Learning*, *63*(2), 330-367.

Janssen, N., & Barber, H. A. (2012). Phrase Frequency Effects in Language Production. *Plos One*, *7*(3), e33202.

Jiang, N. (2007). Selective Integration of Linguistic Knowledge in Adult Second Language Learning. *Language Learning*, *57*(1), 1-33.

Jiang, N., & Nekrasova, T. (2007). The Processing of Formulaic Sequences by Second Language Speakers. *The Modern Language Journal*, *91*(3), 433-445.

Jurafsky, D. (1996). A Probabilistic Model of Lexical and Syntactic Access and Disambiguation. *Cognitive Science*, *20*(2), 137–194.

Jurafsky, D. (2002). Probabilistic Modeling in Psycholinguistics: Linguistic Comprehension and Production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 39–96). Cambridge, MA: MIT Press.

Kawaguchi, S., & Di Biase, B. (2012). Acquiring procedural skills in L2: Processability Theory and skill acquisition. *Studies in Language Sciences*, *11*, 68-95.

Kempe, V., & MacWhinney, B. (1998). The Acquisition of Case Marking by Adult Learners of Russian and German. *Studies in Second Language Acquisition*, *20*(4), 543-587.

Kennedy, C. (2001). Polar Opposition and the Ontology of 'Degrees'. *Linguistics and Philosophy*, *24*(1), 33-70.

Kennedy, C. (2006). Semantics of Comparatives. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics* (2nd Ed.), (pp. 690–694). Oxford: Elsevier.

Kennedy, C. (2007). Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives. *Linguistics and Philosophy*, *30*, 1–45.

Kennedy, C., & McNally, L. (2005). Scale Structure, Degree Modification, and The Semantics of Gradable Predicates. *Language*, 345-381.

Kim, Y. (2006). Effects of Input Elaboration on Vocabulary Acquisition through Reading by Korean Learners of English as a Foreign Language. *TESOL Quarterly*, *40*(2), 341-373.

Kim, Y. (2012). Task Complexity, Learning Opportunities and Korean EFL Learners' Question Development. *Studies in Second Language Acquisition*, *34*(4), 627–658.

Kim, Y., Payant, C., & Pearson, P. (2015). The Intersection of Task-Based Interaction, Task Complexity, and Working Memory: L2 Question Development through Recasts in a Laboratory Setting. *Studies in Second Language Acquisition*, *37*(3), 549-581.

Kormos, J. (1999). Monitoring and Self-Repair in L2. *Language Learning*, *49*(2), 303-342.

Kormos, J. (2000). The Role of Attention in Monitoring Second Language Speech Production. *Language Learning*, *50*(2), 343-384.

Kormos, J. (2006). *Speech Production and Second Language Acquisition*. Mahwah: Erlbaum.

Kormos, J. (2011). Speech Production and the Cognition Hypothesis. In P. Robinson (Ed.), *L2 Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance* (pp. 39–60). Amsterdam: John Benjamins.

Kormos, J., & Préfontaine, Y. (2017). Affective Factors Influencing Fluent Performance: French Learners' Appraisals of Second Language Speech Tasks. *Language Teaching Research*, *21*(6), 699-716.

Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

Krashen, S. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.

Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. London; New York: Longman.

Krashen, S. (1998). Comprehensible Output? *System*, *26*(2), 175–82.

Kraska-Miller, M. (2014). *Nonparametric Statistics for Social and Behavioural Sciences*. Alabama: Taylor & Francis.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The Development of Lexical Fluency in a Second Language. *Second Language Research*, *18*(2), 137-171.

Kuiken, F., & Vedder, I. (2002). The Effect of Interaction in Acquiring the Grammar of a Second Language. *International Journal of Educational Research*, *37*(3-4), 343-358.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University Of Chicago Press.

Lakoff, G. (2004). *Don't Think of an Elephant! Know Your Values and Frame the Debate*. White River Junction, VT: Chelsea Green.

Lakoff, G. (2012). Explaining Embodied Cognition Results. *Topics in Cognitive Science*, *4*(4), 773-785.

Lambert, C. (2010). A Task-Based Needs Analysis: Putting Principles into Practice. *Language Teaching Research*, *14*(1), 99-112.

Lambert, C. (2018). Tasks vs. Exercises. In J. Liontas (Ed.), *the TESOL Encyclopedia of English Language Teaching* (pp. 1-6). Hoboken, NJ: Blackwell Publishing.

Lambert, C. (2019). *Referent Similarity and Nominal Syntax in Task-Based Language Teaching*. Singapore: Springer.

Lambert, C., & Engler, S. (2007). Information Distribution and Goal Orientation in Second Language Task Design. In M.P. García Mayo (Ed.). *Investigating Tasks in Formal Language Learning* (pp. 27–43). Clevedon: Multilingual matters.

Lambert, C., & Kormos, J. (2014). Complexity, Accuracy, and Fluency in Task-based L2 Research: Toward More Developmentally Based Measures of Second Language Acquisition. *Applied Linguistics*, *35*(5), 607-614.

Lambert, C., & Nakamura, S. (2019). Proficiency-Related Variation in Syntactic Complexity: A Study of English L1 and L2 Oral Descriptive Discourse. *International Journal of Applied Linguistics*, *29*(2), 248-264.

Lambert, C. & Zhang, G. (2019). Engagement in the Use of English and Chinese as Foreign Languages: The Role of Learner-Generated Content in Instructional Task Design. *Modern Language Journal*, *103*(2), 391-411.

Lambert, C., Aubrey, S., & Leeming, P. (2020). Task Preparation and Second Language Speech Production. *TESOL Quarterly*, *Early View*, 1-35.

Lambert, C., Kormos, J., & Minn, D. (2017). Task Repetition and Second Language Speech Processing. *Studies in Second Language Acquisition*, *39*(1), 167-196.

Langacker, R. (1987a). *Foundations of Cognitive Grammar*, Vol. 1: *Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. (1987b). Nouns and Verbs. *Language*, *63*(1), 53–94.

Langacker, R. (1990). *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.

Langacker, R. (2008). Cognitive Grammar as a Basis for Language Instruction. In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 66–88). New York, NY: Routledge.

Lantolf, J. (Ed.). (2000). *Sociocultural Theory and Second Language Learning*. Oxford: Oxford university press.

Lantolf, J. (2006). Language Emergence: Implications for Applied Linguistics—A Sociocultural Perspective. *Applied linguistics*, *27*(4), 717-728.

Lantolf, J., & Ahmed, M. (1989). Psycholinguistic Perspectives on Interlanguage Variation: A Vygotskian Analysis. In S. Gass, L. Selinker, & D. Preston (Eds.), *Variation in Second Language Acquisition: Psycholinguistic Issues* (pp. 93 – 108). Clevedon: Multilingual Matters.

Lapkin, S., Hart, D., & Swain, M. (1991). Early and Middle French Immersion Programs: French Language Outcomes. *Canadian Modern Language Review*, *48*(1), 11-44.

Larsen-Freeman, D. (2003). *Teaching Language: From Grammar to Grammaring*. Boston: Heinle & Heinle.

Larsen-Freeman, D. (2002). Language Acquisition and Language Use form a Chaos/Complexity Theory Perspective. In C. Kramsch (Ed.), *Language Acquisition and Language Socialization: Ecological Perspectives* (pp. 33–46). London: Continuum.

Larsen-Freeman, D. (2013a). Complexity Theory and Dynamic Systems Theory. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition* (pp. 103-106). London: Routledge.

Larsen-Freeman, D. (2013b). Transfer of Learning Transformed. *Language Learning*, *63*, 107–129.

Larsen-Freeman, D. (2015). Saying What We Mean: Making a Case for 'Language Acquisition' to Become 'Language Development'. *Language Teaching*, *48*(4), 491-505.

Laufer, B. (1990). 'Sequence' and 'Order' in the Development of L2 Lexis: Some Evidence from Lexical Confusions. *Applied Linguistics*, *11*(3), 281-296.

Leeman, J. (2003). Recasts and Second Language Development: Beyond Negative Evidence. *Studies in Second Language Acquisition*, *25*(1), 37-63.

Lennon, P. (1990). Investigating Fluency in EFL: A Quantitative Approach. *Language Learning*, *40*(3), 387-417.

Leonard, K., & Shea, C. (2017). L2 Speaking Development during Study Abroad: Fluency, Accuracy, Complexity, and Underlying Cognitive Factors. *The Modern Language Journal*, *101*(1), 179-193.

Leow, R. (2007). Input in the L2 Classroom: An Attentional Perspective on Receptive Practice. In R. Dekeyser (Eds.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 21–50). Cambridge: Cambridge University Press.

Leow, R., & Zamora, C. (2017). Intentional and Incidental L2 Learning. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 33-49). New York: Routledge.

Lenzing, A. (2015). Exploring Regularities and Dynamic Systems in L2 Development. *Language Learning*, *65*(1), 89-122.

Lewis-Beck, M., Bryman, A., & Liao, T. (2004). *The Sage Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: Sage.

Levelt. W. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Levelt, W. (1999). Producing Spoken Language: A Blueprint of the Speaker. In C. Brown & P. Hagoort (Eds.), *The Neurocognition of Language* (pp. 83–122). Oxford: Oxford University Press.

Levelt, W. (2001). Spoken Word Production: A Theory of Lexical Access. *Proceedings of the National Academy of Sciences, 98*(13), 464 –13,471.

Li, S. (2010). The Effectiveness of Corrective Feedback in SLA: A Meta-Analysis. *Language Learning*, *60*(2), 309-365.

Li, S. (2018). Corrective Feedback in L2 Speech Production. In J. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching* (pp. 1–9). Hoboken, NJ: Blackwell Publishing.

Lieven, E., & Tomasello, M. (2008). Children's First Language Acquisition from a Usage-Based Perspective. In P. Robinson & N. Ellis (Eds), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 168–196). New York: Routledge.

Lightbown, P. (2008). Transfer Appropriate Processing as a Model for Classroom Second Language Acquisition. In Z. Han (Ed.), *Understanding Second Language Process* (pp. 27 – 44). Clevedon, UK: Multilingual Matters.

Lim, H., & Godfroid, A. (2015). Automatization in Second Language Sentence Processing: A Partial, Conceptual Replication of Hulstijn, Van Gelderen, and Schoonen's 2009 Study. *Applied Psycholinguistics*, *36*(5), 1247-1282.

Loewen, S., & Philp, J. (2006). Recasts in the Adult English L2 Classroom: Characteristics, Explicitness, and Effectiveness. *The Modern Language Journal*, *90*(4), 536-556.

Logan, G. (1985). Skill and Automaticity: Relations, Implications, and Future Directions. *Canadian Journal of Psychology/Revue Canadienne De Psychologie*, *39*(2), 367-386.

Logan, G. (1988). Toward an Instance Theory of Automatization. *Psychological Review*, *95*(4), 492-572.

Logan, G. (2002). An Instance Theory of Attention and Memory. *Psychological Review*, *109*(2), 376.

Long, M. (1981a). Input, Interaction, and Second-Language Acquisition. *Annals of the New York Academy of Sciences*, *379*(1), 259-278.

Long, M. (1981b). Questions in Foreigner Talk Discourse. *Language Learning*, *31*(1), 135-157.

Long, M. (1983a). Does Second Language Instruction Make a Difference? A Review of Research. *TESOL quarterly*, *17*(3), 359-382.

Long, M. (1983b). Native Speaker/Non-Native Speaker Conversation and the Negotiation of Comprehensible Input. *Applied linguistics*, *4*(2), 126-141.

Long, M. (1983c). Linguistic and Conversational Adjustments to Non-Native Speakers. *Studies in Second Language Acquisition*, *5*(2), 177-193.

Long, M. (1985). A Role for Instruction in Second Language Acquisition: Task-Based Language Teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition* (pp. 77-99). Clevedon: Multilingual Matter.

Long, M. (1989). Task, Group, and Task-Group Interactions. In S. Anivan (Ed.), *Language Teaching Methodology for the Nineties* (pp. 31-50, 1990). Singapore: SEAMEO Regional Language Center.

Long, M. (1990). Maturational Constraints on Language Development. *Studies in Second Language Acquisition*, *12*(3), 251-285.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, D. Coste, 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 Eenvironment in Second Language Acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 413-468). San Diego, CA: Academic Press

Long, M. (1998). Focus on Form in Task-Based Language Teaching. *University of Hawai'i Working Papers in ESL, 16*(2), 35-49.

Long, M. (2007). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum.

Long, M. (2009). Methodological Principles in Language Teaching. In M. Long & C. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 373–94). Oxford, UK: Blackwell.

Long, M. (2015). *Second Language Acquisition and Task-Based Language Teaching*. Malden, USA: Wiley-Blackwell.

Long, M. (2016). In Defense of Tasks and TBLT: Nonissues and Real Issues. *Annual Review of Applied Linguistics*, *36*, 5-33.

Long, M. (2017). Instructed Second Language Acquisition (ISLA): Geopolitics, Methodological Issues, and Some Major Research Questions. *Instructed Second Language Acquisition*, *1*(1), 7-44.

Long, M. (2019). Optimal Input for Language Learning: Genuine, Simplified, Elaborated, or Modified Elaborated?. *Language Teaching*, 1-14.

Long, M., & Crookes, G. (1992). Three Approaches to Task-Based Syllabus Design. *TESOL Quarterly*, *26*(1), 27-56.

Long, M., & Robinson, P. (1998). Focus-on-Form: Theory, Research and Practice. In C. Doughty & J. Williams (Eds.), *Focus-on-Form in Classroom Second Language Acquisition* (pp. 15–41). Cambridge: Cambridge University Press.

Long, M., Inagaki, S., & Ortega, L. (1998). The Role of Implicit Negative Feedback in SLA: Models and Recasts in Japanese and Spanish. *The Modern Language Journal*, *82*(3), 357-371.

Loschky, L. (1994). Comprehensible Input and Second Language Acquisition: What Is the Relationship?. *Studies in Second Language Acquisition*, *16*(3), 303-323.

Loschky, L., & Bley-Vroman, R. (1993). Creating Structure-Based Communication Tasks for Second Language Development. In G. Crookes & S. Gass (Eds.), *Tasks and Language Learning: Integrating Theory and Practice* (pp. 123-167). Clevedon: Multilingual Matters.

Lyddon, P. (2011). The Efficacy of Corrective Feedback and Textual Enhancement in Promoting the Acquisition of Grammatical Redundancies. *The Modern Language Journal*, *95*, 104-129.

Lyster, R. (1998a). Negotiation of Form, Recasts, and Explicit Correction in Relation to Error Types and Learner Repair in Immersion Classrooms. *Language Learning*, *48*(2), 183-218.

Lyster, R. (1998b). Recasts, Repetition, and Ambiguity in L2 Classroom Discourse. *Studies in Second Language Acquisition*, *20*(1), 51-81.

Lyster, R., & Mori, H. (2006). Interactional Feedback and Instructional Counterbalance. *Studies in Second Language Acquisition*, *28*(2), 269-300.

Lyster, R., & Ranta, L. (1997). Corrective Feedback and Learner Uptake: Negotiation of Form in Communicative Classrooms. *Studies in Second Language Acquisition*, *19*(1), 37-66.

Lyster, R., & Ranta, L. (2013). Counterpoint Piece: The Case for Variety in Corrective Feedback Research. *Studies in Second Language Acquisition*, *35*(1), 167-184.

Mackey, A. (1999). Input, Interaction, and Second Language Development: An Empirical Study of Question Formation in ESL. *Studies in Second Language Acquisition*, *21*(4), 557-587.

Mackey, A. (2006). Feedback, Noticing and Instructed Second Language Learning. *Applied Linguistics*, *27*(3), 405-430.

Mackey, A. (2007) Interaction as Practice. In R. Dekeyser (Ed.) *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 85–110). Cambridge: Cambridge University Press.

Mackey, A., & Abbuhl, R. (2005). Input and Interaction. In C. Sanz (Ed.), *Mind and Context in Adult Second Language Acquisition* (pp. 207–233). Washington, DC: Georgetown University Press.

Mackey, A., & Gass, S. (2015). *Second Language Research: Methodology and Design*. Routledge.

Mackey, A., & Oliver, R. (2002). Interactional Feedback and Children's L2 Development. *System*, *30*(4), 459-477.

Mackey, A., & Philp, J. (1998). Conversational Interaction and Second Language Development: Recasts, Responses, and Red Herrings?. *The Modern Language Journal*, *82*(3), 338-356.

Mackey, A., & Polio, C. (Eds.). (2009). *Multiple Perspectives on Interaction: Second Language Research in Honor of Susan M. Gass*. Mahwah, NJ: Lawrence Erlbaum.

Mackey, A., Abbuhl, R., & Gass, S. (2012). Interactionist Approach. In S. Gass, & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition*. Abingdon: Routledge.

Mackey, A., Gass, S., & McDonough, K. (2000). How do Learners Perceive Interactional Feedback?. *Studies in Second Language Acquisition*, *22*(4), 471-497.

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional Input and the Incorporation of Feedback: An Exploration of NS–NNS and NNS–NNS Adult and Child Dyads. *Language Learning*, *53*(1), 35-66.

Mackey, A., Philp, J., Egi, T., Fujii, A., & Tatsumi, T. (2002). Individual Differences in Working Memory, Noticing of Interactional Feedback, and L2 Development. In P. Robinson (Ed.), *Individual Differences in Instructed Language Learning* (pp. 181 – 209). Amsterdam: Benjamins

MacWhinney, B. (1987). Applying the Competition Model to Bilingualism. *Applied Psycholinguistics*, *8*(4), 315-327.

MacWhinney, B. (1997). Second Language Acquisition and the Competition Model. In A. de Groot & J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives* (pp. 113–144). Mahwah, NJ: Erlbaum.

MacWhinney, B. (Ed.). (1999). *The Emergence of Language*. Hillsdale, NJ: Erlbaum.

MacWhinney, B. (2001a). Emergentist Approaches to Language. In J. Bybee & P. Hopper (Eds)., *Frequency and the Emergence of Linguistic Structure* (pp. 449–470). Amsterdam: John Benjamins.

MacWhinney, B. (2001b). The Competition Model: The Input, the Context, and the Brain. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 69–90). New York: Cambridge University Press.

MacWhinney, B. (2006). Emergentism—Use Often and With Care. *Applied Linguistics*, *27*(4), 729-740.

MacWhinney, B. (2015). Introduction: Language Emergence. In B. MacWhinney & W. O'Grady (Eds.), *Handbook of Language Emergence* (pp. 1–32). New York: Wiley.

MacWhinney, B. & Bates, E. (1989). *The Crosslinguistic Study of Sentence Processing*. New York: Cambridge University Press.

MacWhinney, B., & O'Grady, W. (Eds.). (2015). *The Handbook of Language Emergence*. Oxford: Wiley Blackwell.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue Validity and Sentence Interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127-150.

Marsden, E. (2006). Exploring Input Processing in the Classroom: An Experimental Comparison of Processing Instruction and Enriched Input. *Language Learning*, *56*(3), 507-566.

Marsden, E., & Chen, H. Y. (2011). The Roles of Structured Input Activities in Processing Instruction and the Kinds of Knowledge They Promote. *Language Learning*, *61*(4), 1058-1098.

Martin, K., & Ellis, N. (2012). The Roles of Phonological Short-Term Memory and Working Memory in L2 Grammar and Vocabulary Learning. *Studies in Second Language Acquisition*, *34*(3), 379-413.

McClelland, J., Fiez, J., & McCandliss, B. (2002). Teaching the/r/–/l/Discrimination to Japanese Adults: Behavioral and Neural Aspects. *Physiology & Behavior*, *77*(4-5), 657-662.

McDonough, K. (2005). Identifying the Impact of Negative Feedback and Learners' Responses on ESL Question Development. *Studies in Second Language Acquisition*, *27*(1), 79-103.

McDonough, K., & Mackey, A. (2006). Responses to Recasts: Repetitions, Primed Production, snd Linguistic Development. *Language Learning*, *56*(4), 693-720.

McDonough, K., & Mackey, A. (2008). Syntactic Priming and ESL Question Development. *Studies in Second Language Acquisition*, *30*(1), 31-47.

McDonough, K., & Kim, Y. (2009). Syntactic Priming, Type Frequency, and EFL Learners' Production of Wh-Questions. *The Modern Language Journal*, *93*(3), 386-398.

McDonough, K., & Nekrasova-Becker, T. (2014). Comparing the Effect of Skewed and Balanced Input on English as a Foreign Language Learners' Comprehension of the Double-Object Dative Construction. *Applied Psycholinguistics*, *35*(2), 419-442.

McDonough, K., & Trofimovich, P. (2015). Structural Priming and the Acquisition of Novel Form-Meaning Mappings. In T. Cadierno & S. Eskildsen (Eds.), *Usage-Based Perspectives on Second Language Learning* (pp. 105–123). Berlin: de Gruyter Mouton.

McEnery, A., Xiao, R., & Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. London: Routledge.

McLaughlin, B. (1987). *Theories of Second Language Learning*. London: Arnold.

Michel, M. (2011). Effects of Task Complexity and Interaction on L2 Performance. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance* (pp. 141–174). Amsterdam: John Benjamins.

Michel, M. (2017). Complexity, Accuracy, and Fluency in L2 Production. In M. Sato & S. Loewen (Eds.), Hanbook of Instructed Second Language Acquisition (pp. 50–68). London/New York: Routledge.

Moors, A., & De Houwer, J. (2006). Automaticity: A Theoretical and Conceptual Analysis. *Psychological Bulletin*, *132*(2), 297–326.

Morgan-Short, K. (2014). Electrophysiological Approaches to Understanding Second Language Acquisition: A Field Reaching Its Potential. *Annual Review of Applied Linguistics*, *34*, 15-36.

Morgan-Short, K. & Bowden, H. (2006). Processing Instruction and Meaningful Output-Based Instruction: Effects on Second Language Development. *Studies in Second Language Acquisition*, *28*(1), 31-65.

Muranoi, H. (2007). Output Practice in the L2 Classroom. In R. Dekeyser (Ed.), In R. Dekeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (pp. 51–84). Cambridge: Cambridge University Press.

Nassaji, H. (2007b). Focus on Form through Recasts in Dyadic Student–Teacher Interaction: A Case for Recast Enhancement. In C. Gascoigne (Ed.), *Assessing the Impact of Input Enhancement in Second Language Education*. (pp. 53–69). Stillwater, OK: New Forums Press.

Nicholas, H., Lightbown, P., & Spada, N. (2001). Recasts as Feedback to Language Learners. *Language Learning*, *51*(4), 719-758.

Nobuyoshi, J., & Ellis, R. (1993). Focused Communication Tasks and Second Language Acquisition. *ELT Journal*, *47*(3), 203-210.

Norris, J., (2009). Task-Based Teaching and Testing. In M.H. Long & C. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 578–594). Oxford, UK: Blackwell.

Norris, J., & Ortega, L. (2000). Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis. *Language Learning*, *50*(3), 417-528.

Norris, J., & Ortega, L. (2001). Does Type of Instruction Make a Difference? Substantive Findings from a Meta-Analytic Review. *Language Learning*, *51*(s1), 157-213.

Norris, J. & Ortega, L. (2003). Defining and Measuring SLA. In C. Doughty and M. Long (Eds.), *The Handbook of Second Language Acquisition*, (pp. 717–61). Oxford: Blackwell.

Norris, J., & Ortega, L. (2013). Assessing Learner Knowledge. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 573–589). New York: Routledge.

O'Grady, W. (2015). Processing Determinism. *Language Learning*, *65*(1), 6-32.

Oh, S. (2001). Two Types of Input Modification and EFL Reading Comprehension: Simplification versus Elaboration. *TESOL Quarterly*, *35*(1), 69-96.

Oliver, R. (1995). Negative Feedback in Child NS-NNS Conversation. *Studies in Second Language Acquisition*, *17*(4), 459-481.

Oliver, R. (1998). Negotiation of Meaning in Child Interactions. *The Modern Language Journal*, *82*(3), 372-386.

Oliver, R. (2000). Age Differences in Negotiation and Feedback in Classroom and Pairwork. *Language Learning*, *50*(1), 119-151.

Oliver, R. (2002). The Patterns of Negotiation for Meaning in Child Interactions. *The Modern Language Journal*, *86*(1), 97-111.

Oliver, R., & Mackey, A. (2003). Interactional Context and Feedback in Child ESL Classrooms. *The Modern Language Journal*, *87*(4), 519-533.

Ortega, L. (2007). Meaningful L2 Practice in Foreign Language Classrooms: A Cognitive-Interactionist SLA Perspective. In R. Dekeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*, (pp. 180-207). Cambridge: Cambridge University Press.

Ortega, L. (2009). Sequences and Processes in Language Learning. In M. Long & C. Doughty (Eds.), *The Handbook of Language Teaching* (pp. 81-105). Oxford, UK: Blackwell.

Ortega, L. (2015). Second Language Learning Explained? SLA across 10 Contemporary Theories. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (2nd Ed.), (pp. 245–272). New York, NY: Routledge.

Pallotti, G. (2009). CAF: Defining, Refining and Differentiating Constructs. *Applied Linguistics*, *30*(4), 590-601.

Palmeri, T. (1997). Exemplar Similarity and the Development of Automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 324–354.

Panova, I., & Lyster, R. (2002). Patterns of Corrective Feedback and Uptake in an Adult ESL Classroom. *TESOL Quarterly*, *36*(4), 573-595.

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and Procedural Determinants of Second Languages*. Amsterdam: John Benjamins.

Park, E. (2011). Learner-Generated Noticing of Written L2 Input: What do Learners Notice and Why?. *Language Learning*, *61*(1), 146-186.

Park, E. & Han, Z-H. (2008) Learner Spontaneous Attention in L2 Input Processing: An Exploratory Study. In Z-H. Han, E. Park, A. Révész, C. Combs, & J. Kim (Eds.). *Understanding Second Language Process* (pp. 106-132). Clevedon: Multilingual Matters.

Parker, K., & Chaudron, C. (1987). The Effects of Linguistic Simplification and Elaborative Modifications in L2 Comprehension. *University of Hawai'i Working Papers in ESL, 6*(2), 107-133

Paulson, C., & Bruder, M. (1976). *Teaching English as a Second Language: Techniques and Procedures*. Cambridge, MA: Winthrop Publishers.

Pawley, A., & Syder, F. (1983). Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191–225). London: Longman.

Pham, T., Kang, J., Johnson, A., & Archibald, L. (2020). Feature-Focusing Constraints on Implicit Learning of Function Word and Meaning Associations. *Applied Psycholinguistics*, *41*(2), 401-426.

Peters, P. (1996). Comparative Insights into Comparison. *World Englishes*, *15*(1), 57-67.

Phillips, N., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic Priming in a First and Second Language: Evidence from Reaction Time Variability and Event-Related Brain Potentials. *Journal of Neurolinguistics*, *17*(2-3), 237-262.

Philp, J. (2003). Constraints on "Noticing the Gap": Nonnative Speakers' Noticing of Recasts in NS-NNS Interaction. *Studies in Second Language Acquisition*, *25*(1), 99-126.

Pica, T. (1988). Interlanguage Adjustments as an Outcome of NS-NNS Negotiated Interaction. *Language Learning*, *38*(1), 45-73.

Pica, T. (1994). Research on Negotiation: What does It Reveal about Second-Language Learning Conditions, Processes, and Outcomes?. *Language Learning*, *44*(3), 493-527.

Pica, T. (1996). Second Language Learning through Interaction: Multiple Perspectives. *Working Papers in Educational Linguistics*, *12*(1), 1-22.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and Using Communication Tasks for Second Language Research and Instruction. In S. Gass &. G. Crookes (Eds.), *Task-Based Learning in a Second Language* (pp. 9-34). Clevedon, Avon: Multilingual Matters.

Pica, T., Kang, H. S., & Sauro, S. (2006). Information Gap Tasks: Their Multiple Roles and Contributions to Interaction Research Methodology. *Studies in Second Language Acquisition*, *28*(2), 301-338.

Pica, T., Young, R., & Doughty, C. (1987). The Impact of Interaction on Comprehension. *TESOL Quarterly*, *21*(4), 737-758.

Pickering, M., & Ferreira, V. (2008). Structural Priming: A Critical Review. *Psychological Bulletin*, *134*(3), 427.

Pickering, M., Branigan, H., Cleland, A., & Stewart, A. (2000). Activation of Syntactic Information during Language Production. *Journal of Psycholinguistic Research*, *29*(2), 205-216.

Pienemann, M. (1984). Psychological Constraints on the Teachability of Languages. *Studies in Second Language Acquisition*, *6*(2), 186-214.

Pienemann, M. (1985). Learnability and Syllabus Construction. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition* (pp. 23-75). Clevedon: Multilingual Matter.

Pienemann, M. (1987). Determining the Influence of Instruction on L2 Speech Processing. *Australian Review of Applied Linguistics*, *10*(2), 83-113.

Pienemann, M. (1989). Is Language Teachable? Psycholinguistic Experiments and Hypotheses. *Applied Linguistics*, *10*(1), 52-79.

Pienemann, M. (2015). An Outline of Processability Theory and Its Relationship to Other Approaches to SLA. *Language Learning*, *65*(1), 123-151.

Pienemann, M., & Johnston, M. (1987). Factors Influencing the Development of Language Proficiency. In D. Nunan (Ed.), *Applying Second Language Acquisition Research* (pp. 45–141). Adelaide, Australia: National Curriculum Resource Centre, AMEP.

Pienemann, M., & Keßler, J. (2012). Processability Theory. In S. Gass & A. Mackey (Eds.), *the Routledge Handbook of Second Language Acquisition* (pp. 228-247). New York: Routledge.

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an Acquisition-Based Procedure for Second Language Assessment. *Studies in Second Language Acquisition*, *10*(2), 217-243.

Plonsky, L. (2017). Quantitative Research Methods. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 505–521). New York: Routledge.

Plonsky, L., & Kim, Y. (2016). Task-Based Learner Production: A Substantive and Methodological Review. *Annual Review of Applied Linguistics*, *36*, 73-97.

Plonsky, L., & Oswald, F. (2014). How Big Is “Big”? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878-912.

Polio, C., & Gass, S. (1998). The Role of Interaction in Native Speaker Comprehension of Nonnative Speaker Speech. *The Modern Language Journal*, *82*(3), 308-319.

Prabhu, N. (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

Pütz, M. & Sicola, L. (Eds.) (2010). *Cognitive Processing in Second Language Acquisition*. Amsterdam & Philadelphia: John Benjamins.

Rasuki, M. (2017). Processing Instruction: A Review of Issues. *International Journal of Education and Literacy Studies*, *5*(3), 1-7.

Reali, F., & Christiansen, M. H. (2007). Processing of Relative Clauses is Made Easier by Frequency of Occurrence. *Journal of Memory and Language*, *57*(1), 1-23.

Rebuschat, P. (2013). Measuring Implicit and Explicit Knowledge in Second Language Research. *Language Learning*, *63*(3), 595-626.

Rebuschat, P. (Ed.). (2015). *Implicit and Explicit Learning of Languages*. Amsterdam: John Benjamins.

Redington, M., & Chater, N. (1998). Connectionist and Statistical Approaches to Language Acquisition: A Distributional Perspective. *Language And Cognitive Processes*, *13*(2-3), 129-191.

Read, J. (2004). Plumbing the Depths: How Should the Construct of Vocabulary Knowledge be Defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a Second Language* (pp. 209-227). Amsterdam: John Benjamins Publishing.

Révész, A. (2009). Task Complexity, Focus on Form, and Second Language Development. *Studies in Second Language Acquisition*, *31*(3), 437-470.

Révész, A. (2011). Task Complexity, Focus on L2 Constructions, and Individual Differences: A Classroom-Based Study. *The Modern Language Journal*, *95*, 162-181.

Révész, A. (2012). Working Memory and the Observed Effectiveness of Recasts on Different L2 Outcome Measures. *Language Learning*, *62*(1), 93-132.

Révész, A., & Han, Z. (2006). Task Content Familiarity, Task Type and Efficacy of Recasts. *Language Awareness*, *15*(3), 160-179.

Révész, A., Sachs, R., & Hama, M. (2014). The Effects of Task Complexity and Input Frequency on the Acquisition of the Past Counterfactual Construction through Recasts. *Language Learning*, *64*(3), 615-650.

Richardson, M. (1993). *Negative Evidence and Grammatical Morpheme Acquisition: Implications For SLA*. Unpublished Manuscript, University of Western Australia.

Rickard, T. (1997). Bending the Power Law: A CMPL Theory of Strategy Shifts and the Automatization of Cognitive Skills. *Journal of Experimental Psychology: General*, *126*(3), 288-311.

Robinson, P. (1995). Attention, Memory, and the "Noticing" Hypothesis. *Language Learning*, *45*(2), 283-331.

Robinson, P. (1996). Learning Simple and Complex Second Language Rules under Implicit, Incidental, Rule-Search, and Instructed Conditions. *Studies in Second Language Acquisition*, *18*(1), 27-67.

Robinson, P. (1997). Generalizability and Automaticity of Second Language Learning under Implicit, Incidental, Enhanced, and Instructed Conditions. *Studies in Second Language Acquisition*, *19*(2), 223-247.

Robinson, P. (2001). Task Complexity, Cognitive Resources, and Syllabus Design: A Triadic Framework for Examining Task Influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction*, (pp. 285-318). Cambridge: Cambridge University Press.

Robinson, P. (2002). Effects of Individual Differences in Intelligence, Aptitude, and Working Memory on Adult Incidental SLA: A Replication and Extension of Reber, Walkenfield, and Hernstadt (1991). In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 211–265). Amsterdam/Philadelphia: Benjamins.

Robinson, P. (2003a). Attention and Memory During SLA. In M. Long & C. Doughty (Eds.), *the Handbook of Language Teaching* (pp. 631-678). Oxford, UK: Blackwell.

Robinson, P. (2003b). The Cognitive Hypothesis, Task Design, and Adult Task-Based Language Learning. *Second Language Studies*, *21*(2), 45-105.

Robinson, P. (2005a). Cognitive Abilities, Chunk-Strength, and Frequency Effects in Implicit Artificial Grammar and Incidental L2 Learning: Replications of Reber, Walkenfeld, and Hernstadt (1991) and Knowlton and Squire (1996) and Their Relevance for SLA. *Studies in Second Language Acquisition*, *27*(2), 235-268.

Robinson, P. (2005b). Cognitive Complexity and Task Sequencing: Studies in a Componential Framework for Second Language Task Design. *International Review of Applied Linguistics in Language Teaching*, *43*(1), 1-32.

Robinson, P. (2007). Task Complexity, Theory of Mind, and Intentional Reasoning: Effects on L2 Speech Production, Interaction, Uptake and Perceptions of Task Difficulty. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 193-213.

Robinson, P. (2010). Situating and Distributing Cognition across Task Demands: The SSARC Model of Pedagogic Task Sequencing. In M. Pütz & L. Sicola (Eds.), *Cognitive Processing in Second Language Acquisition: Inside the Learner's Mind* (pp. 239-264). Amsterdam: John Benjamins.

Robinson, P. (2011a). Second Language Task Complexity, the Cognition Hypothesis, Language Learning, and Performance. In Robinson, P. (Ed.). (2011). *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance* (pp. 3-38). Amsterdam: John Benjamins.

Robinson, P. (2011b). Task-Based Language Learning: A Review of Issues. *Language Learning*, *61*(s1), 1-36.

Robinson, P. (2015). The Cognition Hypothesis, Second Language Task Demands, and the SSARC Model of Pedagogic Task Sequencing. In M. Bygate (Ed.), *Domains and Directions in the Development of TBLT* (pp. 87-121). Amsterdam: John Benjamins.

Robinson, P., & Ellis, N. C. (Eds.). (2008). *A Handbook of Cognitive Linguistics and SLA*. New York: Routledge.

Robinson, P., & Gilabert, R. (2007). Task Complexity, the Cognition Hypothesis and Second Language Learning and Performance. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(3), 161-176.

Robinson, P., & Ha, M. (1993). Instance Theory and Second Language Rule Learning under Explicit Conditions. *Studies in Second Language Acquisition*, *15*(4), 413-438.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and Motion: Measuring the Effects of the Conceptual Demands of Tasks on Second Language Speech Production. *Applied Linguistics*, *30*(4), 533-554.

Rodgers, D. (2006). Developing Content and Form: Encouraging Evidence from Italian Content-Based Instruction. *The Modern Language Journal*, *90*(3), 373-386.

Rodgers, D. (2011). The Automatization of Verbal Morphology in Instructed Second Language Acquisition. *International Review of Applied Linguistics in Language Teaching*, *49*(4), 295-319.

Rogers, J. (2017). Awareness and Learning under Incidental Learning Conditions. *Language Awareness*, *26*(2), 113-133.

Rogers, J., & Révész, A. (in press). Experimental and Quasiexperimental Designs. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics*. New York: Routledge.

Rothman, J., & Slabakova, R. (2018). The Generative Approach to SLA and Its Place in Modern Second Language Studies. *Studies in Second Language Acquisition*, *40*(2), 417-442.

Rott, S. (1999). The Effect of Exposure Frequency on Intermediate Language Learners' Incidental Vocabulary Acquisition and Retention through Reading. *Studies in Second Language Acquisition*, *21*(4), 589-619.

Ruiz, S., Tagarelli, K., & Rebuschat, P. (2018). Simultaneous Acquisition of Words and Syntax: Effects of Exposure Condition and Declarative Memory. *Frontiers in Psychology*, *9*, 1168.

Saffran, J., & Wilson, D. (2003). From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants. *Infancy*, *4*(2), 273-284.

Saffran, J., Newport, E., Aslin, R., Tunick, R., & Barrueco, S. (1997). Incidental Language Learning: Listening (and Learning) Out of the Corner of Your Ear. *Psychological Science*, *8*(2), 101-105.

Saito, K. (2015). Variables Affecting the Effects of Recasts on L2 Pronunciation Development. *Language Teaching Research*, *19*(3), 276-300.

Saito, K., & Lyster, R. (2012). Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of/ɹ/by Japanese Learners of English. *Language Learning*, *62*(2), 595-633.

Samuda, V., & Bygate, M. (2008). *Tasks in Second Language Learning*. Basingstoke: Palgrave Macmillan.

Sanz, C., & Grey, S. (2015). Effects of Conditions on L2 Development. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 301-324). Amsterdam: John Benjamins.

Sanz, C., & Morgan-Short, K. (2004). Positive Evidence versus Explicit Rule Presentation and Explicit Negative Feedback: A Computer-Assisted Study. *Language Learning*, *54*(1), 35-78.

Saragi, T., Nation, I., & Meister, G. (1978). Vocabulary Learning and Reading. *System*, *6*(2), 72–128.

Sapir, E. (1944). Grading: A Study in Semantics. *Philosophy of Science*, *11*(2), 93-116.

Sato, M., & McDonough, K. (2019). Practice is Important but How about Its Quality?: Contextualized Practice in the Classroom. *Studies in Second Language Acquisition*, *41*(5), 999-1026.

Saxton, M. (1997). The Contrast Theory of Negative Input. *Journal of Child Language*, *24*(1), 139-161.

Schmid, M., Verspoor, M. & Macwhinney, B. (2011). Coding and Extracting Data. In M. Verspoor, W. Lowie & K. De Bot (Eds.), *A Dynamic Approach to Second Language Development: Methods and Techniques*, (pp. 39-54). Amsterdam: John Benjamins.

Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*(2), 129-158.

Schmidt, R. (1992). Awareness and Second Language Acquisition. *Annual Review of Applied Linguistics*, *13*, 206-226.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmitt, N. (Ed.). (2004), *Formulaic sequences: Acquisition, Processing, and Use*. Philadelphia: John Benjamins.

Schwartz, B. (1999). Let's Make up Your Mind: “Special Nativist” Perspectives on Language, Modularity of Mind, and Nonnative Language Acquisition. *Studies in Second Language Acquisition*, *21*(4), 635-655.

Segalowitz, N. (2003). Automaticity and Second Language Acquisition. In C. Doughty & M. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 382-408) Oxford: Blackwell.

Segalowitz, N. (2010). *Cognitive Bases of Second Language Fluency*. New York: Routledge.

Segalowitz, N. (2016). Second Language Fluency and Its Underlying Cognitive and Social Determinants. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 79-95.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in Bilingualism and Second Language Learning. In J. Kroll & A. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 371–388). New York: Oxford University Press.

Segalowitz, N., & Lightbown, P. (1999). Psycholinguistic Approaches to SLA. *Annual Review of Applied Linguistics*, *19*, 43-63.

Segalowitz, N., & Freed, B. (2004). Context, Contact, and Cognition in Oral Fluency Acquisition: Learning Spanish in at Home and Study Abroad Contexts. *Studies in Second Language Acquisition*, *26*(2), 173-199.

Segalowitz, N., & Segalowitz, S. (1993). Skilled Performance, Practice, and the Differentiation of Speed-Up from Automatization Effects: Evidence from Second Language Word Recognition. *Applied Psycholinguistics*, *14*(3), 369-385.

Segalowitz, S., Segalowitz, N., & Wood, A. (1998). Assessing the Development of Automaticity in Second Language Word Recognition. *Applied Psycholinguistics*, *19*(1), 53-67.

Shantz, K. (2017). Phrase Frequency, Proficiency and Grammaticality Interact in Non-Native Processing: Implications for Theories of SLA. *Second Language Research*, *33*(1), 91-118.

Sharwood-Smith, M. (1981). Consciousness-Raising and the Second Language Learner. *Applied Linguistics*, *2*(2), 159-168.

Sharwood-Smith, M. (1993). Input Enhancement in Instructed SLA: Theoretical Bases. *Studies in Second Language Acquisition*, *15*(2), 165-179.

Sharwood-Smith, M. (2008). Revisiting the Role of consciousness with MOGUL. In Z-H Han, E. Park, A. Révész, C. Combs, & J. Kim (Eds.). *Understanding Second Language Process* (pp. 1-15). Clevedon: Multilingual Matters.

Sheen, Y. (2004). Corrective Feedback and Learner Uptake in Communicative Classrooms across Instructional Settings. *Language Teaching Research*, *8*(3), 263-300.

Shintani, N. (2011). A Comparative Study of the Effects of Input-Based and Production-Based Instruction on Vocabulary Acquisition by Young EFL Learners. *Language Teaching Research*, *15*(2), 137-158.

Shintani, N. (2012). Input-Based Tasks and the Acquisition of Vocabulary and Grammar: A Process-Product Study. *Language Teaching Research*, *16*(2), 253-279.

Shintani, N. (2015). The Effectiveness of Processing Instruction and Production-Based Instruction on L2 Grammar Acquisition: A Meta-Analysis. *Applied Linguistics*, *36*(3), 306-325.

Shintani, N. (2016). *Input-Based Tasks in Foreign Language Instruction for Young Learners*. Amsterdam: John Benjamins.

Shintani, N., & Ellis, R. (2010). The Incidental Acquisition of English Plural *–s* by Japanese Children in Comprehension-Based and Production-Based Lessons: A Process-Product Study. *Studies in Second Language Acquisition*, *32*(4), 607-637.

Shintani, N., & Ellis, R. (2014). Tracking 'Learning Behaviours' in the Incidental Acquisition of Two Dimensional Adjectives by Japanese Beginner Learners of L2 English. *Language Teaching Research*, *18*(4), 521-542.

Shintani, N., Li, S., & Ellis, R. (2013). Comprehension-Based versus Production-Based Grammar Instruction: A Meta-Analysis of Comparative Studies. *Language Learning*, *63*(2), 296-329.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic Entrenchment: Prior Knowledge Impacts Statistical Learning Performance. *Cognition*, *177*, 198-213.

Skehan, P. (1996). A Framework for the Implementation of Task-Based Instruction. *Applied Linguistics*, *17*(1), 38-62.

Skehan, P. (1998a). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. (1998b). Task-Based Instruction. *Annual Review of Applied Linguistics*, *18*, 268-286.

Skehan, P. (2001). Tasks and Language Performance. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*, (pp. 167-186). New York: Routledge.

Skehan, P. (2003). Task-Based Instruction. *Language Teaching*, *36*(1), 1-14.

Skehan P. (2007). Language Instruction through Tasks. In J. Cummins & C. Davison (Eds.), *International Handbook of English Language Teaching* (pp. 289-301). Boston, MA: Springer.

Skehan, P. (2009a). Modelling Second Language Performance: Integrating Complexity, Accuracy, Fluency, and Lexis. *Applied Linguistics*, *30*(4), 510-532.

Skehan, P. (2009b). Models of Speaking and the Assessment of Second Language Proficiency. In A. Bennati (Ed.), *Issues in Language Proficiency* (pp. 202-215). London: Bloomsburry.

Skehan, P. (Ed.). (2014). *Processing Perspectives on Task Performance*. Amsterdam: John Benjamins.

Skehan, P. (2016). Tasks versus Conditions: Two Perspectives on Task Research and Their Implications for Pedagogy. *Annual Review of Applied Linguistics*, *36*, 34-49.

Skehan, P. (2018). *Second Language Task-based Performance: Theory, Research, Assessment*. New York: Routledge.

Skehan, P., & Foster, P. (1997). Task Type and Task Processing Conditions as Influences on Foreign Language Performance. *Language Teaching Research*, *1*(3), 185-211.

Skehan, P., & Foster, P. (1999). The Influence of Task Structure and Processing Conditions on Narrative Retellings. *Language Learning*, *49*(1), 93-120.

Skehan, P., & Foster, P. (2001). Cognition and Tasks. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 183–205). Cambridge: Cambridge University Press.

Skehan, P., & Foster, P. (2008). Complexity, Accuracy, Fluency and Lexis in Task-Based Performance: A Meta-Analysis of the Ealing Research. In S. Daele, A. Housen, F. Kuiken, M. Pierrard, & I. Vedder (Eds.), *Complexity, Accuracy, And Fluency in Second Language Use, Learning, and Teaching*. Brussels: University of Brussels Press.

Slobin, D. (1996). From "Thought and Language" to "Thinking for Speaking." In J. Gumperz & S. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70–96). Cambridge: Cambridge University Press.

Slobin, D. (1997). The Origins of Grammaticizable Notions: Beyond the Individual Mind. In D. Slobin (Ed.), *The Crosslinguistic Study of Language Acquisition* (pp. 265–323). Mahwah, NJ: Erlbaum.

Sneddon, J., Adelaar, K., Djenar, D., & Ewing, M. (2012). *Indonesian: A Comprehensive Grammar* (2nd Ed.). London: Routledge.

Spada, N., Jessop, L., Tomita, Y., Suzuki, W., & Valeo, A. (2014). Isolated and Integrated Form-Focused Instruction: Effects on Different Types of L2 Knowledge. *Language Teaching Research*, *18*(4), 453-473.

Spinner, P., & Jung, S. (2018). Production and Comprehension in Processability Theory: A Self-Paced Reading Study. *Studies in Second Language Acquisition*, *40*(2), 295-318.

Stassen, L. (1984). The Comparative Compared. *Journal of Semantics*, *3*(1-2), 143-182.

Strapp, C., Helmick, A., Tonkovich, H., & Bleakney, D. (2011). Effects of Negative and Positive Evidence on Adult Word Learning. *Language Learning*, *61*(2), 506-532.

Suzuki, Y. (2017a). The Optimal Distribution of Practice for the Acquisition of L2 Morphology: A Conceptual Replication and Extension. *Language Learning*, *67*(3), 512-545.

Suzuki, Y. (2017b). Validity of New Measures of Implicit Knowledge: Distinguishing Implicit Knowledge from Automatized Explicit Knowledge. *Applied Psycholinguistics*, *38*, 1229–1261.

Suzuki, Y. (2018). The Role of Procedural Learning Ability in Automatization of L2 Morphology under Different Learning Schedules: An Exploratory Study. *Studies in Second Language Acquisition*, *40*(4), 923-937.

Suzuki, Y. (in press). Optimizing Fluency Training for Speaking Skills Transfer: Comparing the Effects of Blocked and Interleaved Task Repetition. *Language Learning*.

Suzuki, Y., & DeKeyser, R. (2015). Comparing Elicited Imitation and Word Monitoring as Measures of Implicit Knowledge. *Language Learning*, *65*(4), 860-895.

Suzuki, Y., & DeKeyser, R. (2017a). Effects of Distributed Practice on the Proceduralization of Morphology. *Language Teaching Research*, *21*(2), 166-188.

Suzuki, Y., & Dekeyser, R. (2017b). Exploratory Research on Second Language Practice Distribution: An Aptitude× Treatment Interaction. *Applied Psycholinguistics*, *38*(1), 27-56.

Suzuki, Y., & Sunada, M. (2018). Automatization in Second Language Sentence Processing: Relationship between Elicited Imitation and Maze Tasks. *Bilingualism: Language and Cognition*, *21*(1), 32-46.

Suzuki, Y., & Sunada, M. (2019). Dynamic Interplay between Practice Type and Practice Schedule in a Second Language: The Potential and Limits of Skill Transfer and Practice Schedule. *Studies in Second Language Acquisition*, 1-29.

Swain, M. (1985). Communicative Competence: Some Roles of Comprehensible Input and Comprehensible Output in Its Development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235–253). Rowley, MA: Newbury House.

Swain, M. (1993). The Output Hypothesis: Just Speaking and Writing Aren't Enough. *Canadian Modern Language Review*, *50*(1), 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, England: 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.), *Handbook of Research in Second Language Teaching and Learning* (pp. 471-483). Mahwah, NJ: Erlbaum.

Swain, M. & Lapkin, S. (1982). *Evaluating Bilingual Education: A Canadian Case Study*. Clevedon, Avon: Multilingual Matters.

Swain, M., & Lapkin, S. (1989). Canadian Immersion and Adult Second Language Teaching: What's the Connection?. *The Modern Language Journal*, *73*(2), 150-159.

Swain, M., & Lapkin, S. (1995). Problems in Output and the Cognitive Processes They Generate: A Step towards Second Language Learning. *Applied Linguistics*, *16*(3), 371-391.

Swain, M & Lapkin, S (2001). Focus on Form through Collaborative Dialogue: Exploring Task Effects. In M Bygate, P Skehan, & M Swain (Eds.), *Researching Pedagogic Tasks in Second Language Learning, Teaching and Testing* (pp. 99-118). London: Longman.

Talmy, L. (2000). *Toward a Cognitive Semantics: Concept Structuring Systems*. Cambridge, MA: MIT Press.

Talmy, L. (2008). Aspects of Attention in Language. In P. Robinson & N. Ellis (Eds.), *Handbook of Cognitive linguistics and Second Language Acquisition* (pp. 37–54). New York, NY: Routledge.

Taraban, R., & Roark, B. (1996). Competition in Learning Language-Based Categories. *Applied Psycholinguistics*, *17*(2), 125-148.

Taylor, J. (2002). *Cognitive Grammar*. Oxford: Oxford University Press.

Tavakoli, P., & Skehan, P. (2005). Strategic Planning, Task Structure, and Performance Testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239-276). Amsterdam: John Benjamins.

Tavakoli, P., Campbell, C., & McCormack, J. (2016). Development of Speech Fluency over a Short Period of Time: Effects of Pedagogic Intervention. *TESOL Quarterly*, *50*(2), 447-471.

Terrell, T. D. (1991). The Role of Grammar Instruction in a Communicative Approach. *The Modern Language Journal*, *75*(1), 52-63.

Thoms, J. (2012). Classroom Discourse in Foreign Language Classrooms: A Review of the Literature. *Foreign Language Annals*, *45*(s1), s8-s27.

Tomasello, M. (Ed.) (1998). *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Lawrence Erlbaum.

Tomasello, M. (2000). First Steps toward a Usage-Based Theory of Language Acquisition. *Cognitive Linguistics*, *11*(1/2), 61-82.

Tomasello, M. (2003). *Constructing a Language*. Boston, MA: Harvard University Press.

Tomasello, M., & Stahl, D. (2004). Sampling Children's Spontaneous Speech: How Much Is Enough?. *Journal of Child Language*, *31*(1), 101-121.

Tomlin, R., & Villa, V. (1994). Attention in Cognitive Science and Second Language Acquisition. *Studies in Second Language Acquisition*, *16*(2), 183-203.

Toth, P. (2011). Teacher-and Learner-Led Discourse in Task-Based Grammar Instruction: Providing Procedural Assistance for Morphosyntactic Development. *Language Learning*, *61*, 141-188.

Towell, R. (2002). Relative Degrees of Fluency: A Comparative Case Study of Advanced Learners of French. *International Review of Applied Linguistics in Language Teaching*, *40*(2), 117.

Tyler, A. (2012). *Cognitive Linguistics and Second Language Learning*. New York: Routledge.

Vanpatten, B. (1990). Attending to Form and Content in the Input: An Experiment in Consciousness. *Studies in Second Language Acquisition*, *12*(3), 287-301.

VanPatten, B. (1996). *Input Processing and Grammar Instruction: Theory and Research*. Norwood, NJ: Ablex.

VanPatten, B. (2002). Processing Instruction: An Update. *Language Learning*, *52*(4), 755-803.

Vanpatten, B. (2011). Stubborn Syntax: How It Resists Explicit Teaching and Learning. In C. Sanz & R. Leow (Eds.). *Implicit and Explicit Language Learning: Conditions, Processes, and Knowledge in SLA and Bilingualism* (pp. 9-22). Washington, DC: Georgetown University Press.

VanPatten, B. (2015a). Foundations of Processing Instruction. *International Review of Applied Linguistics in Language Teaching*, *53*(2), 91-109.

Vanpatten, B. (2015b). Input Processing in Adult Second Language Acquisition. In B. Vanpatten & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 115–135). New York, NY: Routledge.

Vanpatten, B., & Cadierno, T. (1993). Explicit Instruction and Input Processing. *Studies in Second Language Acquisition*, *15*(2), 225-243.

VanPatten, B., Collopy, E., Price, J. E., Borst, S., & Qualin, A. (2013). Explicit Information, Grammatical Sensitivity, and the First-Noun Principle: A Cross-Linguistic Study in Processing Instruction. *The Modern Language Journal*, *97*(2), 506-527.

Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in Second Language Development from a Dynamic Systems Perspective. *The Modern Language Journal*, *92*(2), 214-231.

van den Branden, K. (1997). Effect of negotiation on language learners' output. *Language Learning*, *47*(4), 589–636.

van den Branden, K. (Ed.). (2006). *Task-Based Language Education: From Theory to Practice*. Cambridge: Cambridge University Press.

van de Guchte, M., Braaksma, M., Rijlaarsdam, G., & Bimmel, P. (2015). Learning New Grammatical Structures in Task-Based Language Learning: The Effects of Recasts and Prompts. *The Modern Language Journal*, *99*(2), 246-262.

van Geert, P. (1991). A Dynamic Systems Model of Cognitive and Language Growth. *Psychological review*, *98*(1), 3-53.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wang, Z. (2014). On-line Time Pressure Manipulations: L2 Speaking Performance under Five Types of Planning and Repetition Conditions. In P. Skehan (Ed.), *Processing Perspectives on Task Performance* (pp. 27–62). Amsterdam: John Benjamins.

Wiener, S., Lee, C. Y., & Tao, L. (2019). Statistical Regularities Affect the Perception of Second Language Speech: Evidence from Adult Classroom Learners of Mandarin Chinese. *Language Learning*, *69*(3), 527-558.

Williams, J. (2005). Learning without Awareness. *Studies in Second Language Acquisition*, *27*(2), 269-304.

Williams, J. N. (2012). Working Memory and SLA. In S. Gass & A. Mackey (Eds.), *The Routledge Handbook of Second Language Acquisition* (pp. 427–441). New York: Routledge.

White, L. (1987). Against Comprehensible Input: The Input Hypothesis and the Development of Second-Language Competence. *Applied Linguistics*, *8*(2), 95-110.

White, L. (2003) On the Nature of Interlanguage Representation: Universal Grammar in the Second Language. In C. Doughty and M. Long (Eds), *The Handbook of Second Language Acquisition* (pp. 19-42). Oxford: Blackwell.

White, L. (2015). Linguistic theory, universal grammar, and second language acquisition. In B. VanPatten& J. Williams (Eds.), *Theories in second language acquisition: An introduction*, (pp. 34-53). New York, NY: Routledge.

Wolter, B., & Gyllstad, H. (2013). Frequency of Input and L2 Collocational Processing: A Comparison of Congruent and Incongruent Collocations. *Studies in Second Language Acquisition*, *35*(3), 451-482.

Yamashita, J., & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly*, *44*(4), 647-668.

Yano, Y., Long, M., & Ross, S. (1994). The Effects of Simplified and Elaborated Texts on Foreign Language Reading Comprehension. *Language Learning*, *44*(2), 189-219.

Year, J., & Gordon, P. (2009). Korean Speakers' Acquisition of the English Ditransitive Construction: The Role of Verb Prototype, Input Distribution, and Frequency. *The Modern Language Journal*, *93*(3), 399-417.

Yule, G. (1997). *Referential communication tasks*. New York: Routledge.

Zhang, X., & Lantolf, J. (2015). Natural or Artificial: Is the Route of L2 Development Teachable?. *Language Learning*, *65*(1), 152-180.

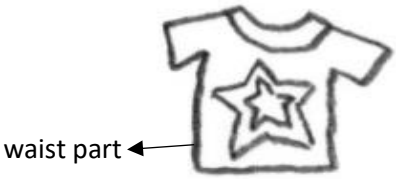
APPENDIX 1. A recruitment task (adopted from Lambert, 2019)

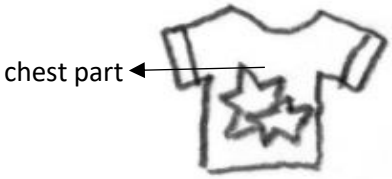
Instruction: You are going to complete a task in which you need to describe a piece of clothing so that I can identify it within a larger group. If your description is not detailed enough for me to select an item from the alternatives, I can only say, “Tell me more” or “I have several like that.” I cannot ask questions or give any hints about what to describe or how to describe it. You can use the vocabulary that has been provided, if necessary. However, you need to speak only English. When I have enough information, I will say, “Okay, I’ve got it.” so that we can confirm the answer.

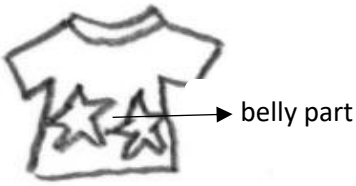
This is NOT a test. Please relax and speak naturally as if you were actually describing a piece of clothing to a friend, and it was important that the friend choose the correct items.

Target referent (to be circled)

Options







APPENDIX 2. Materials for the study

I. A SET OF INSTRUCTION[1]

a. The instruction for the pre-modified input condition

You are going to complete three tasks in which you need to choose three items of clothing accurately based on the description I give you. You have ONLY ONE CHANCE to choose the target items correctly. So, make sure that you listen to all the description I give you carefully and completely. Hence, you are not allowed to choose any item before I finish giving ALL the description about the target item. In the case where you are already able to locate the target item after listening to some part of the description, you need to keep on listening to the description that I give to make sure that it is the item that I really want. Some vocabulary of some specific parts of the target item is already provided in your picture. However, if you still need more information about the target items or do not understand some parts of the description that I give you, you are allowed to ask for clarification by saying 'sorry' or what does . . . mean?'. However, you need to speak only English when asking for more information.

Please keep in mind that this is not a test. So, you do not need to worry about being graded. Just relax and speak naturally as if we were a friend. Do you have any questions before we start?

b. The instruction for the interactional feedback condition

You are going to complete three tasks in which you need to describe three items of clothing clearly and unambiguously to me so that I can choose each of the items you describe correctly from a set of very similar items. The items that I need to choose are the ones that are circled in the paper you are holding. However, in my paper, there are no circles or any other clues. So, I will depend on your description. Because I have only ONE CHANCE to choose the target item correctly, it is important that you describe the target item clearly and unambiguously to me. When I need more information from you, I will ask you a specific part of the item such as how it looks like or how it is different from the same part of the other items. Some vocabulary of some specific parts of the target item is already provided in your picture. Hence, you are asked to describe the item 100% in English. Also, it is important to note that the position of the items in the pictures that we are now holding is different, so you cannot tell me to choose an item based on its position in the picture.

Please keep in mind that this is not a test. So, you do not need to worry about being graded. Just relax and speak naturally as if you were talking to a friend. Do you have any questions before we start?

[1] These instructions will be given in the participants ' L1 (Indonesian)

c. The instruction for the comparison condition (which is the same instruction for the pretest, posttest and delayed posttest)

You are going to complete three tasks in which you need to describe three items of clothing clearly and unambiguously to me so that I can choose each of the items you describe correctly from a set of very similar items. The items that I need to choose are the ones that are circled in the paper you are holding. However, in the picture that I am holding now, there are no circles or any other clues. So, I will depend on your description. Because I have only one chance to choose the target item correctly, you are asked to describe the target item as detailed as possible. I cannot ask questions or give any hints about what to describe or how to describe it. You can use the vocabulary that has been provided, if necessary. However, you need to speak English only.

Please keep in mind that this is not a test. So, you do not need to worry about being graded. Just relax and speak naturally as if we were a friend. Do you have any questions before we start?

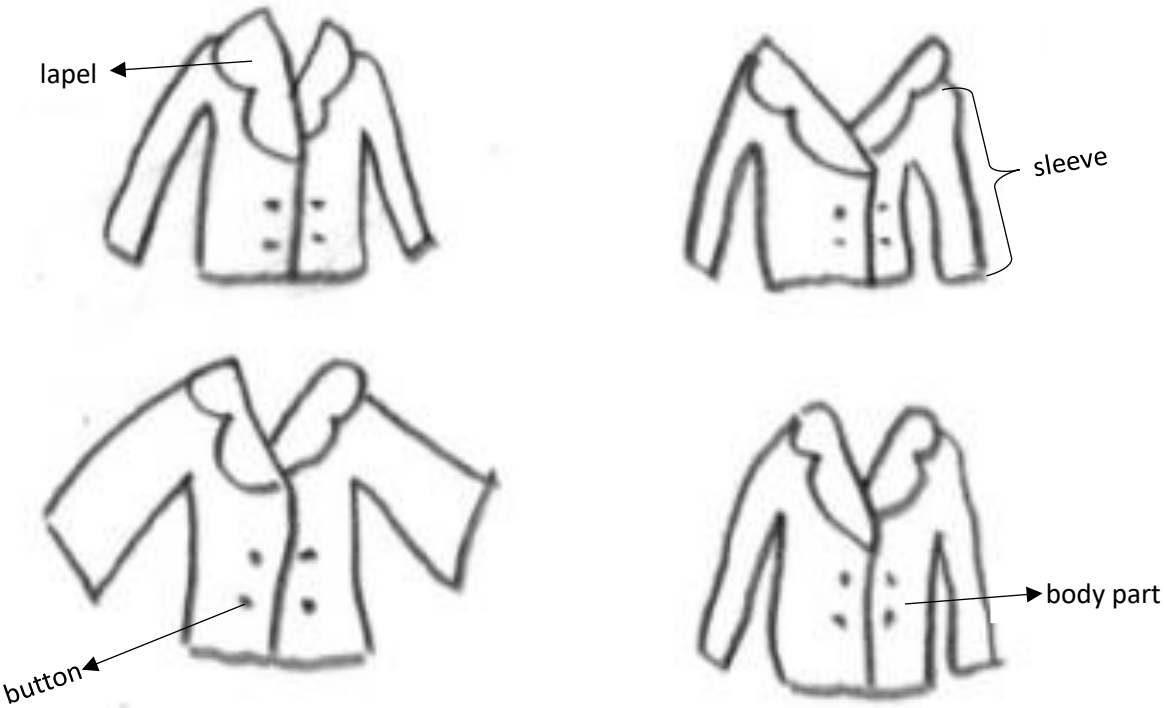
II. A SET OF MATERIALS FOR THE PRE-, POST- AND DELAYED POST-TEST[2]

Set 1

Target referent (to be circled)

JACKET

Right side | Left side

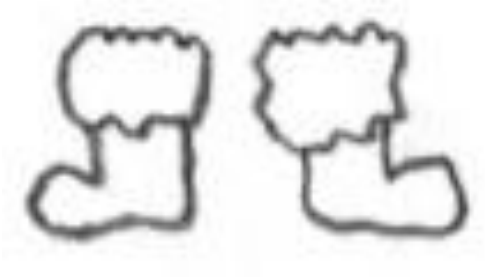


[2] Note: The materials used in this study are adopted and adapted from Lambert (2019).

Set 2

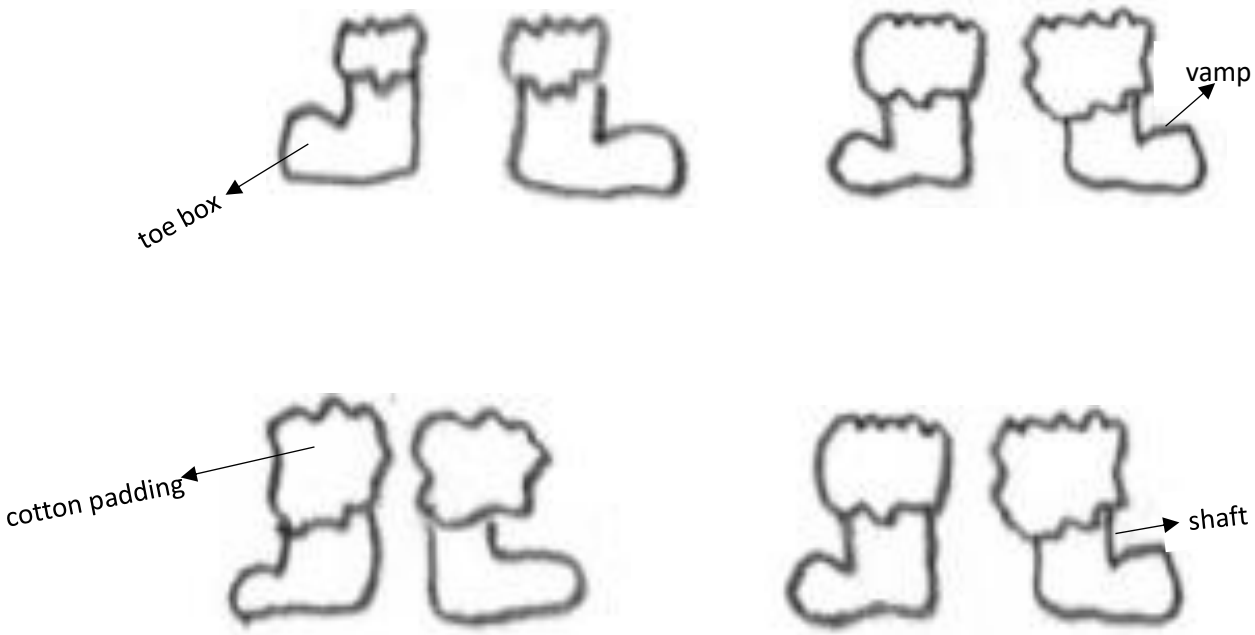
Target referent (to be circled)

BOOTS



Right side

Left side

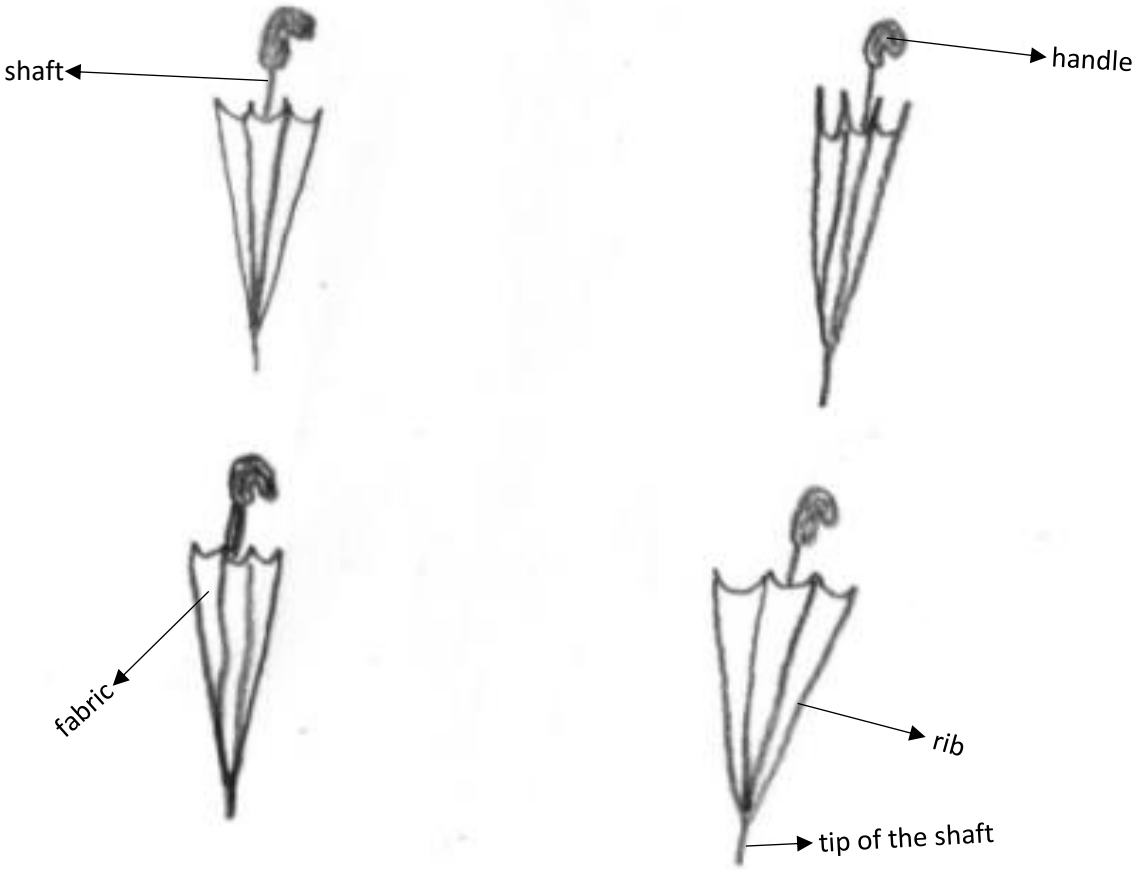


Set 3

Target referent (to be circled)

UMBRELLA

Options

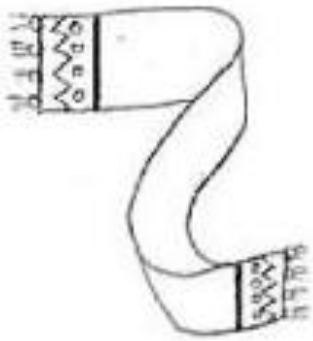


III. A SET OF MATERIALS FOR USE IN THE TREATMENTS

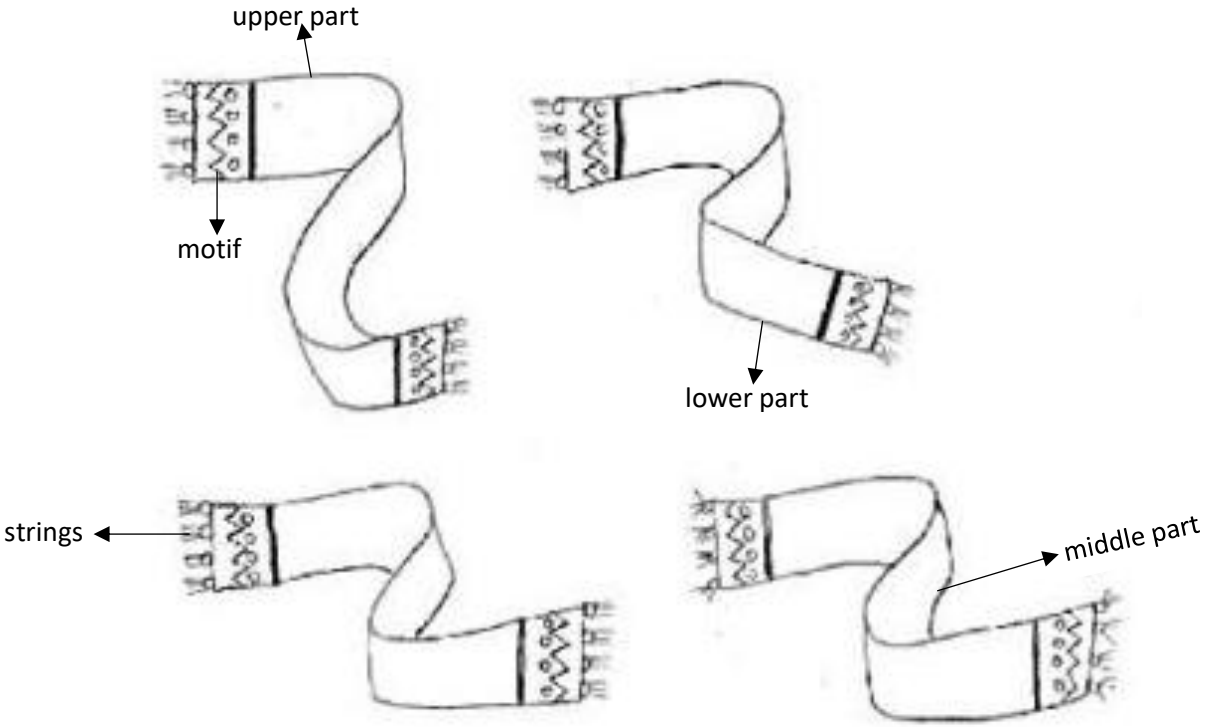
Set 4

Target referent (to be circled)

SCARF



Options



Set 5

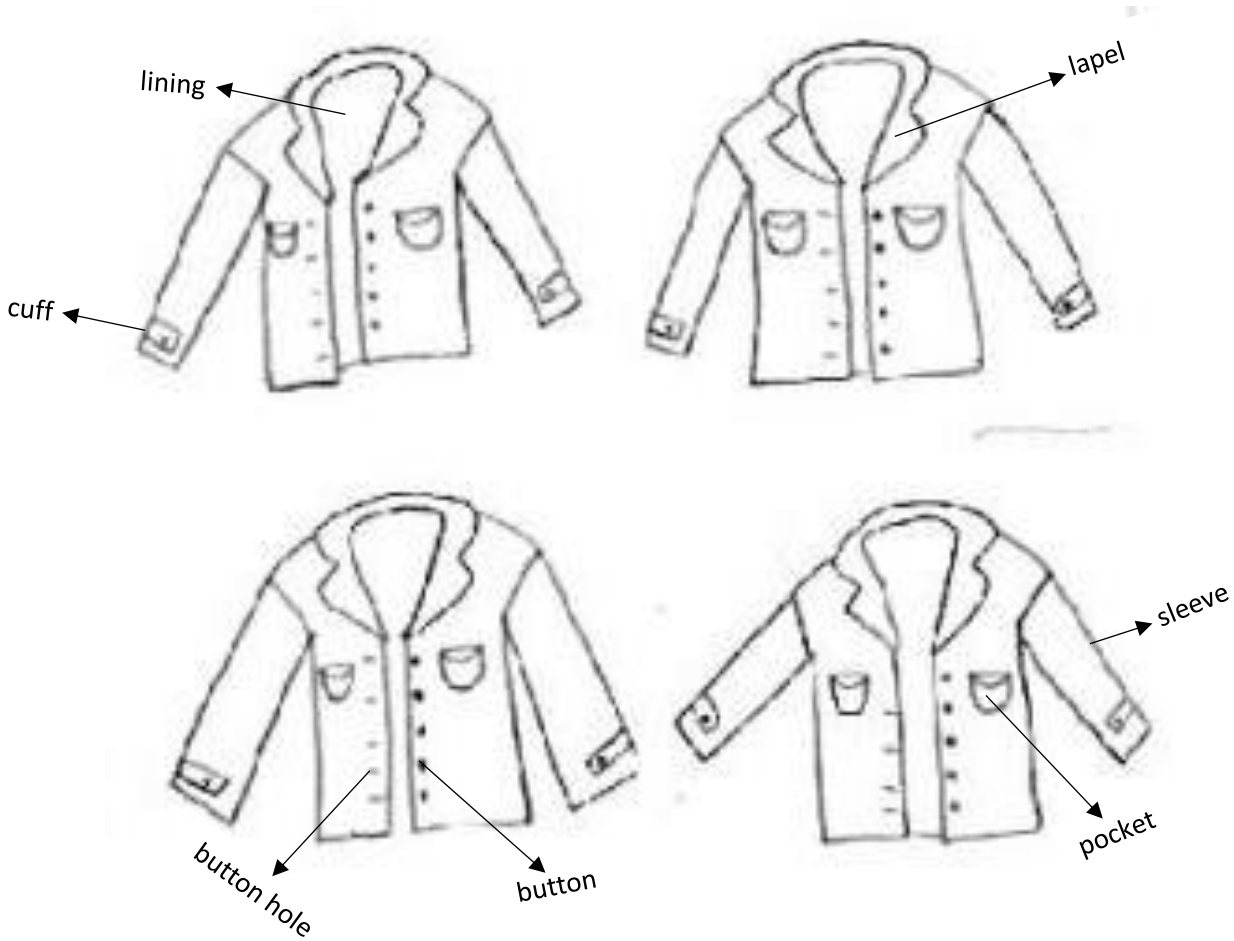
Target referent

JACKET

Options

Right side

Left side

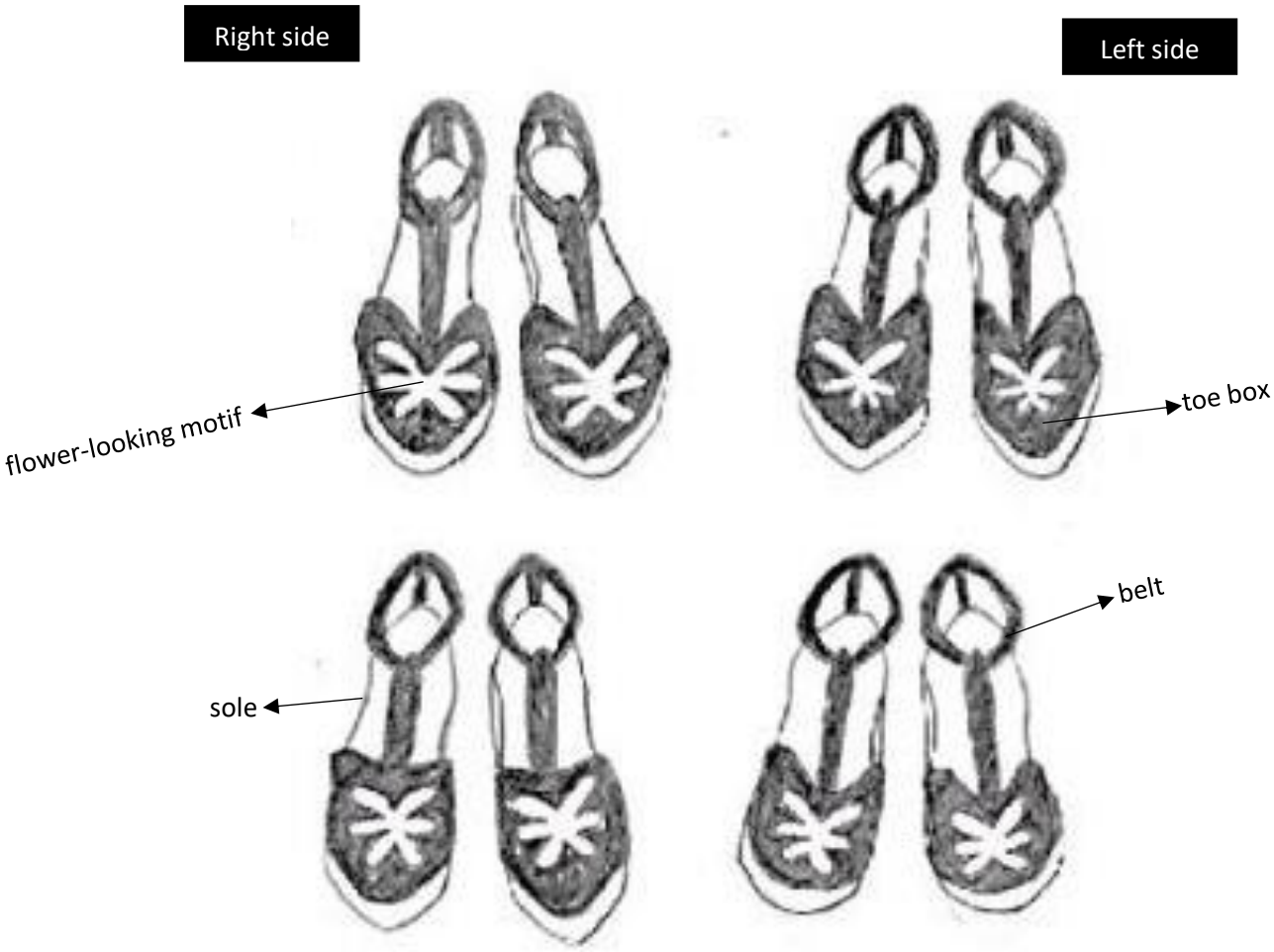


Set 6

Target referent (to be circled)

SHOES

Options

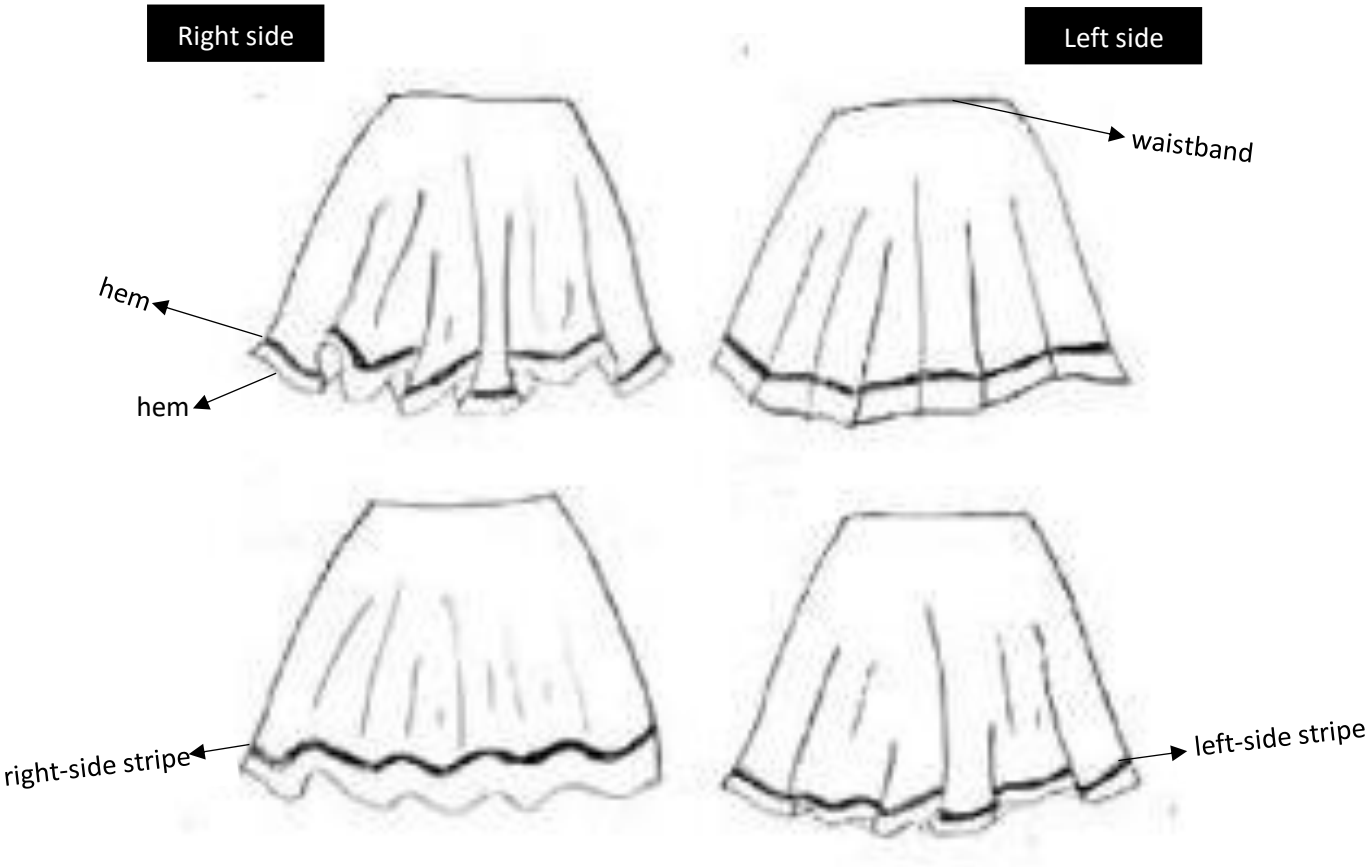


Set 7

Target referent (to be circled)

PLEAT

Options



Set 8

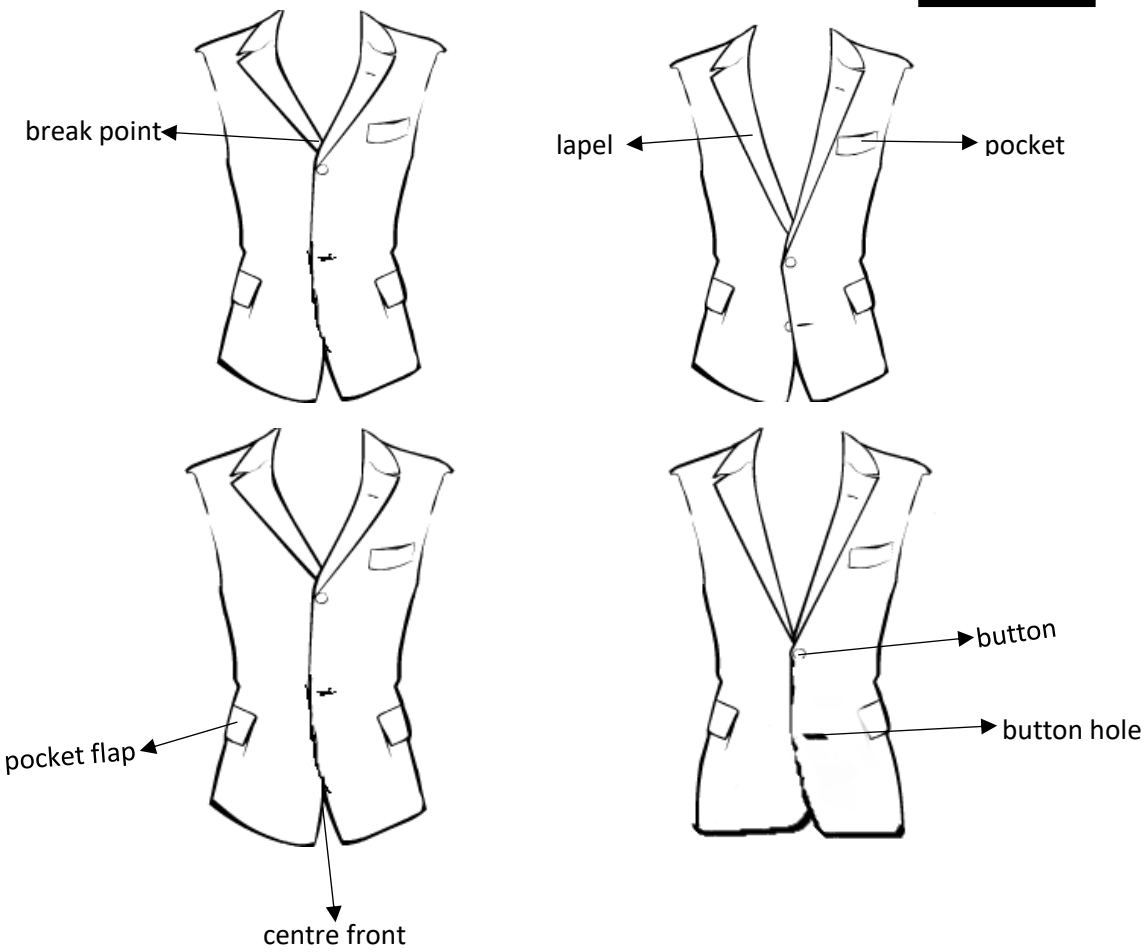
Target referent (to be circled)

VEST

Options

Right side

Left side

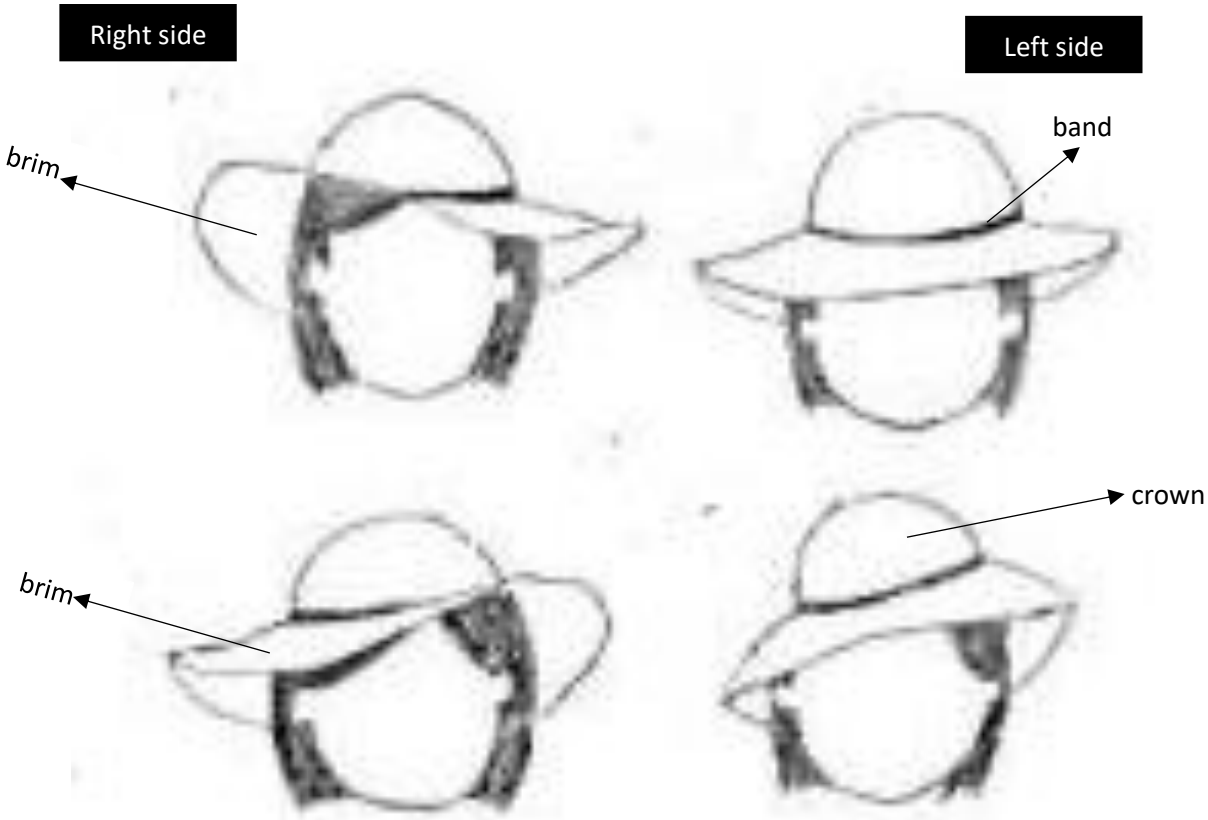


Set 9

Target referent

HAT

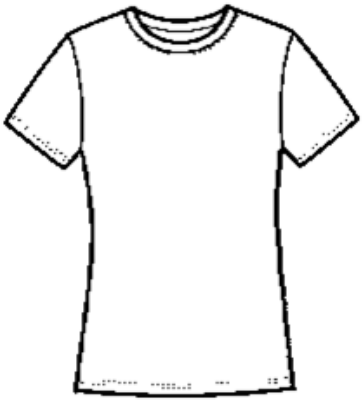
Options



Set 10

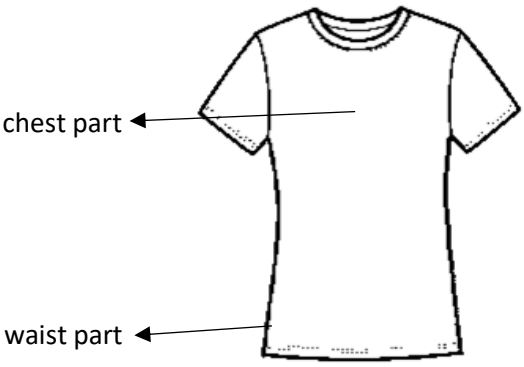
Target referent (to be circled)

T-SHIRT

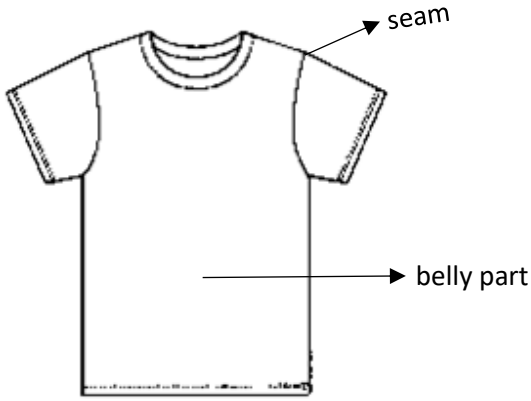


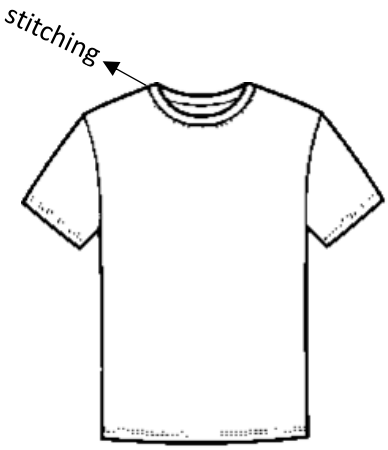
Options

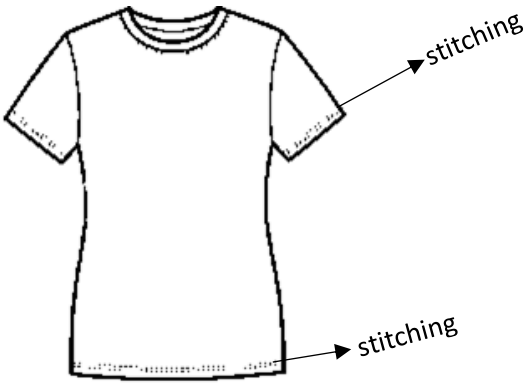
Right side



Left side







Set 11

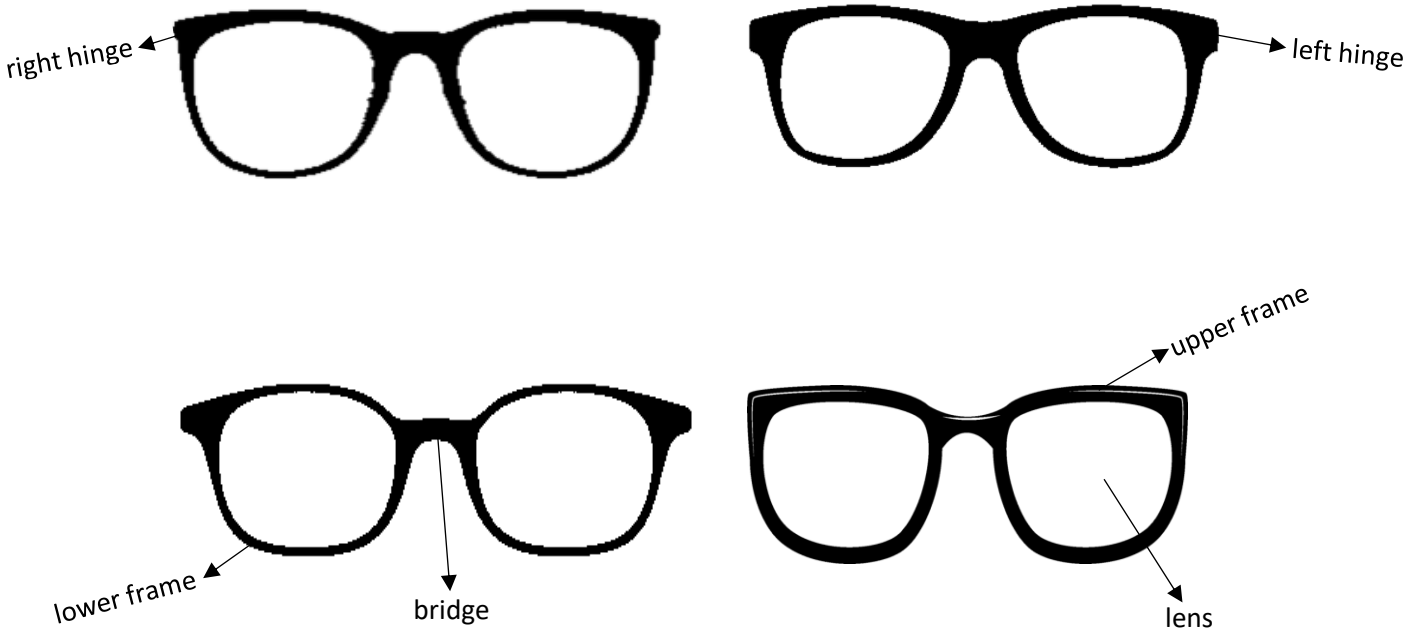
Target referent (to be circled)

GLASSES

Options

Right side

Left side



Set 12

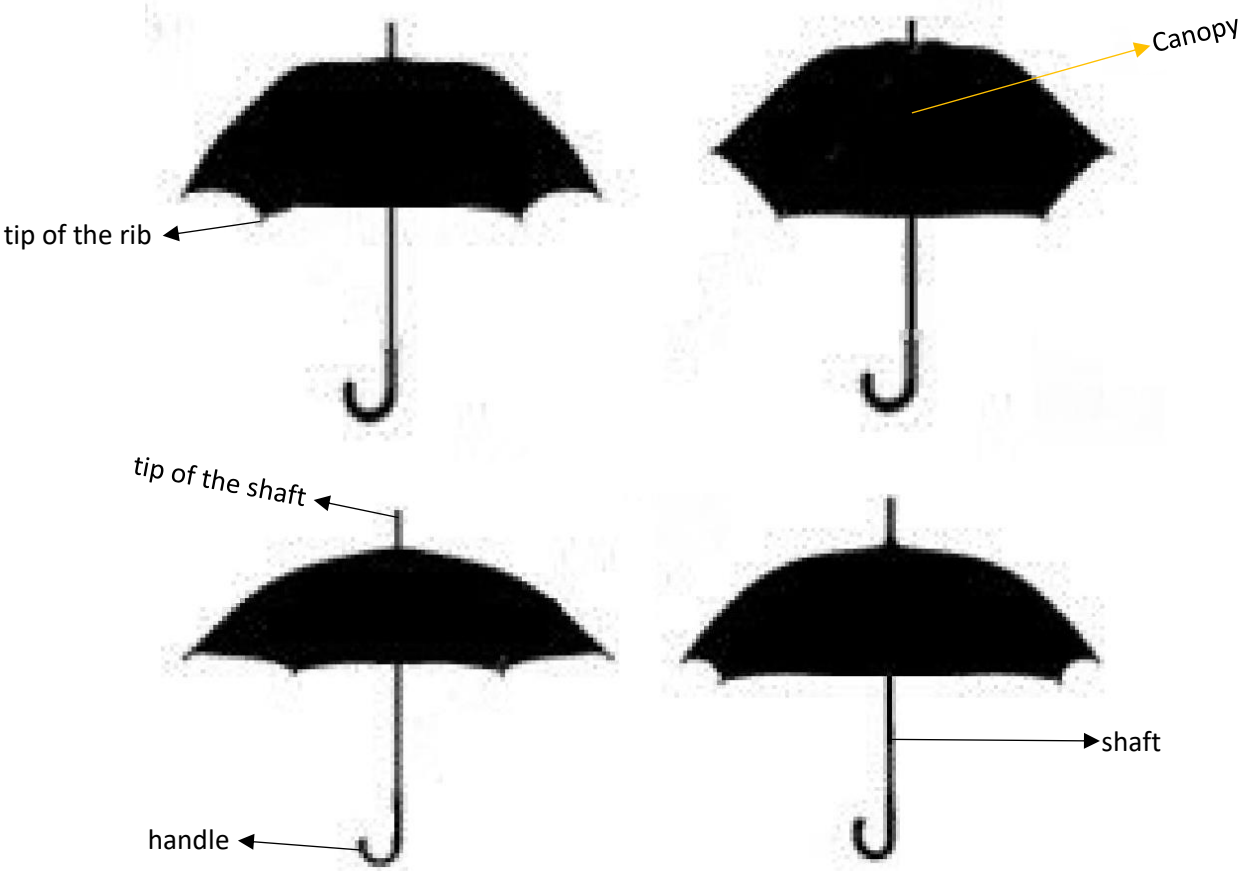
Target referent (to be circled)

UMBRELLA

Options

Right side

Left side



APPENDIX 3. Pre-modified input materials

Task Set 4

Okay, I want you to get me a scarf.

You have four scarves there, but the one that I want you to choose is the one which has a long vertical part.

And if we compare the vertical part of my scarf with that of the other scarves, **the vertical part of my scarf is much longer than that of the other three scarves**.

Okay, now look at the upper part and the lower part of all the scarves.

The upper part and the lower part of my scarf are quite different. That is **the upper part of my scarf is longer than the lower part**.

And **it is also bigger**.

So, **the upper part of my scarf is bigger and longer than the lower part**.

Okay, have you found the scarf that I just described?

I will repeat my descriptions one more time so you can confirm your answer.

Task Set 5

Now I want you to get me a jacket.

You have four jackets there, and they all look similar.

But if you look at the sleeves, you can see that the sleeves of the four jackets are quite different, right?

For example, **there is one jacket whose sleeves are larger or bigger than the sleeves of the other jackets**, but that jacket is not mine.

So, the sleeves of my jacket are not that big. They are slim.

Okay, how many jackets do you have now? Three. Good!

Now focus on the three jackets, and look at the positions of the sleeves of the three jackets. They look quite different, right? Especially the distance between the right-hand sleeve and the body part.

And my jacket is the one whose right-hand sleeve is slightly further away from the main part of the body.

Or the left-hand sleeve of my jacket is closer to the main part of the body than the right-hand sleeve.

So, you can see that the armpit on the right side is wider than that on the left side.

The right-hand sleeve of my jacket also looks very straight.

Okay, now look at the buttons and the button holes.

My jacket is the one which has more buttons than button holes.

That is, while the other jackets have the same number of buttons and button holes, **my jacket has fewer number of button holes**.

Okay, have you found the jacket that I just described?

I will repeat my descriptions one more time for you to confirm your answer.

Task Set 6

Okay, now shoes. I want you to get me a pair of shoes.

You've got four shoes there, but the ones that I want you to take are shoes whose toe-boxes are a little bit pointed. They are not rounded or curved.

Okay, how many shoes do you have now? Three, right? Okay, focus on those three shoes.

Now if you compare the size of the toe-boxes of all the three shoes, you can see that they are not the same. **One of them looks a little bit smaller than the other**, right?

And **my shoes are the ones whose toe-box on the left side is slightly larger**.

Okay, now if you look at the motifs, the flower-looking motifs of all the shoes, you can also see that they are quite different in size.

And for my shoes, **the flower-looking motifs of my shoes are slightly bigger than those of the other shoes**, because, like I said, **the toe-boxes of my shoes are bigger than the toe-boxes of the other shoes**.

Ok, now, the last clue. Look closely at the size between the right and left shoe; **the left side of my shoes is slightly bigger than the right**.

Have you found the shoes?

OK. I will repeat my descriptions one more time for you to confirm your answer.

Task Set 7

Okay, now I want you to get me a pleat.

The pleat has a stripe, just like the other three pleats.

But if you look at the waistband, the waistband of my pleat does not look so symmetrical.

If you look at the left side of the waistband, you can see that it looks slightly higher.

Okay, how many pleats do you have now? Two! Good! Focus on the two pleats.

Now look at the space between the stripe and the hem of the two pleats, **the space on the left side of my pleat is much wider than that on the right side**.

Also, if you look at the hem and the stripe of my pleat, they look wavy and smooth, not abrupt, not flat.

And if you compare the hem and stripe of my pleat with those of the other pleats, **they look much smoother and wavier than the hem and the stripe of the other pleats**.

Have you found the pleats that I just described?

Okay, I'll repeat my descriptions one more time.

Task Set 8

Okay, now I want you to get me a vest.

You have four vests there and all of them have three pockets: two at the bottom and one right at the front of the left chest.

But if you look at the lapels, their lengths are different, right?

That is, **two of the vests there have longer lapels than the others**. And my vest is one of them.

So, now you can focus on **the two vests which have longer lapels than the other two**.

Okay, if you look at the lapels of the two vests, you can see that their positions are quite different.

And if you look at the lapels of these two vests closely, you can see that one of them looks like **it has slightly longer lapels than the other**; and that's my vest.

Okay, I'll give you one more clue.

Look at the centre front of the two vests. You can see that they look quite different as well. It's like **one of them is more widely open than the other**, right? And that's the vest that I want you to take.

Okay, have you found the vest?

I'll repeat my descriptions one more time so you can confirm your answer.

Task Set 9

Now I want you to take me a girl's hat.

The brim of the hat slopes down from the left to the right.

So, you can see that **the left side of the brim is higher**, while **the right side is lower**.

There are three hats that look like the one I just described, right? But the one that I want you to take is the hat which has a pointed brim on the right side.

And the brim of this hat is also longer than the brim of the other two hats.

Also, if we compare this hat with the other two hats, **this hat has a much wider brim**.

Now look at the band between the crown and the brim, you can see that **the band on the right side is more visible than the band on the left side**, right? Because the band on left side is blocked by the brim on that side.

Okay, have you found the hat that I just described?

I'll repeat my descriptions one more time so you can confirm your answer.

Task Set 10

Okay, now I want you to get me a t-shirt.

You've got four t-shirts there, and the one I want you to take it the one whose size is slim.

You have three t-shirts there, right? But the one I want you to take is the one whose belly part is not so straight, but is slightly curved.

How many do you have now? Two. Okay, good! Now focus on the two t-shirts.

Look at the stitching around the waist part now, can you see that **one of the two t-shirts has more stitching, while the other has less stitching?**

The waist part of the t-shirt that has more stitching also looks more curved, right? While the other t-shirt with less stitching is less curved.

And now the t-shirt that I want you to take is **the one which has less stitching** and whose waist part is not so curved.

So, do not chose **the t-shirt which has more stitching**.

Have you found the hat that I just described?

Okay, I'll repeat my descriptions one more time so you can confirm your answer.

Task Set 11

Okay, I want you to get me glasses.

There are four glasses there, right? And they look quite similar.

But if you look at the frames: the upper- and lower-frame. You can see that the four glasses look slightly different.

And for the glasses that I want you to choose is **the glasses whose upper frames are slightly bigger than the lower frames, not so much bigger but are only slightly bigger**.

Okay, how many do you have now? Two, right. Now focus on the two pairs of glasses.

Look at the lens now, **my glasses have more circular lens**, while **the lens of the other glasses that look like mine are a bit more pointed, right? at the lower part, so they are less circular in comparison to the lens of my glasses.**

Now look at the bridge which connects the left and right frame, **the bridge of my glasses is smaller than that of the other glasses**.

The hinges, the left and right hinge of my glasses are also shorter than those of the other glasses.

Have you found the hat that I just described?

Okay, I'll repeat my descriptions one more time so you can confirm your answer.

Task Set 12

Okay, this is the last one.

It is an umbrella: a wide open umbrella.

You've got four umbrellas there, but the one that I want you to take is the one with a large canopy and a big handle.

Okay, if you look at the handle now, you'll see that there are two umbrellas which have a small handle, and two umbrellas that have big handle. And my umbrella is one of the umbrellas which have a big handle.

Okay, now you may focus on the two umbrellas that have a big handle.

Now, if you look at the space between the two ribs at both the right and left side, **the space between the two ribs of my umbrella is closer than that of the other umbrella**.

And **in comparison with the other umbrella, my umbrella also has smaller and flatter canopy**.

So, **the canopy of the other umbrella that looks like mine is slightly bigger and looks more upright or more rounded**.

Okay, one last clue.

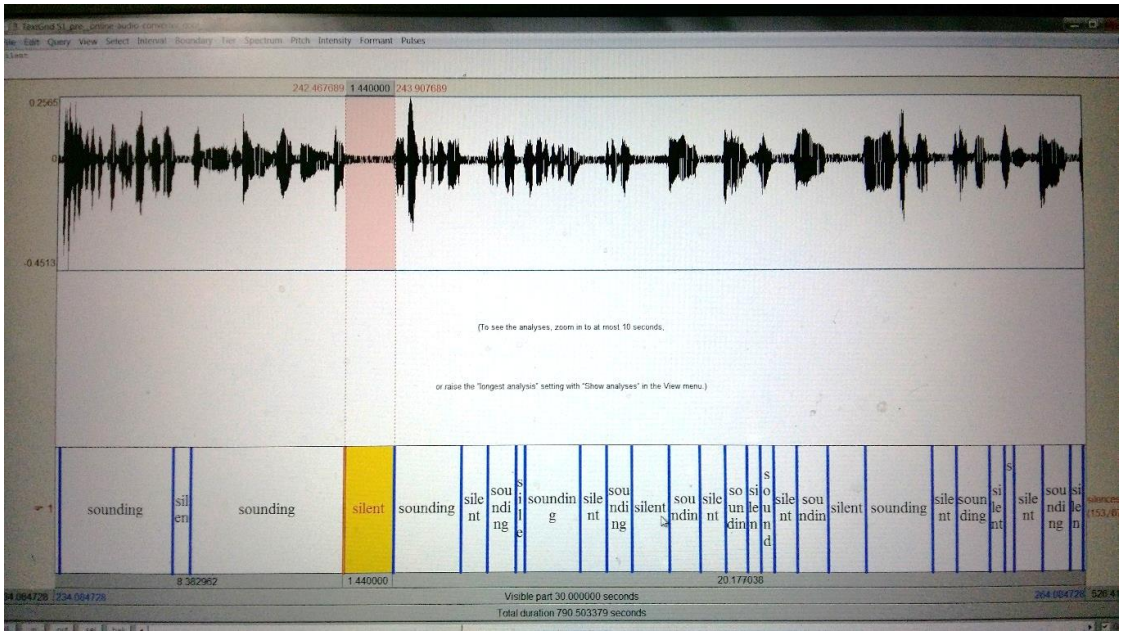
Look at the tip of the shaft. **My umbrella has a longer tip of the shaft than the other umbrella**.

Have you found the umbrella that I just described?

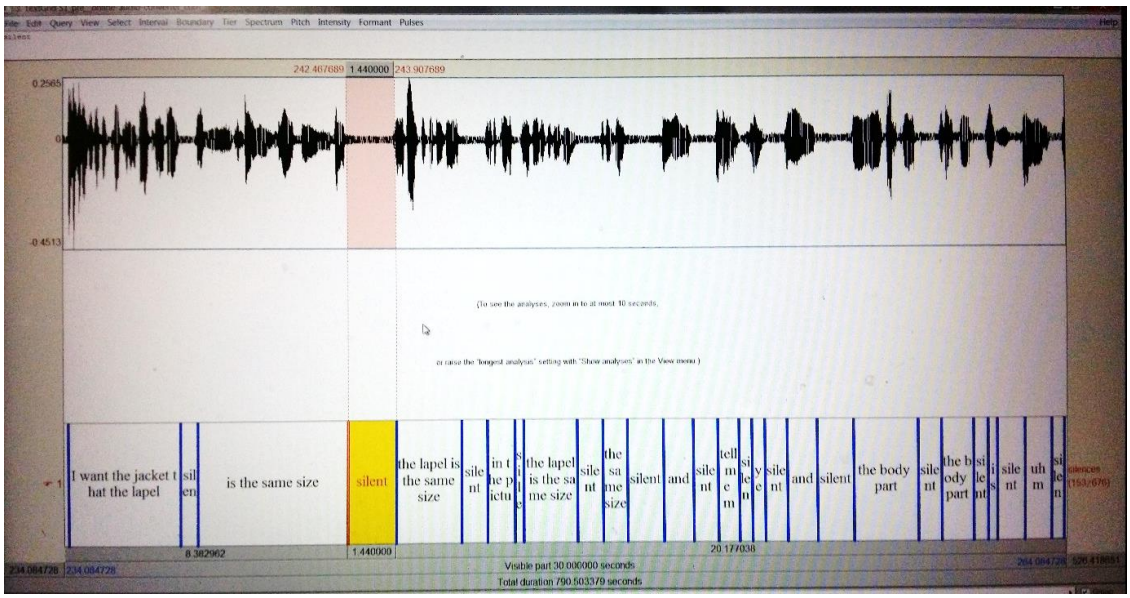
Okay, I'll repeat my descriptions one more time so you can confirm your answer.

APPENDIX 4. Step-by-step procedures of data transcription using PRAAT applied in the study.

1. Boundaries between silences and soundings (utterances) that took place within the learner's task performance in the recording were first created automatically. Silence was set at 0.25 seconds or longer. The picture below indicated where silences and soundings took place in the recording being analysed.



2. Each sounding part was then played and utterances that were produced within it were transcribed manually. In addition, the silent part was also played to make sure that it truly contained no utterance. The picture below shows the results where utterances in the sounding parts have been transcribed.



3. Finally, the complete transcription was saved into the computer using TextGrid to enable the researcher to refer back to the transcription at any time. The column below indicates the conversion of the transcription from PRAAT to TextGrid on the part of the utterances indicated in the picture above.

```
        intervals [150]:
            xmin = 234.2116893424036 (this indicates the time where the utterance started)
            xmax = 237.5396893424036 (this indicates the time where the utterance ended)
            text = "I want the jacket that the lapel"
    intervals [151]:
            xmin = 237.5396893424036 (this indicates the time where the silence started)
            xmax = 238.0516893424036 (this indicates the period where the silence ended)
            text = "silent"
    intervals [152]:
            xmin = 238.0516893424036
            xmax = 242.4676893424036
            text = "is the same size"
    intervals [153]:
            xmin = 242.4676893424036
            xmax = 243.9076893424036
            text = "silent"
    intervals [154]:
            xmin = 243.9076893424036
            xmax = 245.85968934240358
            text = "the lapel is the same size"
    intervals [155]:
            xmin = 245.85968934240358
            xmax = 246.62768934240358
            text = "silent"
    intervals [156]:
            xmin = 246.62768934240358
            xmax = 247.45968934240358
            text = "in the picture"
    intervals [157]:
            xmin = 247.45968934240358
            xmax = 247.71568934240358
            text = "silent"
    intervals [158]:
            xmin = 247.71568934240358
            xmax = 249.34768934240358
            text = "the lapel is the same size"
    intervals [159]:
            xmin = 249.34768934240358
            xmax = 250.11568934240358
            text = "silent"
    intervals [160]:
            xmin = 250.11568934240358
            xmax = 250.85168934240357
            text = "the same size"
    intervals [161]:
            xmin = 250.85168934240357
            xmax = 251.93968934240357
            text = "silent"
    intervals [162]:
            xmin = 251.93968934240357
            xmax = 252.83568934240358
            text = "and"
    intervals [163]:
```

```
        xmin = 252.83568934240358
        xmax = 253.57168934240357
        text = "silent"
    intervals [164]:
        xmin = 253.57168934240357
        xmax = 254.2116893424036
        text = "tell me more"
    intervals [165]:
        xmin = 254.2116893424036
        xmax = 254.59568934240357
        text = "silent"
    intervals [166]:
        xmin = 254.59568934240357
        xmax = 254.9796893424036
        text = "yea"
    intervals [167]:
        xmin = 254.9796893424036
        xmax = 255.65168934240359
        text = "silent"
    intervals [168]:
        xmin = 255.65168934240359
        xmax = 256.5476893424036
        text = "and"
    intervals [169]:
        xmin = 256.5476893424036
        xmax = 257.6356893424036
        text = "silent"
    intervals [170]:
        xmin = 257.6356893424036
        xmax = 259.5556893424036
        text = "the body part"
    intervals [171]:
        xmin = 259.5556893424036
        xmax = 260.2596893424036
        text = "silent"
    intervals [172]:
        xmin = 260.2596893424036
        xmax = 261.2196893424036
        text = "the body part"
    intervals [173]:
        xmin = 261.2196893424036
        xmax = 261.6676893424036
        text = "silent"
    intervals [174]:
        xmin = 261.6676893424036
        xmax = 261.9556893424036
        text = "is"
    intervals [175]:
        xmin = 261.9556893424036
        xmax = 262.7876893424036
        text = "silent"
    intervals [176]:
        xmin = 262.7876893424036
        xmax = 263.6196893424036
        text = "uhm"
    intervals [177]:
        xmin = 263.6196893424036
        xmax = 264.0036893424036
        text = "silent"
```