TITLE
Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency

AUTHOR
Hunter, Ann-Marie

DATE DEPOSITED
23 October 2017

Fluency development in the ESL classroom: The impact of immediate task repetition and procedural repetition on learners' oral fluency

Thesis submitted by:

Ann-Marie Hunter

For the award of Doctor of Philosophy

School of Arts and Humanities

St Mary's University

University of Surrey

May 2017

St Mary's
University College
Twickenham
London

DECLARATION OF ORIGINALITY

Students are reminded that the work that they submit for assessment must be their own.

Please read the following statements and sign and date at the bottom of this form to show that you have complied:

1. This thesis and the work to which it refers are the results of your own efforts. Any ideas, data or text resulting from the work of others (whether published or unpublished) are fully identified as such within the work and attributed to the originator in the text, bibliography or footnotes.

2. This thesis has not been submitted in whole or in part for any other academic degree or professional qualification at this or any other institution.

3. Any chapters that describe the outcomes of joint research should be clearly identified as such with a statement inserted as a footnote on the first page and contributors named. Significant data, images or text resulting from the input of other researchers should be identified as such and attributed to the persons concerned by means of a footnote within the chapter.

4. It is usual to acknowledge the help and guidance of others who have assisted you during your research and preparation of your thesis. Such acknowledgements do not replace or obviate the need for individual attribution as discussed in points 1 and 3.

5. The University College reserves the right to submit electronic versions of your draft documents for assessment of plagiarism using electronic detection software such as 'turnitin'. In addition, whether or not drafts have been so assessed, the University College reserves the right to require an electronic version of the final document (as submitted) for assessment.

SIGNED: ....

PRINT NAME: ANN MARIE HUNTER

DATE: 29/5/17

Acknowledgements

Firstly, I would like to thank my supervisor, Pauline Foster, who has been with me throughout this entire project from conception to printing. I have benefitted from her guidance and insight on so many levels. She has challenged me, offered me opportunities and supported me at every stage, and I am eternally grateful. I would also like to thank my wonderful co-supervisors who have worked with me during different phases of this project: To Martin Bygate, who provided such stimulating ideas and debate at the design stage of the project. To Peter Skehan, whose suggestion to analyse accuracy and complexity greatly enriched the study. To Christine Campbell, whose feedback on the statistical analyses was invaluable.

Thank you also to all the people I have had the good fortune to collaborate and work with over the last five years, in particular Parvaneh Tavakoli and Fumiyo Nakatsuhara, thank you for inspiring me and mentoring me. Thanks also to Frank Boers, Nel de Jong and Craig Lambert - the informal discussions that we have had about this research have enriched and bettered this study.

Very special thanks go to Mauricio Figueroa Candia. Mauricio gave up his own time to train me in how to annotate and code in PRAAT. Without this, the project wouldn't have got off the ground. Mauricio, I can't tell you how grateful I am for your assistance, expertise and encouragement.

Central to my aims in this research was that the research was grounded in pedagogic authenticity. This meant that I was working with genuine students and teachers in intact classes. The students and teachers involved in this study were, without exception, open-minded, enthusiastic and good-natured about being part of this research study. I am therefore indebted to the support of all of the students who were involved in the present study, pilot studies and also all of my students with whom I experimented with the techniques described in these pages. You are all stars! The teachers who let me into their classes and who talked to me about how the task sequences could be tweaked – thank you so much! I would also like to give a particular mention to Andrew Daventry, who was director of studies at the language school where the data collection took place. Andrew went out of his way to help me to conduct this research and allowed me time away from teaching to analyse data. Also, I would like to thank James Baker, the school director, for having absolute confidence in me and allowing me to carry out this research in the school.

Finally, I want to say how much I have been supported by my amazing family and friends. There are too many to mention everyone individually but in particular, I would like to thank my excellent friend Tamsin, my parents, Gareth and Donna, my husband’s family Tessa, Daphne, Michael, Steve and Shirley, who have not only offered moral support but also provided financial assistance, childcare support and been sounding boards for ideas. This brings me to my two favourite people in the world, my husband, Jack and my little boy, Casper. Jack, I love you so much and I cannot thank you enough for your patience, support and creative input over the last five years. Casper, Mammy’s finished now. Let’s go and get an ice-cream.

Fluency development in the ESL classroom: The impact of task repetition and procedural repetition on learners' oral fluency

Ann-Marie Hunter

Thesis submitted for the award of Doctor of Philosophy

St Mary's University

University of Surrey

May, 2017

Abstract

When they have the opportunity to repeat a pedagogic task, learners speak with a higher degree of fluency during the second or subsequent performance. The impact of repetition on learners' fluency on entirely new tasks, however, is less clear. The aim of this study was to investigate the effects of two inherently repetitive pedagogic task sequences on short-term fluency and also the extent to which any increased fluency in the short-term was maintained on a new task (i.e. a 'transfer effect'). One of these task sequences involved repetition of the same task (TR) and the other engaged learners in repeating the task *procedure* but with new content (PR). 64 ESL students were divided among three groups: Task Repetition (TR), Procedural Repetition (PR), and Control (C). Participants in all three groups were recorded speaking during a pre- and post-test stage. In addition, the two experimental groups (TR and PR) took part in a training session between pre- and post-test which was also audio-recorded. For the TR group, the session involved performing a narrative task three times. For the PR group, the session required learners to perform three *different* narrative tasks of the same type (i.e. picture narrative). Participants' oral performances were analysed in terms of a range of fluency measures, representing different aspects of fluency (i.e. speed, breakdown, repair) at pre-test and post-test and also during the training session (i.e. all three performances for both experimental groups). The findings revealed that oral fluency increased during the training session only for the TR group and between pre- and post-test only for the PR group relative to the control. These findings are explained in terms of the different sorts of 'practice' which the two task sequences provide and the different ways they impact upon the speech production process. The methodological and pedagogic implications of the study's findings are also discussed.

Final word count: 92,585

Contents

List of Figures

List of Tables

1 – Introduction

A goal of many second language learners is to speak the language effortlessly in the way that they do their first language. And yet, for the majority of language learners, using the language communicatively remains hard work, even after a substantial period of studying. While certain aspects of their ability might improve over time, it is often their ability to communicate orally that stagnates. Indeed, it is very common to hear someone say of their L2: "I *understand* everything really well, but I just can't speak it". Speaking a second language, then, clearly involves more than simply learning the grammatical rules and vocabulary. It involves somehow *integrating* what is known about the language into communication under the pressures of real-time interaction.

This ability to integrate L2 knowledge into performance is sometimes referred to in lay terms as "thinking in the language", meaning that there is no need to 'translate' L1 thoughts into L2 speech. The popular conception is, at least for low level students, that it takes a long time to conceive of what one wants to say, first in L1 and then laboriously 'translate' the idea into L2 speech. Language researchers, on the other hand, might refer to this ability to 'think in a language' and express ideas effortlessly in another language as 'L2 fluency'.

Understanding and exploring this, at times, elusive concept of L2 fluency is important for a number of reasons. Firstly, 'fluency' features as a criteria for oral assessment in a number of high- and medium-stakes language examinations (e.g. APTIS, IELTS, TOEIC*). This means that many L2 learners are motivated to increase fluency in preparation for examination. Secondly, listeners may find it "tiring and annoying" to interact with highly dysfluent speakers (Derwing, 2017, p. 247; Varonis & Gass, 1982). The knock-on effect of this is that (especially in an ESL context) potential interlocutors may avoid interacting with dysfluent learners (Derwing, Rossiter & Munro, 2002) and therefore learners miss out on the opportunity to get the massive amounts of practice that are necessary for further fluency development (Segalowitz, 2010). Fluency might also impact on interlocutor perceptions about personality and intelligence (Thomson & Isaacs, 2011) and it is also believed to be a key factor involved in predicting overall L2 proficiency (Baker-Smemoe et al., 2014; Iwashita et al., 2008).

Understanding what fluency is, and how it can be achieved by learners, fostered by teachers and measured in tests and in research studies are key concerns in second language acquisition (SLA) and language testing (LT) research. Consequently, research into L2 fluency development has seen great developments in recent decades in terms of conceptualising the construct of fluency (Segalowitz, 2010; 2016), finding new and innovative ways to measure fluency (e.g. N.H de Jong & Wempe, 2009), and developing tools, techniques and task sequences which might be relevant for working with fluency in the language classroom (e.g. Galante & Thomson, 2016; Lynch & Maclean, 1994; Tavakoli et al., 2015; Wood, 2010)

1.1 Task repetition

In recent years, there has been substantial interest in the role that task repetition could play in fluency development (e.g. N. de Jong[1] & Perfetti, 2011; Lambert, Kormos & Minn, 2017). In chapters from a new handbook on instructed second language acquisition (ISLA), both Michel (2017) and Derwing (2017) argue for the usefulness of task repetition in developing fluency. Michel, for example, writes: “Task repetition and familiarity is a fruitful way to foster higher levels of performance in terms of CAF. Repeating a task just once may enhance students’ fluency” (Michel, 2017, p. 59).

Indeed, research findings for the impact of TR on oral fluency are almost unanimous: *task repetition increases fluency*. However, as I will show, there remain a number of unanswered questions in the literature, particularly around the questions of what exactly needs to be repeated in order to benefit learners’ fluency, whether fluency increases in the long-term or whether it is merely temporary, and how other performance markers such as accuracy and complexity are affected by repetition and to what degree. Furthermore, while our conceptualisation and measurement of fluency has developed and become much more nuanced, the majority of studies which have looked at the impact of TR on fluency have conceptualised and measured fluency in a relatively broad sense meaning that it is difficult to make claims about exactly how TR impacts on different aspects of oral fluency and, consequently, how it relates to the underlying speech production process.

[1] The works of *Nel* de Jong and *Nivja* de Jong were unable to be differentiated by first initial therefore Nel de Jong appears as N. de Jong and Nivja de Jong as N.H. de Jong

1.2 The L2 Fluency ‘gap’

Given the desire that teachers have to help students improve their fluency and the findings that TR might perform that function, it might be anticipated that TR features heavily in language classrooms. However, despite the increased *academic* interest and enthusiasm for task repetition, recent research has suggested that TR is not commonly used in L2 classrooms for the purpose of developing learners’ fluency (Tavakoli & Hunter, 2017). In fact, Derwing (2017, p. 253) explains that currently there is a general “consensus that many L2 students do not have much opportunity to enhance their spoken fluency in classrooms” at all (Derwing, 2017, p. 253).

Perhaps the first to acknowledge a gap between research into L2 fluency and pedagogic practice were Gatbonton and Segalowitz (1988). They suggested that the approaches to promoting fluency in the classroom at that time were limited and that pedogogy should focus on enabling learners to automatise utterance and suggested a way to do that in a ‘creative’ way. Much more recently, Rossiter, Derwing, Manimtim and Thomson (2010) conducted a review of teaching materials available to language teachers in an ESL context, investigating student coursebooks and teaching resources, including those specifically geared towards promoting fluency. They found that, particularly in the coursebooks, there was a heavy emphasis on “free communication” activities. These, they point out, are traditionally very popular in English as a Second Language (ESL) teaching and yet they have little empirical support in language teaching research. They found that activities supported by language research such as the teaching of formulaic language, fillers and repetition were present to a much lesser degree. In a similar investigation, Diepenbroek and Derwing (2014) surveyed 48 textbooks that were in use for ESL and Language Instruction for Newcomers to Canada (LINC) courses and, similarly, found that they were not very useful for the development of oral fluency.

What these later studies highlight is that discoveries and developments in language research do not necessarily effect immediate change in language teaching materials. However, an assessment of language teaching *materials* can only go so far towards an understanding of what actually goes on in language *classrooms* with regards to fluency because “it is assumed that teachers will incorporate oral fluency in other ways, rather than relying [solely] on a written textbook" (Diepenbroek & Derwing, 2014, p. 14). Indeed, Foster and Hunter (2016) describe a range of ways that activities in language coursebooks and teaching resources can be easily adapted by teachers to give them a fluency focus.

Tavakoli and Hunter (2017), then, took the investigation a step further by investigating what teachers actually *do* in the classroom with regards to fluency development. Like Derwing et al. (2010), they also found that teachers were considerably more likely to turn to 'free communication activities' to improve students' fluency rather than the sorts of activities which L2 research has suggested might be beneficial for L2 fluency development (e.g. task repetition). Overall, Tavakoli and Hunter (2017) concluded that there was evidence to suggest that the 'gap' which has been discussed in relation to fluency research and pedagogic practice (e.g. Chambers, 1997; Gatbonton & Segalowitz, 1988) was still very much present.

Derwing (2017), however, provides some specific reasons for the fluency gap. She suggests that large class sizes, prioritisation of other skills that need to be taught, time restrictions as well as a lack of awareness about the sorts of activities that may foster fluency all impact on the extent to which there is a 'fluency focus' in language classrooms. In terms of task repetition specifically, Ahmadian et al. (2017) and Hunter (in preparation) found that teachers are reluctant to use TR with classes because they fear that students will find it boring. What all this means is that, while SLA research is attempting to sell TR as a good way to promote fluency in the classroom, teachers do not seem to be buying it.

1.3 The present study

The present study was motivated by a desire to investigate the precise nature of the relationship between different types of task repetition and L2 fluency, as well as looking at how task repetition can be operationalised in the context of a language classroom, the considerations for teachers, and the extent to which contextual, learner and teacher factors might impact on the effectiveness of TR as a pedagogic technique.

This study therefore has a dual purpose and motivation: to understand and advance knowledge in the area of L2 fluency (i.e. a *theoretical* purpose) and to explore practical ways of integrating fluency-fostering activities into language teaching practice (i.e. a *pedagogical* purpose). This dual perspective has implications for the design, data collection, methodology, research questions, and discussion; in other words: *every phase of the project*.

Drawing on the literature in L2 fluency research, it is clear that if a study is to contribute to the *theoretical* debate about TR effects on fluency it will need to fulfil a number of criteria. (1) It will need to draw on existing theoretical models of L2 fluency and speech production (e.g. automaticity, monitoring, proceduralisation) (Segalowitz, 2010). (2) It should use valid and empirically established means of gauging fluency (Michel, 2017; Segalowitz, 2010; 2016). (3) It should be carried out with a suitable number of participants to allow for subsequent statistical analysis to be meaningful (Plonsky, 2013; Plonsky & Gass, 2011). (4) It should be as controlled as possible and demonstrate an awareness of affective factors in order to allow for cross-study comparisons to be made (Revesz, 2014; Samuda & Bygate, 2008). Finally, (5) it should adopt a methodological approach that complements the multi-dimensional nature of L2 fluency (Riazi & Candlin, 2014).

Drawing on research into teacher cognition, perceptions and practice, it was established that if a study is to generate findings that might have *pedagogic* relevance it will similarly need to address certain points. (1) It should define and operationalise key constructs (i.e. fluency) in ways which are meaningful to teachers (Borg, 2009; Tavakoli & Hunter, 2017). (2) It should seek to answer questions which are relevant to language teaching (Andon & Eckerth, 2009; Ellis, 2003; Nassaji, 2012). (3) It should conduct research in an environment that resembles the teaching environment (DeKeyser, 2017; 2010; Foster, 1998; Nunan, 1991). Finally, (4) it should make space for teachers' and learners' views to shape research questions and frame research findings (Nassaji, 2012; Tavakoli & Hunter, 2017)

While most research into L2 fluency and task repetition could be said to adhere to *either* the criteria associated with an academic *or* pedagogic motivation, there have been few attempts to integrate the two, possibly because it presents a number of challenges to the researcher as the following, from DeKeyser (2017), highlights:

> (W)hat is sorely needed from an applied perspective [on skill acquisition] is studies that are... carried out in a classroom context, yet look closely at very specific processes in a controlled design, in other words, studies that combine ecological validity with internal validity… This is, of course, a tall order (DeKeyser, 2017)

This study will therefore answer DeKeyser's (2017) call for studies that combine ecological validity with internal validity. As he implies, however, this will involve overcoming a number of challenges that are specific to carrying out this sort of multi-purpose empirical research.

1.4 What is the layout of this thesis?

Having introduced the overall aim of this research and given a brief introduction to some of the key themes and constructs, I will now present an overview of the layout of the thesis. In Chapter 2, I explore the concept of *fluency,* firstly looking at the speech production process. I look at how fluency can be defined and measured for research purposes and I draw on Segalowitz' (2010) framework for thinking about fluency in terms of three 'domains': *cognitive, perceived,* and *utterance.* In terms of fluency *measurement,* I draw on a number of studies which have identified reliable indices of fluency in order to provide descriptions of key measures and what aspects of fluency (speed, breakdown, repair) they purport to measure.

Chapter 3 provides an overview of sources of influence on L2 fluency development, including the speakers' fluency in L1, proficiency level and *practice*. In particular, I note the pedagogic interventions that have been found to impact on L2 fluency. I suggest that a dynamic systems framework might be most appropriate for thinking about L2 fluency development given its complex, multi-dimensional nature.

Chapter 4 introduces the key concepts of 'task', 'task-based language teaching' and task repetition. I begin with a look at how repetition of language is a part of everyday life and is also key to language learning, and in particular the development of L2 fluency. I look at how task repetition can take different forms, with some studies investigating repetition of the exact same task and others looking at similar tasks. I also provide the theoretical underpinnings for TR.

Chapter 5 provides an in-depth synthesis of the published research on TR with a particular focus on TR effects on L2 *fluency*. First, I consider the ways in which studies have varied in terms of the participants they have worked with and the ways in which the dependent and independent variables have been operationalised. I explain that the body of research can be divided into that which has looked at short-term effects of TR on fluency, long-term effects, and learners' perceptions of TR. I present the main findings of these studies in these three areas, highlighting particular issues which remain unanswered.

In Chapter 6, I outline the rationale for the present research, identifying a number of lacunae in the task repetition/fluency research in terms of empirical questions and methodology. This leads to the three broad research questions which have guided this research and the related hypotheses that these questions have generated.

In Chapter 7, I turn to the methodological approaches which have been used in TR/ fluency studies. I establish commonalities in the methodological approach of studies and identify divergences. I use this as a springboard to discuss the methodological approach of the current research, providing detailed reasoning for the choice of mixed methods, in particular and the research location of a language classroom.

In Chapter 8, I discuss the methodological *procedure* of the present research including how it was informed by a pilot study. I introduce the design of the study, information about the participants, context and tasks. I present the dependent variables which are used in the study along with justification from the literature.

Chapter 9 explains how the data were coded and analysed. In particular there is detailed information about the speech analysis that was carried out using the computer software PRAAT (Boersma & Weeninck, 2013).

In Chapter 10, I present the results of the quantitative analyses while Chapter 11 takes a qualitative look at two nestled case studies from the data set. I use these studies to explore the construct of automaticity in L2 fluency and look for explanations for the trends observed in the quantitative data.

In the discussion section (Chapter 12), I return to my original hypotheses and explore how the findings of the study relate to them. I offer explanations for any deviancies from existing patterns observed in the literature and provide theoretical explanations for findings.

Chapter 13 presents what I see as the theoretical, methodological and pedagogic implications of the study as well as its limitations and directions for future research.

2 – Second Language Fluency

2.1 Introduction

Understanding what L2 fluency is and how it develops is clearly of interest to L2 learners, testers, teachers and researchers. As we will see, however, fluency is a complex, multi-faceted concept in terms of its definition, measurement and development over time. This chapter begins with a short exploration of some current theories related to the speech production process that will form the basis for discussions about what fluency is, how it can be measured and how it develops. This is followed by two main sections. The first section, 'Defining Fluency', will explore what exactly is meant by L2 'fluency' in qualitative terms. This will involve an examination of the L2 speech production process and a discussion of automaticity in L2 speech processing. The second section, 'Measuring Fluency' investigates how L2 fluency might be reliably captured for L2 research and testing purposes. I conclude with a short chapter summary.

2.2 Current theories of L2 speech production

There are a number of models which have been used to capture the process of speech production (Kormos 2006), in terms of both L1 and L2 speech. Many of these models share a common base, drawing on Levelt's oft-cited and hugely influential work on the native speaker (1989). A recent and comprehensive model, and one which focuses specifically upon fluency in second language speech production, is provided by Segalowitz (2010), and will guide the discussion of fluency in this chapter.

Levelt (1989) described the process of L1 speech production and later offered a "blueprint" of the native speaker (Levelt, 1999) (Figure 1). This model posits that the production of speech requires speakers to first *conceptualise* their intended message, then *formulate* the message by drawing on knowledge stores before *articulating* the message as overt speech. He also suggests that the process is *monitored* at a number of different levels.

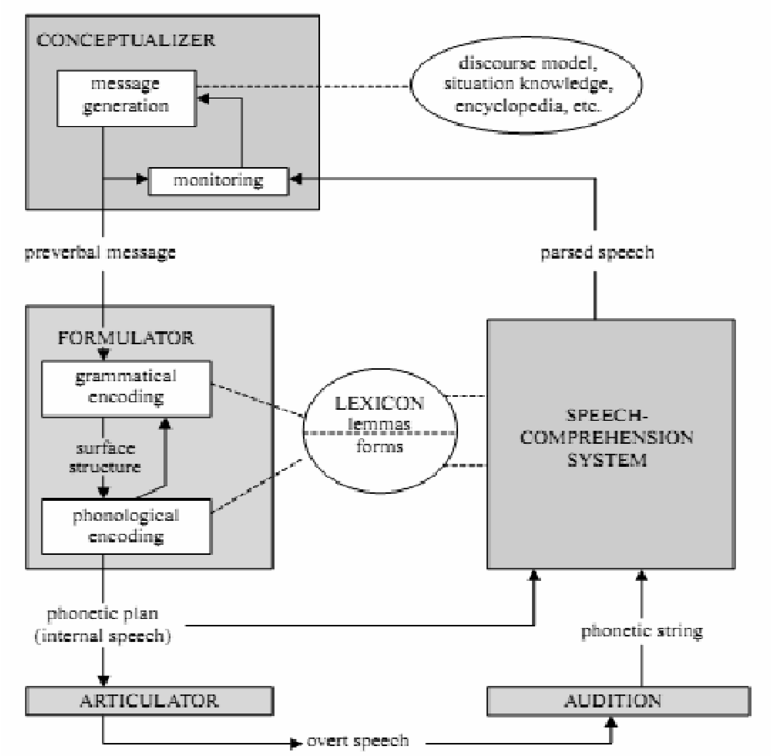


Figure 1: The blueprint for the native speaker (from Levelt, 1999)

This model has since been revised and adapted to incorporate L2 speech (de Bot, 1992; Kormos, 2006; Segalowitz, 2010). These revised models (Levelt, 1989, De Bot, 1992, Kormos, 2006) all agree that the four broad stages of speech production (regardless of whether someone is speaking in L1 or L2) are *conceptualisation, formulation, articulation and self-monitoring* and that these stages happen in the same order (Kormos, 2006). However, these models vary in terms of the exact nature of the formulation process and the way that linguistic knowledge is organised (e.g. whether there is a shared vocabulary store for L1 and L2 knowledge).

Differences do exist between speech production processing in L1 and L2. Hilton (2008) explains that conceptual and discursive planning can be described as 'higher-order' or 'meaning-related' processes, and the linguistic encoding processes such as lexical selection, morpho-syntactic and phonological encoding can be described as 'lower-order' processes. She goes on to explain that in L1, the lower-order processes are highly automatic. She writes:

> "in our L1…we do not have to 'pay attention' to how we are going to articulate a word, conjugate a verb, or place an adverb in an utterance. We may occasionally find ourselves actively 'looking for' a word or language form that momentarily escapes us,

> but this is relatively rare, considering how many thousands of words we produce in our L1 every day" (Hilton, 2008, p. 153-154).

This means that L1 speech is generally smooth and fast. For L2 speech, however, especially for speakers at lower levels of proficiency, both higher-order and lower-order processes may involve conscious attention which results in speech which is slower and more likely to be characterised by hesitancies.

Levelt's (1989; 1999) model and de Bot's (1992) and Kormos' (2006) updated versions provide the theoretical backdrop for much L2 oral production research, however, the current study will make use of Segalowitz' model of the L2 speaker (see Figure 2) which is particularly relevant due to the fact that the focus is on how the speech production process relates to *fluency.* It also incorporates de Bot's (1992) and Levelt's (1999) amendments to Levelt's original model and therefore constitutes an up-to-date and comprehensive point of departure.

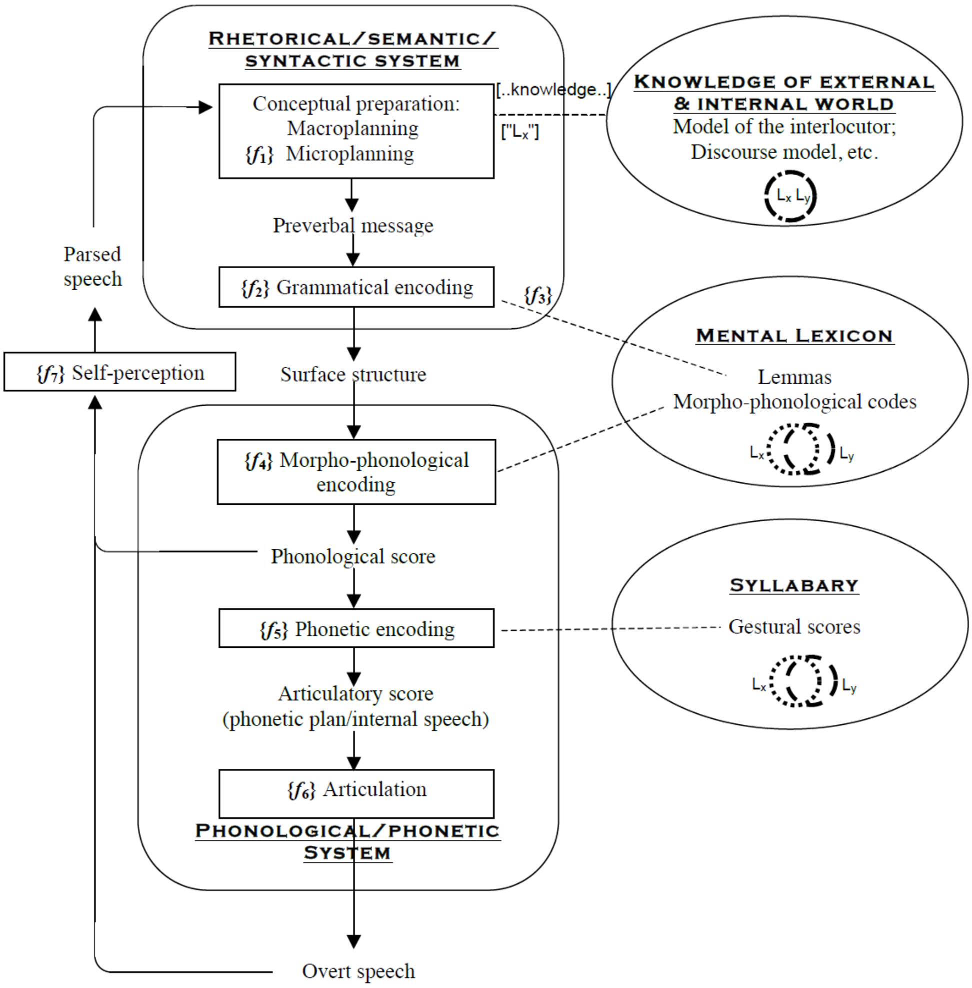


Figure 2: The L2 speech production process (Segalowitz, 2010)

I will look now in a little more detail at each of these stages before discussing how speech production processes relate to L2 fluency. As already mentioned, the first part of the speech production process, *conceptualisation,* involves the speaker deciding what she[2] wants to say (Levelt, 1989). This, in turn, can be broken down into two stages of *macroplanning* and *microplanning* (Segalowitz, 2010). Macroplannning involves planning what to say based on the speakers understanding of the world and her impressions about her interlocutor. It also involves decision-making in relation to register choice (i.e. formal or casual speech). For a bilingual speaker, it has been suggested that it is at this stage that speakers select which of

[2] I have taken a decision to use feminine pronouns for learners and masculine pronouns for teachers throughout

their languages to use (de Bot, 1992; Paradis, 2004) much like they do when selecting the particular register to use.

Although the macroplanning process has established *what* the speaker wants to say, in what language, and in what register, both Levelt (1989) and de Bot (1992) agree that, at this stage the plan is not yet “organised in language specific terms” (Segalowitz, 2010, p. 10). In other words, although the speaker knows that they want to talk about a tree and they know that they need to say this in French and they know that the discourse context is a casual one, they have not yet ‘chosen’ the surface form of the utterance (i.e. *arbre*). This generation of specific surface structures happens at the *microplanning* stage of conceptualisation. Microplanning is still conceptual in nature and involves the creation of a pre-verbal plan for an utterance by making decisions about lexis that will need to be retrieved and also more specific information about the speaker’s attitude, position and so on.

The macro- and microplanning stages of conceptualisation result in a *preverbal plan.* This plan has all the conceptual information needed regarding the subject of the message, the language and register, and the particular words needed to produce the message satisfactorily. De Bot (1992) points out that for an L2 speaker, the pre-verbal message may contain lexis that the speaker does not actually know. For example the preverbal plan that a speaker has may include the concept of an ‘interview’ and the speaker may also be aware that the language required for this communication is French (the speaker’s L2). However, the speaker does not necessarily know that the French word for ‘interview’ is *entretien.* Although, de Bot notes that this is likely to be the case for L2 speakers, we can also imagine that this is the case when, even in our first language, we refer to objects or people as “thingummy, “what’s-it-called”, “what’s-her-name” and “whatchamacallit”. It is important to understand, then, that the conceptualisation stage involves message generation but is preverbal.

The next stage in Levelt’s model is *formulation.* Here the preverbal message is given its surface structure. In other words, the “linguistic shape” is provided by grammatical encoding (Segalowitz, 2010, p. 12) which dictates the order that words will have in an utterance and their relationship with one another. Next, the speaker must draw on the mental lexicon in order to finish the clothing of the message. Segalowitz (2010) notes that this is the stage where concepts become language specific. Although he also suggests that it does not mean that a bilingual has a different store (located in a “distinct, neural region” (p.13)) for L1 lexis and Lx lexis. Instead, he describes how the language exists in close proximity, in the same way that related words and synonyms within a single language are presumed to be located in “neurally related regions” of the brain (Paradis, 2004).

One difference that bilingual speakers do have is a different history of word learning. Whereas for L1, words are learnt implicitly and form what we call the ‘lexicon’, L2 speakers will have some L2 lemmas which are acquired in this way but also a ‘vocabulary store’ which is made up of lemmas which have been learnt *explicitly* and is therefore a type of declarative knowledge. Accessing declarative knowledge, as opposed to procedural knowledge, is presumed to require more attention and time and therefore presents an additional challenge for L2 speakers.

Another difference for L2 speakers is that lemma retrieval may take longer in L2 when compared to L1. Segalowitz (2010), citing work by Kroll and Stewart (1994) and Kroll and Tokowicz (2005), poses the question of whether L2 lemmas are retrieved directly in the L2 (e.g. a speaker has the concept BOY then retrieves the L2 lemma *garcon*) or whether the L2 lemma is accessed *via* the L1 (i.e. speaker has the concept BOY then thinks of L1 *boy* then thinks of L2 *garcon).* He also asks whether this is linked to proficiency or to the particular lemma (i.e. are some accessed in one way and others in another) Lexical retrieval is clearly a point at which disfluency can occur for L2 speakers (de Bot, 1992). Indeed, de Bot argues that it is *the* main point at which disfluency occurs in L2 speech production. If we think back to some discussions that we have had with people who are learning a language it is common to hear things like “I can’t help translating in my head”.

The third stage is *articulation*. Lemmas are selected to fill the syntactic shape that has already been chosen and complete the surface structure of the utterance. And it is therefore ready to be converted into overt speech. Once the lemmas have been selected, it is next necessary to work on translating the surface structure into overt speech. Part of this challenge is accessing information about the phonology associated with lemmas (i.e. what they sound like). Levelt (1999) suggests that each lemma has with it a phonological code (lexeme) that speakers use to create a *phonological score* for their utterance. This presents a further challenge for the L2 speaker, as this is likely to be less automatic in L2 than L1 (de Bot, 1992).

The phonological score must then be converted into an “articulatory score” (Segalowitz, 2010, p. 15) to tell the speech organs (lungs, pharynx, lips, tongue etc.) what they should do in order to produce the required utterance. This relies on the speaker drawing on another information store, this time the *syllabary* which contains information about how to create sounds (gestural scores). Fluency issues may arise at this point when a speaker has to remember to select an L2 gestural score over a (more dominant) L1 score. This works on the assumption that bilingual speakers do not have a separate store for L1 gestural scores and L2 gestural scores. This articulatory score then converts into overt speech by “setting into action the motor activity for articulating the message and creating the overt speech” (Segalowitz, 2010, p. 16)

An additional stage in the speech production process is *monitoring*. In Kormos’ (2006) thorough discussion of the monitoring process in L2 speech (Kormos, 2006), she refers to Levelt’s (1983; 1989; 1993) account of monitoring happens in L1, in which there are three monitoring ‘loops’. The first loop allows speakers to assess the preverbal plan for its comparability to the original intentions of the speaker. In the second loop, the now-clothed message is monitored before it is articulated (covert monitoring). The third loop is an external loop of monitoring, in which the speaker hears the overt speech and checks it for suitability for purpose. As with other aspects of speech production, it is difficult to test specific theories of monitoring. It is clearly problematic from a research point of view because ‘covert’ monitoring (monitoring and altering the message before articulation) is a silent process. One approach is to analyse the speech with the aid of retrospective interviews with the speakers (see Kormos, 2000, 2003) which allows the researcher to understand what was going through a person’s mind as they were speaking (and covert monitoring their speech).

The extent to which a speaker monitors her output and the way she deals with it is believed to change as general proficiency increases (Evans, 1985; O’Connor, 1988; Verhoeven, 1989). It may be that the number of repairs remains fairly constant but the type of repairs changes (O’Connor, 1988). Lennon (1990) found that after a period of residence in the target language (TL) community, learners’ tendency to self-correct actually went up. He took this to be an indication of learners having more attention available to devote to the monitoring process. Based on the assumption that advanced learners make fewer errors due to “the various mechanisms of the development of automaticity” (Kormos, 2006, p. 134), Kormos (2006) suggests that “error repairs signal not yet fully automatized processes; thus they can serve as good indicators of automaticity in L2 speech production” (2006, p. 134)

A final point on monitoring is that it might hold additional value beyond the purely performance-related. Drawing on research by de Bot (1996), Izumi (2003) and Kormos (1999), Kormos (2006) argues that monitoring has a key role to play in the SLA process. It is suggested that the monitoring process helps learners to identify ‘gaps’ in their interlanguage, become aware of errors in their overt speech, “test hypotheses about the L2, trigger creative solutions to problems, and expand [their] existing resources” (Kormos, 2006, p. 135).

Summary of speech production in L2

Speech production processes in L1 and L2 are comparable in terms of the broad stages that a speaker must go through. Differences between the two arise in the sense that there are certain ‘vulnerability’ points in the process in L2 (Segalowitz, 2010) whereby effortful processing causes delays and bottlenecks in the flow of the process and therefore in the delivery of overt speech. It is believed that L1 speakers do not encounter such difficulties because they rely on automatic and parallel processing at these particular points. This explains why L1 speech is, generally speaking, more smooth and speedy than L2 speech. ‘Smoothness’ and ‘speediness’ are often given as definitions of ‘fluent’ L2 speech and it is to this subject that I will now turn.

2.3 Defining fluency

Defining fluency is tricky partly because exactly what fluency means can differ depending on whom you are speaking to. In this section, I look first at what it might mean to be ‘fluent’ in a first language and then what it means to be fluent in an L2. I will explain that ‘fluent’ can have both a ‘broad’ and a ‘narrow’ definition in an L2 before suggesting that Segalowitz’ (2010) conceptualisation of fluency can be useful in pulling together the various qualitative definitions of L2 fluency into a single model which consider fluency in three different domains: *cognitive, perceived and utterance fluency*.

2.3.1 Fluency in L1

Although discussions of 'fluency' are most often reserved for L2 speech, and, indeed all native speakers are described as 'fluent' in their first language(s) (Hilton, 2008), we can all think of certain people who are not quite as talented at putting their thoughts into speech as others. We might talk about these people as being less *articulate* or *eloquent,* notions which are certainly related to being 'fluent'. In a classic paper on the subject of oral fluency, Fillmore (1979, p. 92) explained that in L1 "(t)he word 'fluency' seems to cover a wide range of language abilities, these individually perhaps best described with terms like articulateness, volubility, eloquence, wit, garrulousness, etc."

Fillmore (1979) goes on to identify four ways in which speakers are judged to be fluent. The first of these is "the ability to talk at length with few pauses" (p. 93). He explains that this means not having to pause to think of what to say and cites radio presenters and sports commentators as examples of people who require this sort of fluency to carry out their jobs. Research into L1 oral fluency has found that there are certain fluency parameters within which native speakers tend to perform (e.g. Goldman-Eisler, 1968). In general, native speakers produce speech at a rate of 130-200 words per minute and they pause for about a third of the time they spend speaking (Hilton, 2008). Pausing is clearly a normal part of speaking fluently, then, and it performs a number of functions (see also Chafe, 1980).

The second kind of fluency that Fillmore mentions is "the ability to talk in coherent, reasoned, and 'semantically dense' sentences" (p.93). He explains that people who are fluent in this sense "tend not to fill discourse with lots of semantically empty material and mentions Noam Chomsky as an example of a person who displays this sort of fluency. The actor and presenter Stephen Fry might be another example. The third kind of fluency is "the ability to have appropriate things to say in a wide range of contexts" (p.93) This really refers to adaptability to different situations and audiences, knowing what to say and how to speak in different places with different people. The final type of fluency identified by Fillmore is "the ability some people have to be creative and imaginative in their language use, to express their ideas in novel ways, to pun, to make up jokes, to attend to the sound independently of the sense, to vary styles, to create and build on metaphors and so on" (p. 93). In other words, fluency in L1 can be defined along *temporal* lines (i.e. how *fast* can a person speak; to what extent do they *pause*?) and also along more aesthetic lines (i.e. Fillmore's second, third and fourth kinds of fluency).

2.3.2 Fluency in L2

Having explored what it may mean to be more or less fluent in one's first language, I want to turn now to what it means to be fluent in an L2. The following section will look at how fluency in L2 can be interpreted in both a broad and narrow sense and that even with a narrow interpretation, there still remain different ways of conceptualising fluency. I present Segalowitz' (2010; 2016) conceptualisation of L2 fluency in terms of cognitive, perceived and utterance fluency as a useful and comprehensive base for discussions about L2 fluency.

2.3.3 Broad and narrow definitions of L2 fluency

Most research papers and chapters on the topic of L2 fluency explain that 'fluency' can be interpreted in two different ways. Lennon (1990) suggests that these be termed a 'broad' and a 'narrow' definition. The 'broad' definition, as outlined in Lennon (1990) refers to the (English language) lay persons' understanding of fluency as meaning high proficiency or ability (Chambers, 1997). In other words, at least in the English language, it is common to respond to a question about our ability in an L2 with something like: "oh, I get by, but I'm nowhere near *fluent*" (Derwing, 2017). Here, the assumption is that the person has not yet *mastered* the language. On the other hand, the same term, 'fluency', is also used by language specialists (e.g. teachers, researchers) to refer only to certain qualities of L2 speech such as "the degree to which speech flows, and to what extent that flow is interrupted by pauses, hesitations, false starts and so on" (Derwing, 2017, p. 246). This, in Lennon's terms, would be the 'narrow' definition. Narrow-definition fluency is often discussed alongside other isolated aspects of speech performance such as the accuracy with which one speaks and the complexity of the language used.

However, Tavakoli and Hunter (2017) have suggested that definitions of fluency are not necessarily dichotomous (i.e. either broad or narrow) and instead inhabit space along a continuum from *very* broad (i.e. fluency as mastery of L2) to the *very* narrow (i.e. fluency as rate of speech, for example) (see Figure 3). This observation resulted from their research findings, in which teachers of English as a second language (ESL) responded to a questionnaire which elicited their views on L2 fluency and related classroom practice (i.e. how they approached fluency development in their classrooms). A number of questions were aimed at establishing how these teachers actually *defined* fluency (i.e. if they have a broad or narrow definition in mind). The majority (though not all) of the teachers seemed to interpret 'fluency' as general proficiency or as overall speaking skill. The researchers argued that differences between language researchers' and language teachers' understanding of fluency could mean that communication between these two groups is more difficult.

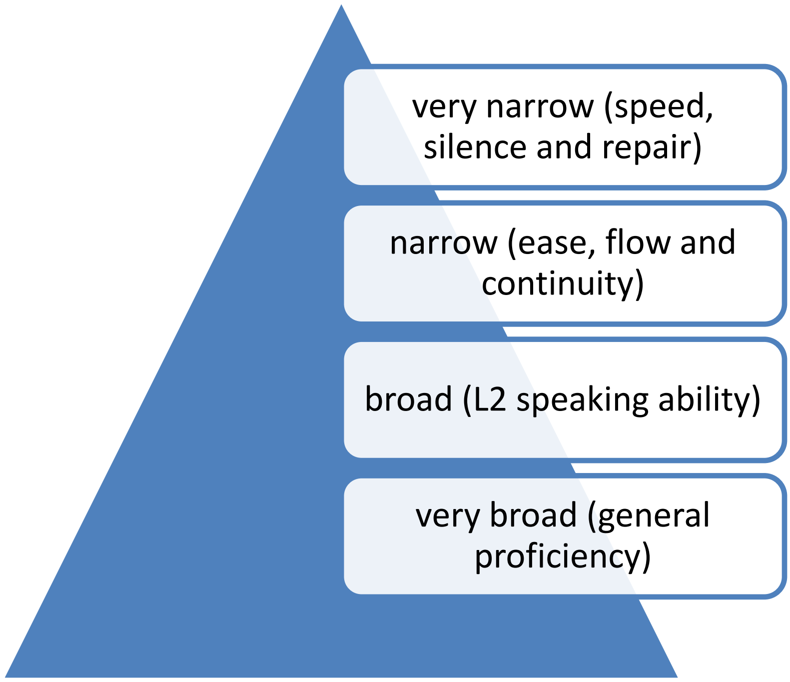


Figure 3: A framework for definitions of fluency (from Tavakoli & Hunter, 2017)

In the SLA literature, a number of qualitative definitions of fluency can be found (see Table 1). Ultimately they present fluency as representative of (a combination of) speed, smoothness and effortlessness and therefore a much *narrower* interpretation than either the lay definition or the teachers' definition (Tavakoli & Hunter, 2017).

As the focus of the current study is fluency in the context of English language teaching, it is worthwhile to also consider qualitative definitions of fluency which can be found in the CEFR

(Common European Framework of Reference for Languages) (table 2). Here we see a similar set of oral characteristics defined. In order to be fluent at the highest level of the CEFR (C2), speakers must speak “smoothly” and “spontaneously” and with a “natural flow” whereas the speech of lower proficiency learners is characterised by “short utterances” and “much pausing”.

Table 1: Definitions of fluency in SLA

	Definition
Derwing & Munro, 2013	Fluency refers to listeners’ perceptions of the flow of the speaker’s language output,
Gatbonton & Segalowitz, 2005	The smooth and rapid production of utterances, without undue hesitations and pauses, that results from constant use and repetitive practice.
Lennon (2000, p. 26)	The rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing
Lennon, (1990, p. 391)	An impression on the listener’s part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently
Pawley & Syder (1983, p. 191)	The native speaker’s ability to produce fluent stretches of discourse
Rehbein (1987, p. 104)	Fluency means that the activities of planning and uttering can be executed nearly simultaneously by the speaker of the language
Sajavaara (1987, p. 62)	The communicative acceptability of the speech act, or ‘communicative fit’
Schmidt, (1992, p. 358).	[Fluent speech is] automatic, not requiring much attention or effort

Segalowitz, 2010	How efficiently the speaker is able to mobilize and temporally integrate, in a nearly simultaneous way, the underlying processes of planning and assembling an utterance in order to perform a communicatively acceptable speech act
Skehan, 1996	[Fluency] consists of the capacity to mobilize one's linguistic resources in the service of real-time communication, i.e., to produce (and comprehend) speech at relatively normal rates, approaching (but not necessarily identical to) one's own native-language rates

Table 2: Fluency as described in CEFR

Band	Fluency descriptor
A1	"Can manage very **short…utterances**, with **much pausing** to search for expressions, to articulate less familiar words and to **repair** communication."
A2	"Can make him/herself understood in very **short utterances**, even though **pauses, false starts and reformulation** are very evident"
B1	"Can **keep going** comprehensibly, even though **pausing** for grammatical and lexical **repair** is very evident."
B2	"Can produce **stretches of language** with fairly **even tempo**, although he/she can be **hesitant** as he/she searches for patterns and expressions"
C1	"Can express him/herself **fluently** and **spontaneously**, almost **effortlessly**. Only a conceptually difficult subject can hinder a **natural, smooth flow** of language."
C2	"Can express him/herself spontaneously…with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it."

In terms of the descriptions of 'fluency' in Table 1, we can see that some have focused on the quality of the speech which is produced (e.g. Lennon, 2000), others focus on the way that the speech is interpreted by another person (e.g. Derwing & Munro, 2013; Sajavaara, 1987) and still others focus on underlying cognitive abilities (Segalowitz, 2010; Skehan, 1996). On the other hand, the CEFR descriptors Table 2) seem to straddle the three different perspectives, sometimes focusing on speech quality (e.g. "short utterances"), other times looking at listener subjective response (e.g. "backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it") and sometimes alluding to underlying cognitive processing (e.g. "effortlessly"). The perspective that one takes to defining fluency will depend on a number of factors including the means that one has available to capture it. Even within a *narrow* interpretation of fluency, then, there are very different ways of conceptualising fluency.

2.3.4 Cognitive, Perceived and Utterance fluency

Although L2 researchers generally work within a narrow definition of fluency, as I explained above, there is still room for argument about what fluency actually *is* (as becomes apparent when looking at the CEFR descriptors). For example, one question is whether fluency is something which is a purely subjective phenomena 'residing in the ear and mind of the listener' (Freed, 2000) or whether it is related to underlying cognitive processes which can be observed and measured objectively. Segalowitz (2010; 2016) suggests that there are *three* different ways of thinking about L2 fluency. He suggests that the example sentence: *Noriko speaks Inuktitut quite fluently for a Japanese* can be said to have "a three-way ambiguity in meaning" (Segalowitz, 2010, p. 47). It could mean that:

1) Noriko "has the ability to mobilize her cognitive system for speaking Inuktitut in a highly effective and efficient manner, similar to what happens with native speakers of Inuktitut"
2) "the utterances Noriko produces in Inuktitut have, objectively speaking, certain characteristics of speech flow in terms of rate, pauses, hesitations, and repair features that render the speech quite fluid"
3) Or "that people who hear Noriko speak Inuktitut will infer, based on their perceptions of her speech, that she has highly efficient cognitive skills for speaking the language, i.e., that she sounds like she is a 'fluent' speaker"

These three different ways of thinking about L2 fluency are presented in Table 3 and Table 4. In Table 4, on the left side of the diagram, in the domain of *cognitive* fluency, there are processes of utterance planning and assembly which, when integrated and assembled produce a communicatively acceptable utterance (the right hand side of the diagram). Here, we can see that this results in the production of speech and therefore measurable features of oral production. This is the domain of *utterance* fluency and is concerned with overt speech and its particular features which reveal something about a speaker's cognitive fluency. At the bottom of the diagram, we see yet another domain. This is the domain of *perceived* fluency, and refers to the reaction that listeners make about a speaker's cognitive fluency based on their overt speech.

Table 3: The three domains of fluency (Segalowitz, 2010)

Cognitive fluency	Refers to the efficacy of the speaker's underlying processes responsible for fluency-relevant features of utterances
Utterance fluency	Refers to the oral features of utterances that reflect the operation of underlying cognitive processes
Perceived fluency	Refers to the inferences that listeners make about a speaker's cognitive fluency based on perception of the utterance fluency features of the speaker's speech output

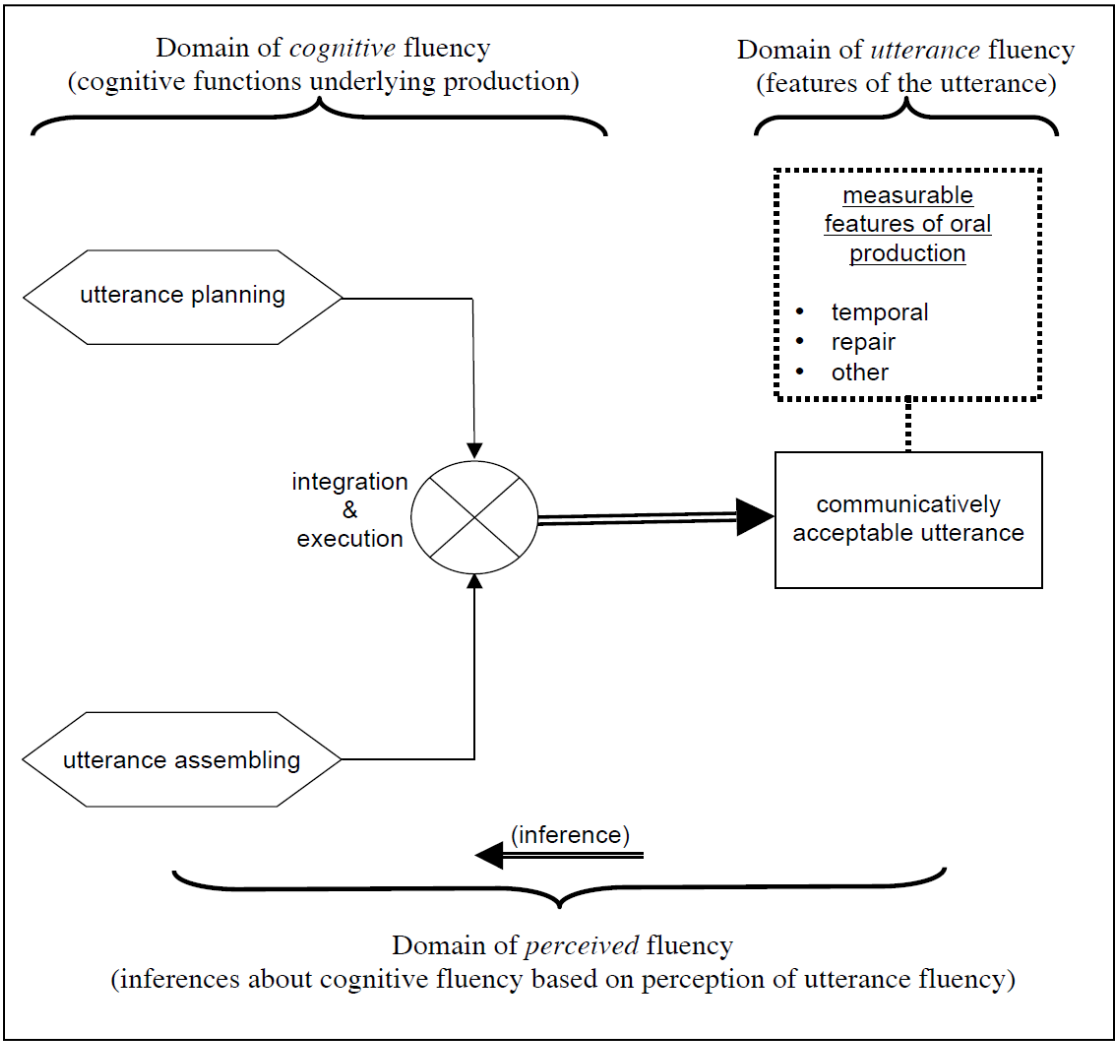


Figure 4: Three domains of fluency. From Segalowitz (2010, p. 50)

The current study will embrace this 'triple perspective' of fluency, with a cognitive 'base'.

Let's turn now to look at each of these domains of fluency in a little more detail.

2.3.4.1 Cognitive fluency

To give a little more detail on each of these domains of fluency, first of all, *cognitive fluency* reflects the speaker's ability to co-ordinate many different processes which underlie speech production (discussed later in this chapter) in real-time in order to produce speech. A fluent speaker will carry out these activities quickly and efficiently, maintaining the natural flow of speech. The kind of activities involved are lexical searches and phonological scoring and the efficacy with which they are carried out is cognitive fluency.

It is generally believed that cognitive fluency (and by extension perceived and utterance fluency) is strongly related to the extent to which lower-order cognitive processes which underlie speech production are automatic (Kormos, 2006; Segalowitz, 2010; Hilton, 2014). And yet, automaticity is, in itself a multi-layered concept with a number of theories which attempt to explain it (DeKeyser, 2017; Kormos, 2006; Segalowitz, 2010). In what follows, I explain what is meant by the term 'automaticity' and theories of how automaticity in L2 speech production comes about.

Automaticity

Automaticity has been explored extensively, in terms of automaticity in general skill acquisition (e.g. Anderson & Lebiere, 1998; Levelt, 1989; Logan, 1988; Schneider & Chein, 2003) and also more specifically to SLA (e.g. DeKeyser, 2001; Hulstijn, 2001; Segalowitz & Hulstijn, 2005). Across these different explorations, a range of interpretations of the term 'automaticity' exists. DeKeyser (2017) explains that "(h)ighly automatized knowledge is usually characterized as unintentional, uncontrollable, unconscious, efficient, and fast" and that automaticity in language skill is "graded". In other words, there is always room for improvement, even in an L1. Within this conception of automaticity, there are different layers. Segalowitz (2010) explains that central to many definitions of automaticity is the notion of processing *speed.* That is, that when a particular process becomes automatic, is somehow carried out *faster* (ballistic or *unstoppable* processing). He also notes, however, that speed of processing "cannot be the sole justification for calling it automatic…Claiming some process has become automatic implies claiming that it has acquired some properties beyond simply becoming faster" (Segalowitz, 2010, p. 79). He explains this as processing *stability.* In other words, that the particular process required to carry out the skill in question has become more efficient and restructured. But how exactly does L2 speech become more speedy and stable (i.e. automatic)?

Routes to automaticity

Kormos (2006) and more recently Derwing (2017) explain that theories of how L2 speech becomes automatic can be grouped under two main labels. The first is what Kormos (2006) refers to as "rule-based" theories of automaticity and the second are "item-based" theories of automaticity.

Rule-based theories often draw on general skill acquisition theory and suggest that the way in which speech processing becomes automatic is when declarative knowledge is converted into procedural knowledge and then eventually it becomes 'habituated'. As DeKeyser (2017) explains "declarative knowledge allows learners to engage with the target behaviour [language form] ..and by drawing on this declarative knowledge repeatedly to engage in this behaviour repeatedly, forming procedural knowledge, establishing a habit after some repetition, and then gradually automatizing this habit" (DeKeyser, 2017). Meanwhile, item-based theories hold that language is memorised as un-analysed chunks.

Rule-based routes to automaticity (proceduralisation)

In order to understand what a rule-based route to automaticity might be, it is necessary to first explain a little bit about 'knowledge' in language learning terms. In SLA it is common to talk about knowledge of the L2 in terms of *declarative* and *procedural* knowledge. Declarative knowledge is sometimes referred to as "knowledge *that*" while procedural knowledge is classed as "knowledge *how*" (DeKeyser, 2017). In many instances of second language learning, knowledge is initially *about* the language and its rules and grammar, in other words, *declarative.* This is true even given the increase in popularity of alternative teaching methods which foreground communication (DeKeyser, 2017).

At the beginning stages of learning an L2, drawing on declarative knowledge about the L2 during online communication can mean that interactions are characterised by dysfluencies. However, DeKeyser (2017) explains:

> as a result of practice [learners] become better at putting their knowledge to use, using it more correctly, more easily, more frequently in a variety of contexts. Sometimes this process is called automatization in a broad sense, but more technically what happens is first developing procedural knowledge and then automatizing it

If we think of a particular skill that we may possess, like being able to tie a shoelace, it is easy to remember that initially this was a difficult task which required a great deal of conscious effort. At one time, we all tied our shoelaces in a way that was effortful and which relied on a set of rules which were carried out carefully and sequentially. It is unlikely that any of us still need a rhyme like 'bunny ears' to tie our shoelaces. In fact, it is unlikely that we think consciously about tying our shoelaces at all. We can tie our shoelaces while doing other things such as watching TV or having a conversation. This is because, through repeated practice, declarative knowledge about tying shoelaces became procedural before being sufficiently habituated as an automatic skill.

It is believed that a similar process is in play in SLA and that through exposure to TL elements (input repetition) and huge amounts of practice (output repetition), cognitive processing skills responsible for language production become proceduralised and then automatised, resulting in speech which is more fluent (Segalowitz & Hulstijn, 2005; Bybee, 2008).

Kormos (2006) gives the example of a language learner who starts out being taught that there is a distinction between the articles “a” and “an” in the sense that one is used before words starting with a consonant and the other before words starting with a vowel. With practice she begins perceive this knowledge as a procedural rule. With further practice, the learner can apply the rule automatically without having to pay conscious effort.

An influential explanation for the process in which rules become automatic is provided by Anderson’s (1983) ACT theory and his revised (1995) ACT-R theory. He suggests that there are five main learning mechanisms at play in the development of automatic skill performance: *composition, proceduralization, generalization, discrimination,* and *strengthening.* Composition and proceduralisation involve the collapsing of smaller production rules into larger ones, For example, whilst writing this section of the thesis, I tried to apply the ‘bunny ears’ approach to shoelace-tying and found it very difficult. This is because these smaller rules (i.e. “make a loop with the right-hand shoelace; make another loop with the left-hand shoelace…”) have been removed from memory and replaced with a larger rule (i.e. “tie shoelaces”). This larger rule is something which can now be applied in one go. Another example which is commonly given is that of remembering telephone numbers. Because working memory is limited in how many numbers it can hold at any one time (somewhere between five and seven is an estimate), when trying to remember phone numbers, people chunk the numbers into groups of two or three numbers in order to be able to hold on to it for longer. The interesting thing that happens sometimes is that if a person tries to say the numbers individually, they can no longer remember the number, even if it is a number they use very frequently. As Kormos (2006) explains, “once a production has become automatic, the initial declarative knowledge underlying it is often not retrievable anymore” (Kormos, 2006, p.41). This also explains why, when would-be English language teachers embark on a training course to learn how to teach what is often their native language to others, they must undergo a significant amount of linguistic training in order to appreciate the grammatical structure of the language. This might also be the case for advanced L2 speakers who once learned rules about the language explicitly but no longer remember them.

The other three processes which are identified as supportive of automaticity in ACT-R theory are generalisation, discrimination and strengthening. These three processes involve being able to apply new production rules wherever they are suitable, and *only* when they are suitable, and that better (more efficient) rules are selected and weaker ones are abandoned.

Item-based routes to automaticity

However, it is likely that the proceduralisation and subsequent automatization of rules is not the only way that speech can be automatic (Raupach, 1987; Robinson & Ha, 1993). Kormos (2006) points out that "phrases and clauses first assembled with the help of syntactic and phonological rules might later be stored as one unit in memory and retrieved as a whole" (Kormos, 2006, p. 156) and also that whole chunks of language may actually be learned and regurgitated as a whole with declarative analysis coming at some later point or not at all.

Overall, cognitive fluency can be seen as dependent on the extent to which speech production is *automatic*[3]. I have also explained, however that there are a number of 'routes' to automaticity in L2 speech production, one being the transformation of declarative knowledge into procedural knowledge which is then automatized and the other the storage of certain linguistic chunks in memory which can be retrieved as an (unanalysed) whole.

2.3.4.2 Perceived Fluency

Having explored what is meant by *cognitive fluency,* the second piece in the jigsaw is *perceived fluency*. It refers to inferences about fluency that listeners make when they hear someone speak. The perceived fluency that we are discussing here are inferences made about a speaker's *cognitive* fluency based on the speed and efficiency of speech production (i.e. their utterance fluency). There is some degree of ambiguity in the way that 'perceived fluency' is

[3] Segalowitz (2010) also considers *attention flexibility* linked to grammatical aspects of the L2 to play a part in cognitive fluency alongside lexical access processing speed and stability (i.e. automaticity). For reasons of clarity and space, I will not expand on this aspect of cognitive fluency here.

used and understood in SLA research. On one hand it can have quite a broad interpretation, as being related to the judgements people make of a person's fluency based on their own idiosyncratic definition of what fluency is (this is the definition presented recently in Derwing, 2017, for example). As we saw in the first section of this chapter, however, human beings are likely to have a range of criteria in mind when they judge a person's 'fluency' which may or may not overlap with an L2 research definition (Brown, Iwashita, & McNamara, 2008). On the other hand, 'perceived fluency' as it is presented in Segalowitz (2010) is much narrower, insofar as it refers only to perceptions of fluency which are based on cognitive fluency.

Perceiving fluency (and measuring utterance fluency) is further complicated by the fact that L2 users can 'compensate' for low levels of cognitive fluency by engaging in communicative strategies which mask deficits in this area (Dörnyei & Scott, 1997). Communication strategies are also used in the L1, of course, although they may be more prevalent in L2 due to increased difficulties in online processing. There has only been limited research into the relationship between the use of communicative strategies and perceived fluency but there are clearly enormous implications for SLA research and language testing and policy.

As Dekeyser (2017) explains:

> (S)killed L2 speakers will often be able to fill the gaps in their procedural knowledge by drawing very efficiently on declarative knowledge without uttering anything that can be detected as nonnative or even nonfluent. High levels of fluency leave enough mental resources to plan ahead, detect possible sources of nonfluency or nonaccuracy, and avoid them by searching efficiently for alternative procedures, including procedures that call on small chunks of declarative knowledge.

2.3.4.3 Utterance fluency

The final domain in Segalowitz' (2010) model is *utterance fluency.* As was the case with perceived fluency, the phrase is often used in two senses. On one hand it is used to refer to any and all features of speech which *might* be related to cognitive and/or 'subjective' fluency. Segalowitz (2010) however, intends a much narrower interpretation referring only to those features of speech which are shown to be directly related to cognitive fluency. On the question of whether or not any specific utterance fluency features have been identified which anre directly linked to cognitive fluency, he writes that "no evidence has been presented yet of a direct association between measures of cognitive fluency and measures of utterance fluency” (2010, p. 104)

Segalowitz (2010) explains that L2 speakers with low levels of cognitive fluency (i.e. automaticity in speech production processing) find themselves in communicative situations where they are aware that their speech production processes cannot keep up with the time-pressure demands of the interaction. In these situations, learners need to stall for time. There are many ways that a speaker can ‘stall for time’ when speaking. Dörnyei and Kormos (1998) identify a number of problem-solving mechanisms that are specifically related to the time-pressure which is experienced by L2 speakers. Firstly they can simplify the message or abandon it completely, they can also replace parts of the message with information that is more quickly encoded (Dörnyei & Kormos, 1998; Dörnyei & Scott, 1997, Kormos, 2006, p. 142; Segalowitz, 2010, p. 41). The other option that L2 speakers have is to make use of *stalling mechanisms* (Dörnyei & Kormos, 1998; Kormos, 2006) (see Table 4).

Table 4: Stalling mechanisms (From Dörnyei & Kormos, 1998)

Stalling mechanism	Description
Unfilled pauses	Remaining silent while thinking
Umming and erring (filled pauses)	Using nonlexicalised filled pauses (er, uh, mmm)
Sound lengthening (drawling)	Lengthening a sound in hesitation
Fillers	Using filling words or gambits to fill pauses, to stall, and to gain time in order to keep the communication channel open and maintain discourse at times of difficulty

Self-repetitions	Repeating a word or a string of words immediately after they were said
Other-repetition	Repeating something interlocutor said to gain time

Dörnyei and Kormos' (1998) overview of stalling mechanisms is a useful departure point from which to discuss utterance fluency. Broadly speaking, it is the investigation of these sort of 'stalling mechanisms' or 'hesitation phenomena' that has formed the basis of the identification of reliable indicators of utterance fluency.

2.4 Measuring fluency

If fluency is a key component (or maybe even *the* key component) of second language skill and proficiency, then it is likely that researchers, policy-makers, examination-designers, teachers and learners will all be keen to know how fluency can be tracked and measured.

The key question over the past few decades has been: What are the specific features of speech that tell us something about cognitive fluency? Unsurprisingly then, this is an area of research that has enjoyed substantial interest over time (e.g. Bosker et al.., 2013; N.H. de Jong, 2016; N.H. de Jong et al., 2012; 2013;; Derwing et al.., 2004; Derwing et al.., 2009; Freed, 2000; Freed, Segalowitz, & Dewey, 2004; Ginther, Dimova, & Yang, 2010; Iwashita et al.., 2008; Kormos & Dénes, 2004; Préfontaine, 2013; Riggenbach, 1991; Rossiter, 2009; Towell, Hawkins, & Bazergui, 1996). Slowly, a number of features of fluent speech are being identified which seem to be good candidates for utterance fluency measures.

This section will firstly consider how fluency fits into a bigger framework for performance assessments, namely, complexity, accuracy and fluency (CAF). Next, I will look at how fluency can be measured in terms of the *speed* of the speech, the extent to which there is *breakdown* in communication, and the evidence of *repair.* I will also explain how potential utterance fluency variables have emerged from research studies which has compared language performance before and after a period in the TL community, language in L1 and L2 or else has looked at how temporal aspects of speech correlate with rater judgements of fluency. The section will conclude with an up-to-date overview of utterance fluency measurement and the measures that emerge as the most robust and reliable.

2.4.1 Measuring performance in L2 (CAF)

Fluency is commonly seen as one of three core components in L2 performance, alongside accuracy and complexity. The origins for this way of thinking about performance triad emerged from a need in the field of SLA, language assessment and, to some extent, language teaching needed to be more specific about what it means to speak (or write, listen and read) 'well' in a second language. To focus just on L2 *speaking* for a moment, Hakuta (1975) and Larsen-Freeman (1976) were among the first to call for ways in which L2 performance could be reliably and objectively measured; "the construction of an SLA index of development" (Larsen-Freeman, 2009, p. 580). Drawing on research on L1 development (e.g. Brown, 1973; Hunt, 1965; 1970), they wanted to find ways to measure objectively the development of L2 language learners. At the same time, in the world of second language teaching, a similar distinction was being made between learner language which was accurate and that which was fluent (see Brumfit & Bennett, 1979).

Skehan (1996; 1998) introduced an L2 proficiency model which brought everything together under a triadic framework known as CAF. CAF stands for *complexity, accuracy, fluency* and the definitions suggested by Skehan are still used in language research, language testing and language teaching (e.g. Ellis, 2003) and they have since been found to be both theoretically and empirically valid as aspects of L2 performance (Ellis & Larsen-Freeman, 2006; Norris & Ortega, 2003; 2006; 2009; Skehan, 2003). Table 5 provides working definitions of the three constructs as they appear in Ellis (2009)

Table 5: Definitions of complexity, accuracy and fluency (from Ellis, 2009, p. 475)

Complexity	"The capacity to use more advanced language, with the possibility that such language may not be controlled so effectively"
Accuracy	"The ability to avoid error in performance, possibly reflecting higher levels of control in the language as well as a conservative orientation, that is, avoidance of challenging structures that might provoke error"

Fluency	"the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems"

These three constructs, as a way of discussing second language performance have been popular in studies which have tried to answer research questions about the impact of a wide range of variables on performance (e.g. Foster & Skehan, 1996; Robinson, 2011). More recently, researchers have even begun to investigate the constructs themselves in order to understand more about cognitive processes underlying L2 performance and development (e.g. Bosker et al., 2012; N.H. de Jong et al., 2012; DeKeyser, 1998)

CAF has also been used as a way to gauge L2 *acquisition* by investigating whether performance gains 'transfer' to a different task, although Larsen-Freeman (2009) asks whether this is the best approach, given that it is possible that transfer happens in a way that should be measured differently. CAF indicators have come to explain oral performance different proficiency levels in ACTEFL and CEFR, two important second language frameworks and these benchmarks form the basis of materials for language teaching and high-stakes L2 examinations (e.g. IELTS, TOEFL).

Despite the popularity of these constructs for assessing L2 performance, the three constructs are not uncontroversial and a number of researchers have expressed concern about their ubiquity (e.g. Housen & Kuiken, 2009; Lambert & Kormos, 2014; Larsen-Freeman, 2009; Pallotti, 2009). Housen and Kuiken (2009) explain that issues with the CAF framework include the fact that the dimensions are operationally defined in different ways in different studies, that there is a question mark over how well CAF can capture L2 *development* as opposed to performance. Others have also pointed out the fact that these aspects of performance are influenced by so many additional factors as well as by each other that it makes it difficult to identify individual effects. Housen, Kuiken and Vedder (2012) point out that the main concern with CAF is that studies do not define constructs in sufficient detail.

Although the focus of the current research is fluency, it is important to acknowledge the interrelated nature of CAF as well as the importance of each of these components to contributing to overall performance (Tavakoli & Skehan, 2005). The study of complexity and accuracy in L2 performance have rich and substantial bodies of literature (see e.g. Norris &

Ortega, 2009; Pallotti, 2009; Foster & Wigglesworth, 2016; Lambert & Kormos, 2014) however, given the aims of the current research, the review of the literature here and in the chapters that follow is therefore focussed primarily on fluency.

2.4.2 Fluency as speed, breakdown and repair

While L2 performance is seen as a complex and multidimensional construct (Housen et al., 2012), fluency itself is also believed to comprise (at least) three sub-constructs. Indeed, the notion that fluency might be measured as "a single unitary concept" (Koponen & Riggenbach, 2000:17) has long been abandoned. Skehan (2003) suggested that L2 fluency is related to *speed, breakdown* and *repair* aspects of speaking (see Table 6). This was later repackaged as two categories of *temporal* fluency and *repair* fluency (Tavakoli & Skehan, 2005). They suggested that certain characteristics of L2 speech reflected temporal aspects of speech (e.g. speed of speaking; pausing), others reflected the tendency to repair one's speech online (e.g. a tendency to reformulate one's message).

Table 6: Aspects of fluency and typical measures

	Typically measured by..
Speed	Speech and articulation rates
Breakdown	Frequency and length of pauses, amount of speaking time
Repair	Frequency of reformulations, self-corrections, repetitions

Skehan's (1998) original categories of speed, breakdown and repair are often used to group utterance fluency indicators. *Theoretically speaking* there is a link between cognitive fluency (automaticity) and key features of the speech which is produced (e.g. speed, breakdown, repair). To give an example, increased automaticity in speech processing *should* result in increased speed of speech. And yet, as I have explained, learners are able to compensate for lower levels of cognitive fluency by employing a range of communicative strategies. Very few empirical studies have been carried out which attempt to link cognitive processing (in terms of speed, stability and flexibility) to oral performance (although see Segalowitz & Freed, 2004 and, more recently, NH de Jong et al., 2015). Furthermore, there are a wide range of internal and external factors which may exert influence on a speaker's speech at any given moment. It is important to bear in mind, then, that the use of 'utterance fluency' as a way to tap cognitive processing is very much in its infancy. There is much that we do not know about this relationship. It is to this topic that we now turn.

2.4.3 Identifying reliable L2 utterance fluency measures

Many researchers have sought to identify reliable indicators of oral fluency and the approach taken by researchers has varied considerably. The research suggests that certain temporal measures consistently emerge as most reliable. Some have sought to identify utterance fluency markers by comparing L2 speech before and after an intervention which was supposed to impact on fluency (e.g. Hansen, Gardner & Pollard, 1998; Lennon, 1990; Towell et al., 1996;). Others have attempted to identify reliable measures of fluency based on differences between performance of a task in L1 and L2 (e.g. Deschamps, 1980; N.H. De Jong, 2016; Raupach 1980; Tavakoli, 2011). These studies work on the assumption that differences in utterance features between L1 and L2 may be indicative of similar differences in cognitive fluency. Other researchers have attempted to ascertain the characteristics of speech which correlate most robustly with perceived fluency ratings by native speakers of the TL (Ejzenberg, 1992; Freed, 1995; Lennon, 1990b; Riggenbach, 1991; Wennerstrom, 2000), non-native (expert) speakers of the TL or teachers and language professionals (Cucchiarini, Strik & Boves, 2002; Freed, 1995; Riggenbach, 1991).

These studies taken together point to the importance of both speed and breakdown phenomena as gauges of fluency. However, as Kormos and Dénes (2004) point out there are certain issues with the reliability of these findings due to the way that pauses were calculated and small sample sizes. I will therefore look in detail at two influential and robust studies into perceptions of the fluency of L2 speech: Derwing et al. (2004) and Kormos and Dénes (2004).

Derwing et al. (2004) recorded 20 high-beginner Mandarin-speaking ESL learners performing a narrative, personal monologue and conversation task. These speech samples were rated for fluency on a 9-point Likert scale by 28 English native speakers who were on an English language teaching course at university. The listeners were told to pay attention to “temporal variables such as filled and unfilled pauses, false starts, and self-repetitions” (Derwing et al., 2004, p. 664) and that the researchers were interested in their “perceptions of fluency in terms of the flow and smoothness of speech rather than in terms of overall proficiency” (2004, p. 664). They found that length of run (average number of syllables between pauses) and, in particular, speech rate (pruned syllables/min) correlated highly with rater judgements. Their explanation for this was that pruned syllables provide a “composite measure” of fluency which incorporates repair. They found that self-repetitions correlated the least highly of all the measures, leading them to hypothesise that “self-repetition…depending on how it is used by the speaker, could reflect a way to buy time that actually gives an impression of fluency”.

Kormos and Dénes’ (2004) study investigated the relationship between temporal features of speech and native and non-native rater judgements of fluency. In their study, 16 learners from 2 proficiency groups performed a narrative task which was audio-recorded. The speech samples were judged for fluency by two groups of raters: a native speaking rater group (n=3) and a non-native (Hungarian) speaking rater group (n=3). An impressively large range of temporal and linguistic features were considered as potential predictors of fluency ratings in this study. Variables selected for analysis were: 1) speech rate 2) articulation rate 3) phonation time ratio 4) mean length of run 5) number of silent pauses 6) mean length of pauses 7) number of filled pauses 8) number of disfluencies 9) pace 10) space 11) total number of words 12) D-formula for lexical diversity 13) accuracy.

The best predictors of fluency scores (both individual scores and grouped scores) in this study were found to be: speech rate (syllables per minute), mean length of run (average number of syllables between pauses), phonation time ratio (percentage of speaking time) and pace (stressed words per minute). While length of pause seemed to be important for some raters, for others it was not. Similarly, accuracy and lexical diversity was an important predictor only for some of the judges. Through the comparison of a number of speakers’ performances, the researchers also acknowledge that on some measures participants performed very differently from each other and received surprising scores. Despite this, the researchers are confident in the ability of these four variables to predict rater judgements of fluency. An important distinction between this study and that of Derwing et al. (2004) is that the raters in this study were not told to focus on any particular aspect of the speech. They were therefore not guided towards a particular interpretation of fluency (i.e. either broad or narrow). When these studies are taken together, then, there is support for the claim that (narrow definition) fluency indicators might be the same as those that indicate broader definition fluency (i.e. nativelikeness/proficiency). However, in Kormos and Dénes' study, many of the raters were actually language teachers and Applied Linguistics experts, it is therefore possible that they instinctively took a narrower definition despite not being guided to do so. Indeed, from the qualitative data that the researchers collected, it seems that these teachers did focus heavily on temporal features of speech.

In both of the studies above, pausing did not appear to be particularly reflective of the perceived fluency scores. However, others have suggested otherwise. Rossiter (2009), for example, found that pause phenomena were the best predictors of perceived fluency (see also Hilton, 2014). More recently, there has been considerable research which has attempted to identify similarities and differences in the way in which pauses are used in L1 and L2 speech. This is in an attempt to isolate pauses which are related to L2 cognitive fluency, rather than for other reasons. Tavakoli (2010), for example, showed that L2 speakers generally pause more often and for longer than L1 speakers. More recently de Jong (2016) investigated the difference between L1 and L2 pause distribution and found that L2 speakers are more likely to pause within AS units (De Jong, 2016; Skehan & Foster, 2007), clauses (Tavakoli, 2011) and constituents (Riazantseva, 2001).

Fulcher (1996) found that dysfluent speakers were more likely to pause because they have problems relating to grammar and lexical retrieval whereas higher fluency speakers pause because they have something conceptually difficult to say. In a similar vein, De Jong (2016) and Bosker et al. (2014) provide evidence that L1 and L2 speakers differ in terms of what their pauses *precede*, with L1 speakers pausing before less frequent words, presumably due to lexical retrieval speeds. In other words, L1 speakers 'earn' the right to pause by pausing before words with a big pay-off (i.e. low frequency words). L2 speakers, especially those at lower levels of proficiency, on the other hand, pause before words which, though they may be low frequency for the speaker herself, are actually high frequency in general usage. An excellent example of this can be seen in a conversation from the BBC comedy *Mongrels* which is set in the alleyways of London and features urban foxes and stray cats as the lead characters. In one episode, Nelson the fox has his French penfriend come to visit and they have the following conversation:

> Nelson: and how was it getting through the Tunnel?
>
> Christian: It was…er…er how you say… er uh…errrr er er er oh er…. Fine!

This exchange is humorous because the extent to which Christian pauses leads us to think that the word he is searching for will be a sophisticated one.

Overall, then, there is a certain degree of agreement about reliable predictors of speed and breakdown fluency. However, the picture seems less clear when it comes to repair. It may be that different types of repair impact differently on overall perceptions of fluency (Olynyk et al., 1990) or that there is a need to distinguish between error repairs, that is, repairing errors of linguistic form; and appropriateness repairs (i.e. presenting a new or rephrased message) (Kormos, 1999; Levelt, 1983).

Both Lennon (1990b) and Freed (1995) suggest that repair is actually evident in more developed language and that "part of fluency development in the advanced learner may involve *increased* ability to reformulate, monitor and self-correct production on-line (Lennon, 1990b, p. 412). Similarly, Freed (1995, p. 142) writes:

> There is a tendency for students who have been abroad , especially those whose speech is more advanced, to attempt linguistic expressions which they sometimes find don't work: they reformulate their speech producing more false starts than is evidenced in the speech of those who have never been abroad.

Very recently Tavakoli, Nakatsuhara and Hunter (2017) showed that repair fluency did not develop alongside overall proficiency in a linear fashion. Instead, it seems that learners only begin to engage in repair behaviours once they have reached a particular level of proficiency and that once an advanced level of proficiency is achieved, repair behaviour begins to decrease.

I have shown, then, how a substantial body of research has attempted to identify the best oral features that might reliably measure fluency. A range of approaches have been adopted by researchers including longitudinal studies, L1 versus L2 studies, proficiency and rating studies. The measures that consistently emerge as most reliable are: 1) speech rate 2) length of run, possibly because they incorporate multiple areas of fluency (i.e. speed, breakdown and repair). However, I would argue that if the nature of *cognitive* fluency is going to be better understood then more nuanced measures of fluency (e.g. measures related solely to speed) are needed.

2.4.4 Issues in L2 fluency measurement

There are a number of pertinent issues in L2 fluency measurement, which are explored below. Firstly, it is not yet entirely clear how fluency can be measured in interaction. Sato (2014) and Tavakoli (2016) have recently explored the question of how fluency differs in monologic and dialogic modes (see also Wolf, 2008; Macarthy, 2010; Ducasse & Brown, 2009). Sato (2014, p. 80) explains that "neither SLA nor LT research has extensively examined the effects of elicitation tasks used to measure oral fluency during interaction despite the fact that the improvement of fluency during oral interaction is a major goal for L2 learners". In his study, Sato compared temporal and rater judgements of fluency in monologic and in interaction. He found that raters' perceptions of fluency in interaction were based on *interaction-specific* fluency such as turn-taking and scaffolding behaviours. In addition, he found that individual fluency was very weakly correlated with interactional fluency, leading the researcher to suggest that fluency in both of these contexts was "fundamentally different" (Sato, 2014, p. 88). Tavakoli (2016) similarly investigated fluency in monologue and dialogue and found that the latter elicited generally more fluent performance in terms of utterance fluency measures. She also explains, however, that certain decisions which a researcher makes when analysing fluency in dialogue (e.g. what to do about between-turn pauses) can have an impact on the outcome. She argues for a more systematic approach to analysing fluency in interaction. An ongoing challenge for fluency researchers is to find ways to conduct research on interactive speech which also allows for individual performance to be measured.

A second issue is that there have been few fluency studies which have studied effects of language instruction on fluency (Michel, 2017). Michel (2017) explains that this may be partly to do with the fact that CAF measures in general are not sensitive enough to what might be very subtle changes in learner language after a short period of pedagogic intervention (longitudinal studies are less common in general). She also notes that, often, the focus of a particular pedagogic intervention is a specific language focus which would not be suitable for CAF analysis. I would add that the classroom environment, in the sense that it is often noisy and unpredictable, provides additional challenges for the CAF researcher who would probably prefer to collect reasonably stable and clear data. The goal for fluency researchers, then, is to find ways to overcome these challenges in order to ensure that this research is carried out.

Another issue and one which is identified by many researchers (e.g. Segalowitz, 2010; Ellis, 2009; Michel, 2017) is that as a result of the search for the "best or better" measure of fluency (described above) an astonishingly large range of metrics are used in fluency studies. Some of these are "of unknown reliability and validity" (Michel, 2017, p. 64). This means that it is often difficult or impossible to compare results among studies. This is further complicated by the fact that even measures with the same name (e.g. 'speech rate') can be calculated in different ways (i.e. *words* per minute versus *syllables* per minute). What is needed is for researchers to make use of empirically-tested measures and provide thick descriptions of the ways they went about calculating them.

A fourth point relates to the growing trend towards the use of computer technology to analyse fluency. Historically, temporal fluency analysis has relied on orthographic transcription and manual counting of syllables. However, the precise temporal nature of utterance fluency measures, combined with a desire to analyse larger data sets, has precipitated the use of computer technology and specialist software such as PRAAT (Boersma & Weenink, 2008) which has been used to automatically detect silence in speech samples (Cuchiarrini 2000; 2002; N. de Jong & Perfetti, 2011) or used in conjunction with a specific computer script to identify syllable nuclei thereby allowing for the computation of speech rate (de Jong & Wempe, 2009). Automatic fluency analysis makes dealing with larger amounts of data feasible, more objective and more precise (Segalowitz, 2010). However, a drawback is that it requires relatively clear, straightforward speech data of the type that is elicited in a language laboratory and not in a working language classroom, where background noise and recording irregularities are likely to feature.

A final point and one that has been already alluded to is the interaction among fluency and other aspects of performance. This interaction has been at the heart of much of the research into Skehan's (1998) 'trade-off hypothesis' which states that complexity, accuracy and fluency compete for attentional resources and that form (accuracy and complexity) is likely to be sacrificed for meaning (fluency). More recently, Larsen-Freeman (2009) and others have suggested that if fluency is studied in isolation then we miss important observations about the interaction among CAF and, indeed, other aspects of performance such as communicative adequacy (Revesz, Ekiert, & Torgerson 2014; Michel, 2017; Pallotti, 2009). This complicates matters because, as Michel (2017, p. 64) explains: "the future calls, on the one hand, for greater standardization and theory-driven use of constructs and metrics and, on the other hand, for the acknowledgement of variability and dynamicity of CAF in L2 use" (Michel, 2017, p. 64).

2.5 Chapter summary

In this chapter, I have discussed the underlying rationale for research into L2 fluency development. I have discussed popular models of L1 and L2 speech production and described how they might provide a framework for thinking about L2 fluency. I have showed how definitions of fluency vary greatly and I explained that I would be adopting Segalowitz' (2010) framework within which there are three dimensions of fluency: *cognitive, perceived, and utterance* fluency. I presented each of these in turn, focussing in particular on the elusive concept of *automaticity* which is key to understanding cognitive fluency. I showed that automaticity in L2 speech can be seen as a result of a movement from declarative knowledge through procedural knowledge and/or the storage of chunks of (unanalysed or previously analysed) language that are now retrieved in a single step from memory.

Next, the chapter discussed another complex area in L2 fluency research, namely, fluency *measurement.* I situated fluency measurement within the broader context of oral performance measurement (CAF) and then explained how detailed analysis of fluency has involved breaking the notion of fluency down even further into dimensions of *speed, breakdown* and *repair.* I then discussed the search for appropriate measures of fluency and presented the measures that have been found to reflect subjective judgements of fluency most reliably. However, in terms of understanding the measures which are most indicative of *cognitive fluency,* there is still no unequivocal measure. Instead it was argued that *speech rate and mean length of run* may have a part to play in understanding cognitive fluency in a 'global' sense but that peeling back the layers of L2 fluency means looking at different aspects of fluency related to speed, pausing and repair. I also explained current approaches to measuring utterance fluency, making use of speech analysis software and automated analysis.

3 - L2 fluency development

3.1 Introduction

While understanding *what fluency is* and *how fluency can be measured* may be important questions for language researchers, a crucial question for many learners and teachers, as well as researchers, is understanding how fluency *develops*. As Derwing (2017, p. 248) explains:

> Applied linguists typically want to identify ways in which learners' fluency can be enhanced through manipulation of tasks in the classroom, the effects of study abroad or other forms of immersion in the L2, the fluency trajectories of learners, and the interrelationships that effect the fluency of utterances produced by L2 speakers.

In this chapter, I explore the myriad sources of influence on L2 fluency. First, we look at how fluency develops, or fails to develop, in line with increasing proficiency. Second, we turn to the influence of the L1 on L2 fluency. Third, I explore the impact of *practice* on fluency development. This involves looking at different kinds of practice including the sort of practice that comes through *immersion* in the TL and also the more narrow, focused sort of practice that might come about in the language classroom. Fourth, I offer Segalowitz' (2010) dynamic systems framework as a useful way to explain the complex interplay between affective factors in L2 fluency development. I conclude by offering a brief summary of the chapter.

3.2 Sources of influence on L2 fluency

There are myriad potential sources of internal and external influence on L2 fluency, for example: the age of exposure to the L2 (Trofimovich & Baker, 2006; Llanes & Munoz, 2013), the speaker's current age (Mora, 2006), interlocutor (Ejzenberg, 2001), L1 speaking style (N.H. de Jong et al., 2012; Derwing & Munro, 2009), linguistic skill (N.H. de Jong et al., 2015), working memory (Kormos & Safar, 2008), vocabulary knowledge (Hilton, 2008), task structure and demands (Ellis & Yuan, 2004; Skehan & Foster, 1996; Tavakoli & Foster, 2011), anxiety (MacIntyre et al., 1998), and syntactic complexity (Mirdamadi & N.H. De Jong, 2015).

Some of these sources of influence might be temporary or transient sources of influence (e.g. interlocutor; task demands), which might impact on fluency *on a particular task* or *during a particular performance* while others may be a more fixed or permanent source of influence (e.g. vocabulary knowledge, age of first exposure to the L1), which might impact performance *on all tasks* and *at all times*. In what follows, I single out a handful of factors which have a particularly salient relationship with L2 fluency: *proficiency, L1 speaking style,* and *practice.*

3.2.1 Proficiency and fluency

Some researchers have attempted to identify a relationship between utterance fluency features and proficiency (Cucchiarini et al., 2000; 2002; Ginther, Dimova & Yang 2010; Iwashita, Brown, McNamara & O'Hagan, 2008; Tavakoli, Nakatsuhara & Hunter, 2017). Understanding the link between proficiency and L2 fluency is important for a number of reasons. Firstly it may provide information on how the L2 system develops over time. Secondly, and from a language testing point of view, it may provide a case for the use of measures of fluency in place of the "laborious and expensive manual rating systems" (Baker-Smemoe, Dewey, Bown & Martinsen 2014, p. 708) which are currently in place.

In some early investigations into the relationship between fluency and proficiency, Higgs and Clifford (1982) and De Jong & Van Ginkel (1992) found that fluency is only a concern at higher levels at proficiency. Cucchiarini et al. (2000; 2002) found strong correlations between fluency and proficiency but, as Ginther et al. (2010) point out, this could mean either that raters pay attention to fluency when they rate speech samples *or* that fluency co-occurs with other aspects of speech production (e.g. vocabulary size or grammatical accuracy) which are indicative of proficiency level.

Ginther et al. (2010) found a strong correlation between proficiency and utterance fluency, especially for the measures of speech rate and mean length of run. While they suggest that these temporal measures *could* be used as a basis for automated assessment of proficiency, they warn that temporal measures of fluency only represent part of the picture and that other variables are clearly taken into consideration when raters make their judgements.

Baker-Smemoe et al. (2014) similarly found that utterance fluency features could broadly predict proficiency level especially at higher levels but that more refined measures would be needed to separate lower bands of proficiency. Tavakoli et al. (2017) on the other hand found that while certain utterance fluency measures (e.g. pruned speech rate) could accurately differentiate between lower levels (A2, B1, B2), they found a 'ceiling effect' for utterance fluency measures at B2 level, with no clear utterance fluency feature which could differentiate between the B2 and C1 students. In sum, it seems that fluency measures might be able to accurately sort learners into proficiency bands from the lower intermediate level to upper intermediate but some other distinction (probably based on grammatical and lexical accuracy) is needed to differentiate among lower (Baker-Smemoe et al., 2014) and higher (Tavakoli et al., 2017) levels.

Kahng (2014) also indicated that proficiency impacted on the extent to which learners were able to remember (during stimulated recall) difficulties in their production during a speaking task. This is likely because the cognitive process underlying utterance assembly for the lower learners relies on declarative knowledge.

3.2.2 Influence of L1 on L2 fluency

A second source of influence is that of a speaker's L1 fluency. The L1 might be expected to impact on a person's fluency in L2 in a number of different ways. Firstly, a person's fluency in their L1 in terms of their speaking style and preferences (e.g. speed of speech, tendency to pause) might impact on their fluency in L2. In other words, if a person speaks with a lot of pauses in their first language then these might be expected to carry over into the speech in L2. A question is therefore the extent to which it is 'fair'[4] to judge individuals' fluency in the L2 without taking into account their speaking style in L1.

Towell, Hawkins and Bazergui (1996) compared the speech rate of English L1 French learners. In general, L1 speech rate correlated strongly with speech rate in the L2. In other words, the faster people spoke in English, the faster they spoke in French. More recently, this was supported by Derwing et al. (2009) and De Jong et al. (2013).

[4] In particular, I am thinking about fluency judgements in language assessment but of course there are also implications for SLA research more generally.

First language might influence L2 fluency in another sense in that certain languages might be more dissimilar from the TL than others, meaning that learners from certain L1 backgrounds might find it more difficult to achieve fluency in a particular L2 than others (Derwing, 2017). Derwing et al. (2009) found that Mandarin learners of English made fewer gains in fluency than L1 speakers of Slavic languages. Derwing (2017, p. 251) explains that "it is conceivable that L1 was a factor here…[because] English and the Slavic languages are both Indo-European whereas Mandarin is unrelated to English".

Finally, L1 might play an inadvertent role in L2 fluency development because it may have a bearing on the extent to which speakers seek out opportunities to interact in the L2. This could be due to a number of factors such as L1 networks, motivational factors and so on (Derwing, 2017).

3.2.3 The role of practice on fluency development

A third source of influence on fluency development is that of practice. As DeKeyser (2017) explains:

> (T)he role of practice in getting to a sufficiently high level of automatization to enable second language use that is both fluent and almost completely accurate is one of the most central topics in instructed second language acquisition (DeKeyser, 2017).

Practice comes in many forms, from a broad view (i.e. general contact and experience with the TL) to the very narrow (i.e. deliberate repetitive practice e.g. classroom drills) (Lightbown, 2008). In this section, I will begin with a look at the role of practice in a broad sense, followed by practice in the more narrow context of formal language instruction.

Language experience

It seems intuitive that the more exposure to and experience of the target language that a learner has, the greater the impact on her fluency. In the main, research has supported this supposition. Segalowitz (2010) explains that "a central theme in virtually all discussions about L2 acquisition is that frequent exposure to elements in the target language (input repetition) and massive production practice (output repetition) are critical for attaining proficiency and fluency" (Segalowitz, 2010, p. 75). It seems obvious that this sort of exposure and practice would be greatest when a learner is *immersed* in the target language and uses the target language on a day-to-day basis (DeKeyser, 2007).

A particular area of fluency studies that has generated a considerable amount of interest is that which has attempted to evaluate the effectiveness of a study abroad (SA) period for fluency development (Llanes & Munoz, 2009; Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004; Valls-Ferrer & Mora, 2014). Studying abroad is common for university students who are undertaking an undergraduate degree in a foreign language. Often, these students will spend a year of their studies living and attending university in a target language-speaking country. For example, a student at a UK university who is studying for a BA in Spanish might spend her second year in Madrid. This works on the assumption that living among speakers of the language will provide the sort of input and interaction necessary to drive language development (and fluency) forward.

> "It is the type of linguistic practice made available to the foreign language learner through SA, usually in the form of opportunities for meaningful interaction with native speakers and massive exposure to L2 input, that may enhance the automatization of already proceduralized linguistic knowledge, provided learners pursue such opportunities for practice and find themselves in a situation where the conditions for quality and quantity of practice are met" (Mora & Valls-Ferrer, 2012, p. 611)

In one of the earliest pieces of SA research, Lennon (1990) used both utterance fluency measures and rater judgements of fluency to compare the speech of four L1 German learners of English before and after a six-moth stay at a university in the UK. The results suggested that the experience of living abroad did impact positively on the fluency of all four participants to varying degrees.

Indeed, fluency is generally believed to develop to greater extents in environments that provide opportunities for meaningful TL interaction and authentic L2 input (Freed, Segalowitz & Dewey, 2004). However, learners' attitudes and willingness to interact can impact upon this (DeKeyser, 2010). As the quotation above from Mora & Valls-Ferrer (2012) suggests, even though learners undergo a period of study abroad in which they are *expected* to seek out plentiful opportunities to interact in the TL, it is not a given that they actually *will* (DeKeyser, 2007). Indeed, it is likely that a wide range of individual factors such as the learner's personality, extroversion, and motivation, are likely to impact on the extent to which they seek out authentic communication during SA. Kinginger (2011; 2013), for example, found that often students do not have a sufficient level in the TL to allow them to interact confidently and that identity plays a significant role in the success of SA programmes. If students lack the necessary communicative competence in L2 (as a result of not being socialised within the TL community) they may find interaction with speakers of the TL uncomfortable and “unpleasant” (Segalowitz, 2010, p. 112). If this is the case, Segalowitz points out that these learners may therefore be further discouraged from interacting with TL speakers, leading to a “downward spiral in L2 fluency acquisition” (p113) in which, because of deficiencies in this dimension of communication, students miss out on opportunities which might otherwise have served to increase their skills in this particular area. This belief is also supported in Segalowitz and Freed (2004) in which two groups of students, an at home (AH) and a study abroad (SA) group were recorded before and after a semester in which the SA group went to live in a TL country (Spain) while the AH group remained in the US. The researchers found significant gains in oral fluency for the SA group when compared to the AH group however they found that there was a non-linear and complex relationship with extra-curricular activities. They suggest that initial performance level may have influenced the extent to which these learners interacted in the TL.

DeKeyser (2017) argues for the centrality of *practice* over and above immersion in the TL. He writes:

> (I)n spite of the change in context from the classroom to study abroad, it is declarative knowledge and practice to proceduralize and automatize it that determine how much fluency is gained, not a completely independent process of acquiring procedural (let al.one implicit) knowledge “from scratch,” without drawing on declarative knowledge (DeKeyser, 2010)

A further point that should be considered is practical in nature. Although time spent interacting with native speakers in a target language-speaking country may be optimal for fluency development, it is not a realistic possibility for many or *most* language learners. These learners typically attend language classes in schools, colleges and universities in the home country and often have little or no opportunity to interact in the TL outside of class time. An important question, then, is *can fluency be taught in the classroom?*

Formal instruction

Chambers (1997) suggested that many people assume that fluency develops naturally over time and that it cannot be 'taught' in a formal sense. Despite this, over the past decades, certain pedagogic interventions have been found to be beneficial for L2 fluency, at least in the short term. Rossiter et al. (2010), Derwing (2017), Tavakoli and Hunter (2017) and Wood (2010) outline some of these activities and I will explore a selection of these in this section.

Formulaic language

Firstly, language teaching may be useful in developing learners' formulaic language. A store of formulaic sequences is considered by many to be a key influence on L2 fluency (Meunier, 2012; Wood, 2010; Wray, 2002; 2012). Defining a formulaic sequence is challenging (Guz, 2016) but it is thought that formulaic chunks of language can either be summoned at the conceptualiser stage of speech production, effectively bypassing the formulation process (Kormos, 2006), or else are retrieved at the formulator stage as a whole chunk which can be articulated as such (i.e. with no internal pauses or hesitation) (Wood, 2010).

There are too many formulaic sequences to simply teach them all in the classroom (Boers & Lindstromberg, 2009). Instead, classroom approaches to the development of formulaic language have focused on raising awareness of "the ubiquity of formulaic language rather than spending time on the direct teaching of particular sequences" (Boers & Lindstromberg, 2012). One such approach is 'text chunking' (Lewis, 1997). Text chunking and other awareness-raising activities are at the heart of 'The Lexical Approach' in L2 teaching. Popularised by Lewis (1993) and Lewis and Gough (1997) and championed more recently in a teaching handbook entitled 'Teaching Lexically' by Dellar and Walkley (2016), the Lexical Approach holds that, since fluency in L1 is supported by the knowledge of stock phrases (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983) then the same may be true of fluent L2 speech. The focus of instruction, it follows, should be on helping learners to build their store of stock phrases rather than focusing on grammatical structure that lexis must then slot into. In other words "language is grammaticalised lexis, not lexicalised grammar" (Lewis, 1993).

Boers et al. (2006) set out to investigate the effectiveness of a formulaic or lexical approach on perceptions of proficiency (see also Boers & Lindstromberg, 2009; 2012). Their hypothesis was that the teaching of formulaic sequences to learners would enhance their perceived proficiency on the fact that 1) formulaic language helps learners to come across as more 'nativelike', 2) because formulaic language is accessed in a single chunk, the reduction in processing time will manifest in more fluent sounding speech (fewer pauses) and 3) formulaic language provides learners with a 'safe zone' in which error is less likely. The 32 participants in their study were split between two groups, an experimental group who underwent training in noticing collocations and formulaic language, and a control group who did not have any training. All participants were rated by two native-speaker judges and additionally, the learners' speech was analysed for evidence of formulaic language. The experimental group were rated as sounding more proficient and their speech was also found to have evidence of more formulaic language. Although this study did not single out fluency as a dependent variable, it is assumed that their notion of proficiency incorporates fluency to some extent. They took their results as evidence for a teaching approach that foregrounded awareness-raising in relation to formulaic language.

Communication strategy training

Another opportunity for teaching to affect fluency is through communication strategy training. A handful of research has looked at the fluency effects of particular strategy training techniques (Guillot 1999; Rossiter, 2003; Tavakoli, Campbell & McCormack, 2015)

Rossiter (2003) investigated the effects of communication strategy instruction on second language performance (communicative success, speech rate, message abandonment) and on learners' ability to paraphrase. Two classes of ESL adult learners were involved in the study. One class received communication strategy training, and the second class received no training. Two different speaking tasks (a picture story narrative and object description) were given as pretests, immediate post-tests, and delayed post-tests. Post-test results showed that the learners who had received communication strategy training had higher fluency (speech rate) in the immediate and delayed post-tests.

Tavakoli et al. (2015) investigated two groups of English for academic purposes (EAP) learners: a control group and an experimental group. All participants performed a monologic task at the beginning of the study and again at the end of the four-week intervention. Over the course of the four weeks, those participants in the experimental group received two sessions a week of specialised communication training which was aimed at enhancing fluency. The training included a combination of fluency awareness-raising activities (e.g. listening to the speech of a non-native speaker and evaluating it for fluency), strategies for improving fluency (e.g. using lexical fillers and being told to avoid repetitions and hesitations) and speaking practice. The researchers found that the experimental group had higher fluency (speech rate, articulation rate, length of run and phonation time ratio) in the post-test when compared to the control group. They concluded that the effective instruction and ample opportunities for practice may have facilitated automatisation of language to some degree.

Drama techniques

In a recent study, Galante and Thomson (2016) compared two forms of instruction and the impact on oral fluency. What is interesting about this study is that the researchers investigated fluency gains on a whole range of tasks, adding to the robustness of the findings. They tested the oral fluency of Brazilian pre-intermediate learners of English before and after 4 months of instruction. While half of the participants received typical, communicative instruction, the other half were taught using 'drama methods' (role plays, improvisation). The pre- and post-test recordings were played to 30 Canadian university students who judged them on a 9-point Likert scale. The students from the 'drama' group were perceived as more fluent at the post test than the traditional group.

Planning

One area of task-based research that has received considerable interest is the impact of 'planning' on task performance. First investigated by Eliis (1987) and Crookes (1989), a large body of research has since emerged (e.g. Foster & Skehan, 1996; 1999; Mehnert, 1998; Ortega, 1999; Wang 2014a; 2014b). It was initially suggested that planning time would impact on L2 task performance because it allows students to do some of the necessary conceptualising work in advance of the task performance itself. Skehan (1998; 2003) suggests that, because humans attention capacity is limited and meaning competes with form for this attention, L2 speakers find it difficult to produce speech which is, at once, complex, accurate and fluent. Instead there are sacrifices or trade-offs made, so that if the learner (either consciously or sub-consciously) focuses on one aspect of performance, there is a negative impact on other areas. Put simply, when an L1 or L2 speaker is called upon to speak spontaneously, they need to do a lot of things nearly simultaneously. They need to think about what they want to say and how they are going to say it. The 'how' part is arguably more difficult for L2 speakers for whom processes of lemma selection and phonological encoding are not automatic. Allowing a student to plan, however, relieves some of the 'what' pressure), providing them with a chance to produce better-than-usual speech.

In terms of fluency specifically, it has been suggested that providing learners with time to plan their performance beforehand results in increased fluency in terms of speed of speech (Tavakoli & Skehan, 2005) and mean length of run (Skehan & Foster, 2005) and also results in reduced mid-clause pauses (Foster & Skehan, 1996). Bui (2014) and Bui and Huang (2016) have also suggested that content familiarity (i.e. having to speak about a familiar topic) also constitutes "latent planning" (Bui & Huang, 2016, p.2) and found that this form of planning also results in increased speech rate. As pointed out in Bui and Huang (2016) however, very few planning studies have focused on fluency specifically and few have looked at multiple indices of fluency in order to identify which aspects of fluency are impacted by the provision of planning time.

Task rehearsal/repetition

Another, related, area of research is task rehearsal, or repetition. In 1988, Gatbonton and Segalowitz (1988; 2005) put forward a framework for teaching fluency which was called "ACCESS" (Automatization in Communicative Contexts of Essential Speech Segments) (p. 328). Central to this framework is the tenet that students should work with tasks that are "genuinely communicative, inherently repetitive and functionally formulaic" (Gatbonton and Segalowitz, 2005, p. 331).

Task *rehearsal* involves speakers performing a task before performing it publically. Task *repetition* involves doing the same task more than once (Bui, 2014). Both could be considered a form of pre-task planning because, in the case of 'rehearsal' the private performance provides the speaker with time to orient themselves to the task etc. In the case of 'repetition' it is the first performance that acts as a 'dry-run' before subsequent performances. Thus, both rehearsal and repetition provide speakers with the opportunity to plan a subsequent performance.

However, task repetition differs from pre-task strategic planning in a number of important ways which are relevant to L2 fluency. First, with repetition, the speaker not only has the opportunity to plan (i.e conceptualise and formulate) the message *mentally* but actually goes through the entire process of speech production; conceptualising, formulating, articulating and monitoring the output. The impact that repetition has on speech production might be expected to differ somewhat from that of strategic (silent) planning alone.

Although task design and methodology, as well as the task itself, have varied considerably between studies, fluency has been found to be both consistently and robustly affected by task repetition (e.g. Ahmadian, 2011; Bygate, 2001; Lynch & Maclean, 2000; 2001; Sample & Michel, 2014; Wang, 2014b). Like the planning studies, however, there have been few studies which have explored fluency in a fine-grained way, adopting multiple measures of fluency in order to identify which aspects of fluency might be affected. Two notable exceptions are Lambert, Kormos and Minn (2017) and N. de Jong and Perfetti (2011). In a very recent study, Lambert et al. (2017) looked at the effect of task repetition on speech rate, filled pause frequency and location, and repair. They found a positive effect for TR on fluency, irrespective of task type and proficiency level, with improvements still being observed after as many as five iterations of the same task. They concluded that while certain aspects of fluency are affected by a single repetition, other aspects (which represent deeper cognitive processing) benefit from a higher number of repetitions. N. de Jong and Perfetti (2011) used multiple measures of fluency to explore both short-term and transfer effects of TR. They found that TR drove learners to speak in longer runs of speech with stable phonation time ratio even on a new task of the same type. They took this as evidence that proceduralisation had taken place during the TR.

3.4 A dynamic systems approach to L2 fluency development

The picture that begins to emerge when one wants to discuss how fluency develops is a complex one in which different factors interact with each other. This was identified by Segalowitz (2010), who put forward a different way of thinking about L2 fluency development which acknowledges the interconnectedness of factors related to it:

> (F)luency can be thought of as a property of L2 use that emerges from the complex interplay of many factors interlinked in a dynamical system. The components of this

system include the neurocognitive machinery for speech production, a motivational system that supports engagement in L2 communication, the social context in which L2 communication is embedded, and environment of communicative experiences that shape the development and processing capacities of the neurocognitive processing system (Segalowitz, 2010, p. 28)

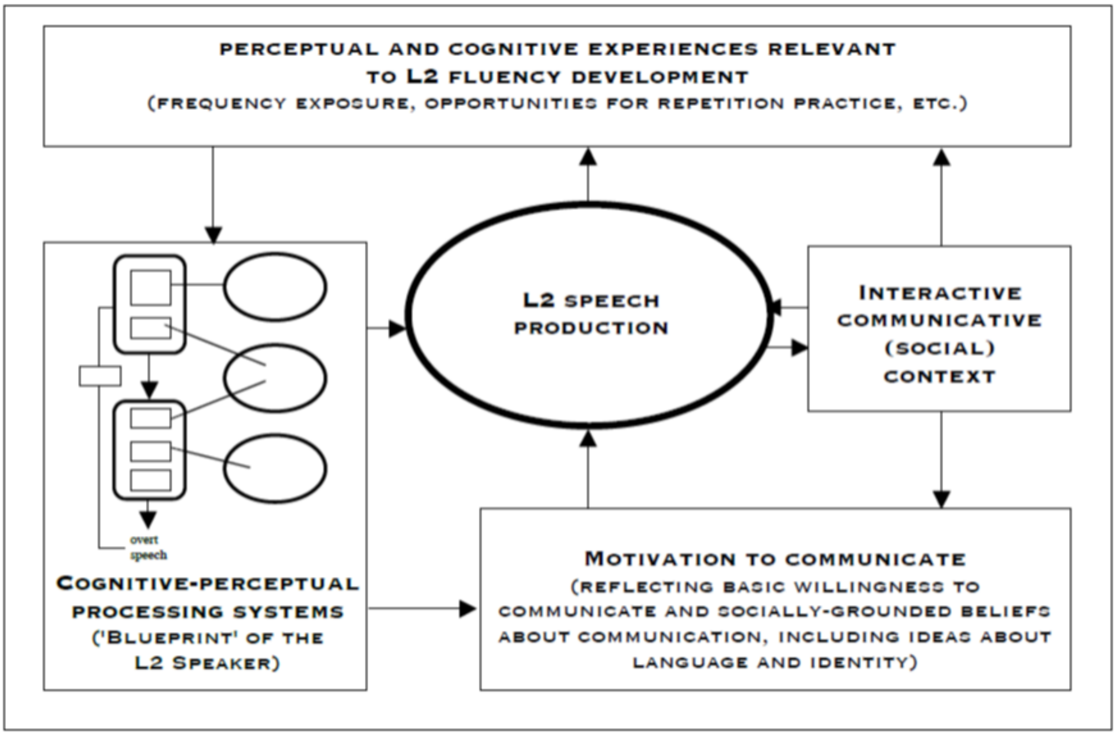


Figure 5: A dynamic systems perspective on L2 fluency (Segalowitz 2010)

Segalowitz (2010) describes the act of producing speech (of which utterance fluency is a component) as being subject to “at least four broadly defined influences that *interact* with one another, creating a system that is continuously changing and adapting over time, and that has the features of a dynamical system” (2010, p. 131). Figure 5 above presents Segalowitz’ framework.

This way of thinking about fluency have implications for language teaching. If language instruction can be delivered in such a way that learners have the opportunity to speak at higher levels of fluency, this could be expected to impact on their sense of self and therefore their willingness to interact with TL speakers which would in turn benefit fluency. Formal language instruction therefore provides an opportunity at which to intervene in this potentially virtuous circle.

3.5 Chapter summary

In this chapter, I presented a number of factors that have been suggested to have a bearing on L2 fluency. In particular, I focused on the impact of proficiency, L1 speaking style and L1 language background and practice. I explained that practice in language learning could be viewed in two different ways: in a broad sense as *experience* and in a more narrow sense as *formal instruction.* It was suggested that many factors interact which influence the extent to which a person gains the experience necessary to foster fluency development and that spending a period of time in a TL country does not necessarily result in increased fluency. In terms of formal instruction, a number of approaches, techniques and activities were discussed including a lexical or formulaic approach, fluency strategy instruction, drama techniques, planning and task repetition. Given the apparently complex and interactive relationship among factors which influence L2 fluency development, it was suggested that a dynamic systems framework might be most appropriate to understand it.

4 – Defining Tasks, TBLT and Task repetition

4.1 Introduction

This chapter will first look at what it means to repeat oneself both in everyday life and in the context of language learning. Next I define and explore the key terms ‘task’, ‘task-based language teaching’ and ‘task repetition’. I then look at the theoretical underpinnings for TR research. I conclude with a short chapter summary.

4.2 Repetition in life and language learning

Every-day language is repetitious (Larsen-Freeman, 2012). We repeat and re-use stock phrases and pleasantries such as “how’s it going?” “fancy a cuppa?” “Lovely weather”. We might also repeat ourselves when we are not understood and need to reiterate something for clarity. Sometimes we may need to give the same information to a sequence of different people in order to get a job done or get some information (Foster & Hunter, 2016). Of course, another way we repeat things in our everyday lives are when we want to commit them to memory. We might, for example, repeat aloud a list of grocery items that we need to get from the supermarket. We may choose to actively *practice* at certain times in our lives, asking a friend or colleague to listen to a presentation that we are working on. We might practice a wedding speech many times in private before the big day. Other times we find ourselves giving the same friend, the same advice time and time again (Foster & Hunter, 2016).

As explained in the previous chapter, not only is practice a part of everyday life, it is also a key factor involved in language learning (DeKeyser, 2010; Lightbown, 2000) The focus of this research - L2 fluency - in particular, is believed to develop through intensive input and output repetition (i.e. practice) (DeKeyser, 2007; 2017; Segalowitz, 2010).

It follows, then, that practice should be at the very heart of L2 pedagogy. Indeed, Dekeyser (2010) and Larsen-Freeman (2012) explain that language teachers have long understood the importance of practice in language teaching. However, exactly *how* this practice should be provided in the classroom has been the focus of much academic and pedagogic debate (DeKeyser, 2010; 2017).

Arguably, the strongest association that people may have for the word ‘practice’ or ‘repetition’ as it applies to the language teaching context, is with audiolingual ‘drills’ and pattern practice (Lynch & Maclean, 2000). Indeed for many years it was widely believed that drilling in the language classroom would lead to fluency (Gatbonton & Segalowitz, 1988). ‘Drilling’ in pedagogic terms, involves the teacher providing a model which is then echoed by the entire class or by selected individual students, as in the following example from Larsen-Freeman (2000, p. 36):

> **TEACHER**: Repeat after me: post office.
>
> **CLASS**: Post office
>
> **TEACHER**: To the post office
>
> **CLASS**: To the post office
>
> **TEACHER**: Going to the post office
>
> **CLASS**: Going to the post office.
>
> **TEACHER**: I’m going to the post office.
>
> **CLASS**: I’m going to the post office

Sometimes (as in substitutive drills) once the drill is established, the teacher can simply suggest different lexis which students then insert. The support for this approach to ‘practice’ came from behaviourist models of language learning in which language learning was seen as the acquisition of ‘behaviours’ that could be strengthened through positive reinforcement and intensive repetition.

However, this account of language learning (both for L1 and L2) is no longer considered valid (Lightbown & Spada, 1999, p. 36) and drills of this kind are limited in their usefulness for language teaching because “they do not teach the students to engage in the target language behaviour of conceptualising a meaning and expressing it through linguistic means, let al.one doing so in creative ways" (DeKeyser, 2010). Nonetheless, elements of this form of repetition continue in language classrooms under the guise of the "highly durable" (Ellis, 2003, p. 29) *Presentation, Practice, Production* (PPP) method in which language is first *presented* to the learner before being *practised* through controlled (scripted) activities, drills and role plays (Gower & Walters, 1983, in Hedge, 2001). There then follows a period of freer *production* in which students are encouraged to use the *practiced* language in more natural ways (i.e. unscripted; spontaneously).

Dekeyser (2010) argues that, while repetition in the form of drills and pattern practice may have limited use for language learning purposes, this does not mean that we should reject practice *in any form.* Larsen-Freeman (2012, p. 206), similarly, argues that "(l)earning takes place not by repeating forms of a closed, static system, but by meaningfully playing the game while revisiting the same territory again and again." A goal of language teaching, then, is to provide learners with this opportunity to revisit the same territory again and again (Gatbonton & Segalowitz, 1988). Given that a dominant approach in language teaching is a task-based or task-supported one, which seeks to foster language learning through the execution of pedagogic tasks (see below), one way of providing practice is through *task repetition.* Task repetition could come in the form of a 'private' enactment of a task before a 'public' one (Essig, 2005). It can also take the form of 'mingles', 'find someone who…', 'speed-dating' and 'fair or exhibition carousel' activities (Denes, 1994; Gatbonton & Segalowitz, 1988; Harsch, 1994; Lynch & Maclean, 1994; Maurice, 1983; 1994; Murray, 1994; Wong, 1994). These sorts of activity have emerged from the need to give learners a chance to attempt a task another time. According to Murray (1994, p. 81):

> (I)n many fluency activities…students have the opportunity to react in a natural communicative manner. However, because of the flow of topics, students may rarely get the chance to improve their performance by trying to get the same information across twice. A fair or exhibition, where people walk around and stop to listen to different short presentations, gives students a chance to give the same talk to the listeners several times in a row. This task can improve both fluency and accuracy

Repetition in this sense does not require students to repeat discrete words and phrases (as in audiolingual drills) but instead encourages students to engage and then re-engage in entire communicative sequences of linguistic performance. It was exactly this sort of repetition that was suggested by Gatbonton and Segalowitz (1988) as particularly conducive for both fluency development and language acquisition more generally. In order to shine some light on what exactly is meant by task repetition, I will first explain what is meant by 'task' and 'a task-based approach to language teaching'.

4.3 Tasks and task-based language teaching

In a recent review of the task-based literature, Ellis (2017) explains that in order to understand what a task-based approach to language teaching might be, it is important to first define what is meant by 'task'.

4.3.1 Defining 'task'

A pedagogic 'task' in language-learning terms can be seen as "a *goal-oriented* activity that people undertake and that involves *meaningful* use of language" (Van den Branden, 2016, p. 240 *my emphasis*). In this sense a 'task' represents the way we use language in our day to day lives. In general, we use language to achieve some sort of personal, social or professional *goal* and the language we use therefore has *meaning.* This is in contrast to an activity which requires learners to focus on the linguistic *form* as is the case with a discrete grammar or vocabulary exercise.

Traditionally, however, language instruction (pre-1970s) largely revolved around 'studying' language in a more analytical fashion: learning about the rules and patterns in language "like an exercise focusing on a single verb form…with the learner being required to work with the target forms at some point of focus isolated by the exercise" (Samuda & Bygate, 2008, p. 8). However, in the 1970s language researchers were exploring theories of SLA which pointed to the importance of communication and, at around the same time, practitioners were beginning to integrate certain 'communicative activities' into their classes in order to encourage learners to communicate in more authentic ways. This was based on the theory that "learners need the opportunity to practise language in the same conditions that apply in real-life situations – in communication, where their primary focus is on message conveyance rather than linguistic accuracy…" (Ellis, 2003, p. 113). These communicative or 'holistic' activities which "involve the learner in dealing with the different aspects of language *together* (my emphasis), in the way language is normally used” (Samuda & Bygate, 2008, p. 8) can be referred to as *tasks.*

Examples of pedagogic tasks include activities in which pairs or groups of learners must sort, rank or order things, solve problems and puzzles, tell stories, give advice, opinions and instructions (Willis & Willis, 2007). An example of a pedagogic task is presented in Samuda and Bygate (2008): the “Things in Pockets” task. In this task, small groups of students are given a bag of small items and told that these items are the contents of a mystery person’s pockets. Each group is given a small form to fill in in which the group must insert their ideas about the mystery person’s identity based on the items they have seen. They also have to say how confident they are about their assumptions as a percentage (i.e. we are 90% sure that the mystery person is a man). Asking students to perform a task like this with whatever linguistic resources they possess is a very different way of thinking about language teaching when compared to ‘traditional’ approaches in which specific language forms are selected for explicit study by the teacher. There are, however, certain disagreements about the exact criteria which can define a ‘task’ in language learning.

Many different definitions of “task” can be found in the SLA literature (Bachman & Palmer, 1996; Ellis, 2003; Nunan, 1989; Willis, 1996) which can range from the very broad (Long, 1995) to the more precise (Willis, 1996). For example, Skehan (1998b) states:

“A task is an activity in which:

- Meaning is primary
- Learners are not given other people’s meanings to regurgitate
- There is some sort of relationship to comparable real world activities
- Task completion has some sort of priority
- The assessment of the task is in terms of outcome”

Building on this definition, Willis and Willis (2007, p. 13) offer a number of questions which teachers can ask about a particular classroom activity to determine how ‘task-like’ it is. They argue that the more confidently a teacher can answer “yes” to these questions, the more ‘task-like’ the activity is. I reproduce these questions here along with my responses as they might apply in relation to the “things in Pockets” task outlined above.

1. Does the activity engage learners' interest? *Yes, the activity is likely to engage learners' interest. I can imagine that learners would be keen to look at each item and speculate about its owner*
2. Is there a primary focus on meaning? *Yes, there is no specific language focus during the task itself (although there is the opportunity for a post-task focus on modal verbs "he must be an electrician" etc)*
3. Is there an outcome? *There is a clear outcome – students have to fill in a form which gives their opinions about the mystery person's identity*
4. Is success judged in terms of outcome? *Yes, there is no other 'test' of linguistic ability*
5. Is completion a priority? *Yes, students are expected to complete the task within the specified time frame*
6. Does the activity relate to real world activities? *Yes, these kinds of conversations are normal in everyday life.*

Willis and Willis (2007) argue that the first of these, *engagement,* is particularly important because "without engagement, without genuine interest, there can be no focus on meaning or outcome. Learners have to want to achieve an outcome, they have to want to engage in meaning" (Willis & Willis, 2007, p. 13).

To offer a short summary, then, a pedagogic task is a classroom activity which engages learners' interest and which requires them to use their existing linguistic resources to carry out meaningful activities in order to achieve a non-linguistic goal. This contrasts with analytical activities which might require learners to focus on the form and meaning of language, as in grammar activities.

4.3.2 Tasks in research versus tasks in the classroom

As Ellis (2003) explains, tasks are of interest to language *teachers* and language *researchers* (and language *testers* (Bygate, Skehan & Swain, 2001)). Researchers are interested in tasks because they elicit 'samples' of learner production (in the case of productive tasks) which are similar to the way that language is produced in naturalistic settings. This means that researchers can analyse the language which learners produce during tasks and use that as a base for discussions about language acquisition (conditions, individual differences etc.)

In fact, tasks have been central to research in SLA with studies asking a wide range of questions in relation to task design (Pica, Kanagy & Falodun 1993; Robinson, 2001; Skehan, 2001), the processes with which learners engage during task performance (Bygate, 1988; 1999; Ellis, 2001; Mackey, 1999), the way that learners negotiate for meaning during tasks (e.g. Doughty and Pica, 1986) and construct meaning (Ohta, 2000; Swain & Lapkin, 2000; 2001), provide feedback to one another (Lyster, Saito and Sato, 2013) display individual differences in terms of motivation on tasks (Dörnyei & Kormos, 2000), the impact of different implementation conditions (Bygate & Samuda, 2005; Foster & Skehan, 1996; Skehan & Foster, 1997; Yule & McDonald, 1990) and the impact of task engagement on language learning (Izumi & Bigelow, 2000; Kim & Tracey-Ventura, 2013; Mackey, 1999; Ellis, 2001).

Bygate et al. (2001) and Gatbonton and Segalowitz (2005) however, point out that often the sorts of tasks used for research purposes often fail to meet certain criteria which would qualify it as being a 'true' pedagogic task. This has important ramifications if the aim of research is to make recommendations for task-based language teaching (Samuda & Bygate, 2008). A further difference between how tasks are implemented for research purposes and teaching purposes is that, in the latter, tasks are hedged with pre- and post-task activities. In other words they are part of a pedagogic *sequence.* A final difference between tasks in research and tasks in classrooms is that the former are often implemented in a laboratory environment which can have different characteristics from a language classroom (Nassaji, 2012; Nunan, 1991; Samuda & Rounds, 2001).

4.3.3 Task-based versus task-supported approaches

The use of pedagogic tasks in language teaching has a long history. In the 1970s, communicative language teaching began to generate a following of teachers who believed in the need to base language learning in genuine communication (Widdowson, 1978, Brumfit & Johnson, 1979). However, as Van den Branden (2016) explains, the communicative approach quickly became absorbed into more traditional PPP approaches. Prabhu (1987), Long (1985) and Pica (1987) were among the first to suggest that a teaching methodology could be entirely centred around tasks and task performance. In other words, that communicative activities need not only *supplement* language teaching, they could be the *foundation.*

Task-based language teaching (TBLT) is now considered a "widespread approach to L2 teaching in many contexts" (Andon & Eckerth, 2009, p. 287) and has been "vigorously promoted" as a part of language teaching policy worldwide (Samuda & Bygate, 2008, p. 195)

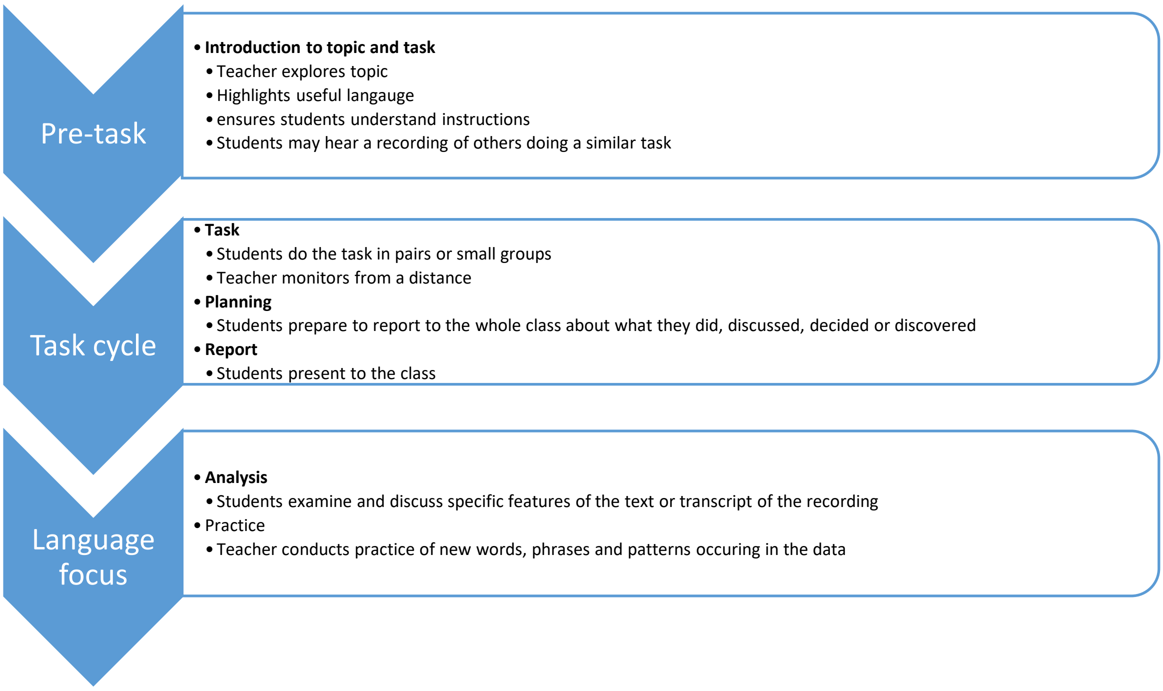


Figure 6: Sequence for a task-based lesson (adapted from Willis, 1996, p. 38)

Although, specific 'official' guidelines for a task-based approaches exist (Figure 6), teachers "do not follow 'official' TBLT-related pedagogic recommendations in a slavish way…[they] reject some of them, embrace others, and combine all of them with other pedagogical elements" (Andon & Eckerth, 2009, p. 305). This is an observation which is supported, more recently, by Van den Branden (2016, p. 246) who explains:

> "in most cases tasks have been integrated in eclectic, hybrid approaches which appear to present teachers with an acceptable mix of the traditional approaches they are familiar with and the innovations they are able, and willing to digest"

In terms of using tasks in the classroom, then, it might be best to speak of a task-*supported* approach to language teaching (Long, 2014; Ellis, 2003; Samuda & Bygate, 2008). In a task-supported syllabus, tasks are used as and when teachers feel they may be useful, alongside other methods, in a complementary fashion. Certainly, the English language school which was the setting for the present research would probably be described as favouring a task-

supported approach, with teachers using a grammar-based coursebook and using tasks to provide additional practice and consolidation or to generate interest in themes.

Part of the reason for the unwillingness of teachers to universally adopt a strong version of TBLT (such as that which is described in Willis, 1996) is that they are unsure as to how tasks will facilitate language learning because of the apparent lack of specific language focus. While a strong version of TBLT would hold that the tasks themselves are all that is required to drive acquisition, more recent research in SLA would point to the fact that there is an important role for the teacher (Samuda, 2001) and that task conditions can be manipulated in order to ensure that learners are able to focus on the language which occurs during task performance to develop their interlanguage (Skehan, 1998; 2003; Foster, 2009) Indeed, much SLA research has pointed to the need for tasks to encourage learners to notice certain aspects about TL form, receive corrective feedback, and negotiate for meaning if they are to promote language acquisition (Mackey, 2012; Gass & Mackey, 2015; Mackey, Abbuhl, & Gass, 2012; Swain, 2005). A challenge which exists in language pedagogy, then, "is to choose, sequence and implement tasks in ways that will combine a focus on meaning with a focus on form" (Foster, 1999, p. 69).

4.4 Defining task repetition

A possible solution to this problem is offered by Willis and Willis (2007). They explain that teachers are, in fact, able to manipulate the *conditions* of the task and the way that a task-based lesson is *sequenced* in order to bring about certain changes to the task demands (for example to provide the opportunity to focus on form). For example, a teacher may provide students with the chance to prepare or 'plan' what they are going to do in the task in order to give rise to higher quality of performance during the pre-task phase in the task procedure (Figure 6). Another option would be to give students the chance to listen to recordings of themselves performing a task which they then transcribe (Lynch, 2007) during the 'language focus' phase (Figure 7). Yet another option would be to provide learners with the opportunity to perform the task another time after the first performance. This could be termed *task repetition.*

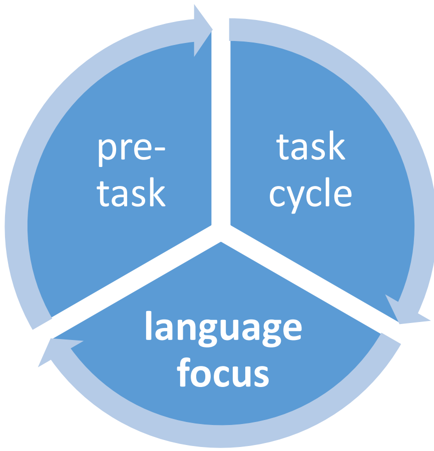


Figure 7: Willis' 1998 task sequence adapted to include task repetition

In very simple terms, then, task repetition requires learners to carry out a task on more than one occasion. If we return to the ‘Things in Pockets’ task which I introduced at the start of this section, exact repetition of this task would involve learners looking at the items and making judgements about the life of the person that owned them, and then doing it all over again, possibly after some kind of language focus. In terms of authenticity, repeating this task with exactly the same ‘items’ and with exactly the same peers might be seen as inauthentic and learners may no longer be engaged (Michel, 2017; Willis & Willis, 2007). However, one can imagine ways in which the task could be altered slightly to allow learners to ‘try again’ in order to build on what they had learnt but at the same time preserve the authenticity of the activity. Arguably, there are at least three ways of achieving this: (1) the same groups could look at the contents of *another person’s pockets.* In other words, each group could be presented with a different collection of items and go through the same procedure again. (If there was more than one group in the class, this could be easily manipulated by simply rotating the items between the groups). (2) The same groups of students could be asked to perform a slightly different task but using the same ‘items’. This could take the form of, for example, asking students to rank the items in terms of how valuable they might be to the owner. This time some of the *content* of the task is repeated but the task itself is new. (3) The task could be repeated with the same content and procedure but with new combinations of students in each group. This would mean that the task outcome may be different because of the different opinions that these new students will bring to the discussion.

In his doctoral research, Patanasorn (2010) identified *three* types of task repetition: 1) task repetition, which involves repetition of the exact same task; 2) procedural repetition, which involves the repetition of the procedure but with different content; and 3) content repetition, in which the procedure changes but the content or material remains the same. These latter two correspond with options 1 and 2 in the preceding paragraph.

Table 7: Different types of repetition and characteristics

	Same content/material?	**Same procedure?**
Task repetition	YES	YES
Procedural repetition	NO	YES
Content repetition	YES	NO

As Table 7 above shows, task repetition involves repeating the exact same task on a subsequent occasion, with the same material (e.g. 'things' in pockets), the same procedure (e.g. speculate on identity) and the same interlocutor (e.g. same group of learners). Procedural repetition involves repeating the same task procedure on a subsequent occasion, with new material (e.g. a new bag of 'things'), the same procedure (e.g. speculate identity) and the same interlocutor (e.g the same group of learners). Content repetition involves reusing the same material on a subsequent occasion, with the same material (e.g the same 'things') but different procedures (e.g. 'which items are most valuable and why?').

What Patanasorn (2010) did not investigate is the type of repetition which I outlined above which requires students to repeat the same task but with fresh interlocutors. This arguably constitutes a fourth type of repetition which I will term 'ecological' repetition[5].

[5] I use the term 'ecological' because I believe it reflects the 'ecology' of a task-based language classroom, in which it is common to interact with different interlocutors. An alternative may have been something like 'mingle' repetition, but it was felt that this would be too niche a definition.

Because of their centrality to the current study, I will look in a little more detail at each of these types of repetition.

4.4.1 Exact repetition of the task (task repetition)

None of Patanasorn's (2010) definitions is entirely unambiguous or uncontroversial. Task repetition, for example, is intended to refer to the repetition of the *exact* same task but this begs the question if any subsequent task can ever be *exactly* the same because the very fact that the task has been carried out previously means that the task demands are now different and the learner's orientation to the task is likely to be different (Larsen-Freeman, 2009; 2012). This is without even considering the impact that a different day or time of day or all the myriad factors that might influence performance might have on the execution of a task. There is also a question mark over the pedagogic validity of repeating a task exactly. As we have seen, one of the aims of a TBLT approach is to encourage the use of tasks which have interactional authenticity. It may be suggested, therefore, that exact TR does not really fulfil this criteria.

4.4.2 Repetition of the type of task (procedural repetition)

Procedural repetition, too, can clearly have multiple interpretations. In Sample and Michel (2014) the participants carried out a spot the difference task in pairs. The picture remained the same but the location of the items that the learners had to 'spot' changed on each iteration. In Pinter (2007), learners engaged in a spot the difference task in which the location of items, and *the items themselves and the picture* changed each time. Both of these studies could be said to be dealing with 'procedural' repetition, and yet the demands and outcome is likely to be different (i.e. for lexical selection).

4.4.3 Repetition of the task content (content repetition)

Content repetition, although not commonly investigated, could be said to have similar issues to procedural repetition. To my knowledge only two studies have considered this approach to repetition (Garcia-Fuentes & McDonough, 2016; Patanasorn, 2010). It therefore remains to be seen how other researchers will interpret this type of repetition in their studies and what the findings will be. A review of the studies in the area of task repetition reveals that the majority have investigated exact task repetition and/or procedural repetition.

4.4.4 Repetition of the same task or task procedure but with different interlocutors (ecological repetition)

As explained above, *exact* task repetition could be said to have lower validity because in real life it is uncommon to re-engage in a task with the same person once the goal has been achieved. However, if the task is carried out with a novel interlocutor each time then "they can rely on their previous performances of the same task to a limited extent only since their interlocutor's contributions will always bring some novelty to the joint interaction" (Pinter, 2007, p. 190)

As suggested in Lambert et al. (2017) and Lynch and Maclean (2000) when the interlocutor changes each time, so too will the exact demands of the task. Therefore it would be incorrect to refer to this sort of repetition which involves alternating interlocutors as simply 'task repetition'.

> "(T)he arrival of a new visitor should present the [speaker] with a novel challenge…the [interlocutor] is not simply a cipher, but a communicative partner who takes the initiative by asking the questions and then reacts to the adequacy and comprehensibility of the [speaker's] response" (Lynch & Maclean, 2000, p. 242)

While Kim and Tracy-Ventura (2013) suggest that Lynch and Maclean's version of TR actually constitutes PR, I propose that this sort of repetition which is interactive and in which the interlocutor changes each time actually constitutes a different type of repetition. It could be applied to all three types of repetition outlined above (task, procedural and content). In other words, it is possible to repeat the same task with a different interlocutor, the same procedure with a different interlocutor and the same content but with a different interlocutor.

Table 8: Studies which have explored repetition effects on fluency shown in terms of type of repetition investigated

Study	Task repetition	Procedural repetition	Ecological?
Ahmadian & Tavakoli, 2011	YES	NO	NO
Ahmadian, 2011	YES	NO	NO
Boers, 2014	YES	YES	YES
Bygate, 1996	YES	NO	NO
Bygate, 2001	YES	YES	NO
Fukuta, 2016	YES	NO	NO
Hsu, 2017	YES	NO	NO
Kim & Tracy-Ventura, 2013	YES	YES	YES
Lambert et al., 2016	YES	NO	YES
Lynch & Maclean, 2000; 2001	YES	NO	YES
N. de Jong & Perfetti, 2011	YES	YES	NO
Sample & Michel, 2014	NO	YES	NO
Thai & Boers, 2016	YES	YES	YES
van de Guchte et al., 2016	NO	YES	N/A
Wang, 2014b	YES	NO	NO

4.4.5 Inherently repetitive task sequences

While task repetition can be *engineered* in the sense that either for research or pedagogic purposes, learners are simply asked to engage with a task another time (as in the Things in Pockets example above), there is an interesting alternative in that certain existing classroom activities involve sequencing tasks in such a way that there is an “inherently repetitive” (Gatbonton & Segalowitz, 2005) element built in. Gatbonton and Segalowitz (1988; 2005) and Segalowitz (2010) argue that these sorts of task sequences are particularly conducive to language learning because the “learner remains involved in genuine communication while engaged in repetition” (Segalowitz, 2010, p. 176). It is best to talk about these inherently repetitive pedagogic activities as ‘task sequences’ rather than ‘tasks’ because they involve repeated engagement with individual ‘tasks’ as well as particular pre- and post-task stages.

In “New Ways in Teaching Speaking” (Bailey & Savage, 1994), there are a number of ‘inherently repetitive’ task sequences put forward by different contributors (e.g Murray,

The Fluency Workshop (from Maurice, 1994)

Procedure

1. **Choose a few topics for class discussion that are interesting to and at the level of your students**
2. **Pair off the students and assign half the students Topic A and the other half Topic B**
3. **Explain the rationale of the activity and draw a chart like the one below on the board.**

	A person/topic	B person/topic	Changeover
Partner 1	4 minutes	4 minutes	1 minute
Partner 2	3 minutes	3 minutes	1 minute
Partner 3	2 minutes	2 minutes	Finish

4. **Have A students speak about their topic for 4 minutes to one B students (Speakers commonly struggle in the first 4 minutes, but encourage them to keep speaking and encourage the listeners to ask questions to keep the speakers talking.)**
5. **Be the time keeper and stop the discussions at the appropriate time.**
6. **Staying with the same partner, have B students speak about their topic for 4 minutes to the A students.**
7. **During a 1-minute changeover, rotate students to form different A/B pairs, and repeat the discussion process for 3 minutes.**
8. **Rotate students to form different A/B pairs and repeat the discussion process for 2 minutes.**
9. **Monitor the time and listen for recurring errors and interesting content that can be used for later discussion. Clarify important points at the changeovers, if necessary.**

Figure 8: The fluency workshop (from Maurice 1994)

1994; Deines, 1994; Wong, 1994). I will focus here on two contributions which are particularly relevant to L2 fluency development and task repetition: Maurice's (1994) 'Fluency Workshop' (Figure 8) and Lynch and Maclean's (1994) 'Poster Carousel' task sequences (Figure 9).

The Poster Carousel (from Lynch & Maclean, 1994)

Procedure

1. **Find suitable short research articles (one for each pair of students)**
2. **Divide the class into pairs and give each pair an article to read.**
3. **Have partners agree on a summary of content and findings in the form of a poster, using no more than two flipchart sheets of paper.**
4. **Tape or pin the finished posters on the classroom wall far enough apart to allow viewing and discussion without interfering with a neighbouring pair of students.**
5. **Have one student from each pair stand by their poster to respond concisely to any questions or points raised by other students visiting the poster. (However, hosts should not initiate discussion or explanation.)**
6. **Have the other member of each pair join the pool of visitors who each go to a different station every 5 minutes or so to read and absorb the information in the poster. Encourage the visitors to raise questions and points of disagreement with the presenter.**
7. **Signal (with a bell or whistle) when time is up and have each visitor move on clockwise to the next poster in the carousel.**
8. **Repeat the time cycle as many times as necessary for each visitor to view and ask questions about all the posters**
9. **Have the members of the original pairs then switch roles, so that the ones who have been answering questions can work their way around the carousel as visitors.**
10. **Monitor the question-and-answer interaction and make notes of points to comment on in plenary at the end of the activity**

Figure 9: The Poster Carousel (taken from Lynch & Maclean, 1994)

The procedure for each of these task sequences is outlined above (Figures 8 and 9). There are similarities between them both in the fact that the task sequences *necessitate* repetition in order to complete the task. The repetition in both task sequences is also *ecological* because it requires learners to interact with different interlocutors each time they repeat the task. In both cases the repetition is immediate (or with a 1 minute 'changeover'). Differences between the two task sequences are that the Fluency Workshop requires learners to repeat the task with increasing time pressure (i.e. allotted time to complete the task reduces from 4 minutes, to 3 and finally, to 2). A further difference is that in the Poster Carousel, the stimuli for the interactions is external because each poster is based on a journal article.

4.5 Theoretical underpinnings of TR and PR research: Planning, Priming and Proceduralisation

Having established what is meant by task repetition and the different ways that it might be interpreted by teachers and researchers, I will now turn to the theories underpinning TR research. As I will show, some TR studies have focused on fluency on the same task (short-term effects[6]) while others have looked at whether fluency transfers to a new task. The theoretical underpinnings are different for each of these.

4.5.1 Short-term performance effects

Planning

Bygate (1996; 2001) explains that task repetition might have an impact on performance of a repeated task because the initial task performance gives learners the chance to orient themselves to the task and familiarise themselves with task demands. It also provides them with the opportunity to work out what they are going to say and "part of the work of conceptualisation, formulation and articulation carried out on the first occasion is kept in the learners' memory store and can be reused on the second occasion" (Bygate, 2001, p. 29). In terms of linking this thinking to Levelt's model of speech production, which was discussed earlier, Bygate suggests that during the first performance of any task, learners need to *conceptualise* the semantic *content* of their message as well as formulate and articulate their message in real time. During a second or subsequent performance, however, much of this planning work has been taken care of. This, in turn, 'frees-up' attentional capacity which can be diverted towards delivering the same message at a higher level of performance (i.e. CAF). In this sense, repetition is seen as a form of pre-task (strategic) planning. Bygate (2001) explains that *procedural* (task type, in his words) repetition could have similar effects because discourse features can be the same across different tasks of the same type and therefore PR allows learners to familiarise themselves with the discourse features of the task:

> "different examples of the same task type share certain characteristic discourse features, such as elements of narrative structure… experience of handling discourse

[6] Lambert et al. (2017) refer to this as "immediate" effects but I prefer short-term to avoid confusion with 'immediate' repetition

types is stored in long-term memory, contributing to communicative competence, and affecting the effectiveness with which a speaker carries out a task" (Bygate, 2001, p. 29)

The theoretical background for so-called 'planning' effects is Skehan's limited attention hypothesis (sometimes referred to as the 'trade-off' hypothesis). The limited attention hypothesis holds that because human beings' attentional capacity is limited, language learners are unable to simultaneously focus on all aspects of performance (i.e. CAF). The result is that form and meaning compete for attentional resources which in turn means that some aspect(s) of performance are reduced. Much research has focused on how tasks and task conditions can be altered such that they allow learners to overcome such 'trade-offs' between aspects of performance (see, for example, the discussion of planning in the previous chapter).

Priming

Others (e.g. Wang, 2014; Skehan, 2014; Fukuta, 2016; Lambert et al., 2017; N. de Jong & Perfetti, 2011) have suggested that in addition to providing *the* opportunity to plan content, task repetition also impacts on subsequent performance because the act of having assembled the appropriate language and articulated it once before means that a 'blue-print' of the utterance is stored in short-term memory (Fukuta, 2016) allowing learners to "activate recently used linguistic constructions" (Lambert et al., 2017, p. 5) in a slicker fashion the second time round. This explanation is often referred to as 'priming effects'. Exactly how priming impacts on speech is seldom explored in TR studies (although forthcoming research by Bui, Ahmadian and Hunter will hopefully address this). Wang (2014) and Skehan (2014), however, argue that priming effects are stronger in task repetition than in planning studies because learners have gone through the entire speech production process (i.e. conceptualisation, formulation, articulation) as opposed to planning silently.

4.5.2 'Transferal' of performance effects

As we will see, while some studies have looked exclusively at the changes in CAF which are found on the repeated performance of the task (short-term effects), others have looked at the impact that TR might have on performance of a completely new task (transfer). In many studies, analysing CAF on the performance of an entirely new task is a means to capturing the extent to which the intervention has impacted on *acquisition.*

Proceduralisation

While short-term effects are often explained in terms of content planning and lexical, syntactic and phonological priming (see above), fluency *transfer* is argued to be related to *proceduralisation* and/or the subsequent automatization of linguistic knowledge.

As explained in Chapter 2, fluency is thought to be related to automaticity in speech production, and automaticity in speech production can be understood in at least two different ways. It can understood as the ability of a speaker to access chunks of language in a single-step (as in item-based accounts of automaticity in language processing) or it can be seen as the conversion of linguistic rules from declarative to procedural knowledge and execution (as in rule-based accounts of automaticity) (Kormos, 2006). N. de Jong & Perfetti (2011), drawing on Anderson and Lebiere's (1989) ACT-R model (a rule-based account) explain that task repetition might provide the necessary practice to help learners collapse smaller production rules into larger rules and then allowing these new, more efficient rules to gain strength "to be able to compete with (and defeat) their "parent" production rules" (N. de Jong & Perfetti, 2011, p. 538). In other words, their view is that task repetition might provide sufficient practice to enable this conversion from declarative knowledge to procedural.

4.6 Chapter summary

I began the chapter with a look at how 'repetition' is a part of life and of language teaching. Next I defined 'task' and presented four different types of task repetition: (1) repetition of the exact same task (task repetition), (2) repetition of the task procedure (procedural repetition), (3) repetition of the task content (content repetition) and (4) task repetition/procedural repetition which is carried out with different interlocutors each time (ecological repetition). I explained how some pedgagogic tasks can be classed as 'inherently repetitive' and I looked at the theoretical underpinnings of TR/PR research, which generally link short-term fluency gains during TR to content planning and/or priming, and fluency transfer gains to proceduralisation. I will turn, now, to explore the available TR literature in more detail.

5 - Task repetition research

5.1 Introduction

The body of research in the area of task repetition is already substantial and rapidly growing. In order to identify the literature that was particularly relevant to the current study, it was first important to take a broad look at how task repetition has been investigated. Studies have differed in terms of (1) the participants, (2) the operationalisation of independent variables and (3) the operationalisation of dependent variables used[7]. I then go on to look at key questions in TR research and the studies that have sought to shed light on them. I conclude this chapter with a brief chapter summary.

5.2 Participants in TR research

In terms of participants, TR research can be divided into that which has looked at TR effects with young learners (Azkarai & Garcia Mayo, 2016; Kim, 2013; Kim & Tracy-Ventura, 2013; Pinter, 2005; 2007; Sample & Michel, 2015; Shintani, 2012; 2014; Van de Guchte et al., 2016) and with adults (Ahmadian, 2011; Ahmadian & Tavakoli, 2011; Arevart & Nation, 1991; Boers, 2014; Bygate, 1996; 2001; Bygate & Samuda, 2005; N. de Jong & Perfetti, 2011; Fukuta, 2016; Gass et al., 1999; Hsu, 2017; Lambert et al., 2017; Lynch & Maclean, 2000; 2001; Nation, 1989; Plough & Gass, 1993; Thai & Boers, 2016). The reason that age might be important is that children are likely to engage in tasks (particularly those which are interactive) in very different ways from adults because "(t)he ability to take full responsibility for ones' own utterances as well as one's understanding of the partners' utterances are skills gradually increasing with age" (Pinter, 2007, p. 191). Mackey and Silver (2005) write that no SLA finding should be generalised to children without sufficient empirical support and presumably the same is true

[7] Methodological differences among TR studies in terms of research design and methodological approach is taken up in Chapter 7.

vice versa. No study has yet investigated age interaction with TR empirically, although doing that will prove difficult in terms of keeping other variables constant (i.e. age of onset, proficiency etc).

Repetition effects have been investigated with participants with a range of L1s, for example Japanese (Fukuta, 2016; Lambert et al., 2017), Dutch (Van de Guchte et al., 2015), Spanish (Azkarai & Garcia Mayo, 2016), and Korean (Kim & Tracy-Ventura, 2013). L1 background might have an influence not only on *what* learners do when they repeat a task but also on their *attitude* to task repetition as we know that "different societies can vary drastically in their emphasis on or tolerance of repetitiveness" (DeKeyser, 2010, p. 160). Furthermore, studies have suggested that L1 background might impact on L2 fluency in a number of ways (Derwing, 2017; Derwing et al., 2008). Very few studies have looked at multilingual groups of learners, although exceptions are Bygate (2001) Bygate and Samuda (2005) and Lynch and Maclean (2000; 2001).

There is variation, too, in the proficiency level of participants that have been involved in TR research. In some studies the participants cover a range of proficiency levels (Boers, 2014; Sample & Michel, 2014; Lynch & Maclean, 2000; 2001) within a heterogeneous group. It is fair to say that most studies have used participants who are towards the higher end of the CEFR (i.e intermediate to advanced). There is no indication that task repetition might have different effects with learners of different levels of proficiency (Lambert et al., 2017).

5.3 Independent variables

In terms of how TR is operationalised, the research can be split into that which has looked at *task* repetition (repetition of the exact same task) (Ahmadian, 2011; Ahmadian & Tavakoli, 2011; Arevart & Nation, 1991; Azkarai & Olivier, 2016; Boers, 2014; Bygate & Samuda, 2005; de Jong, 2012; Hsu, 2017; Lambert et al., 2017; Lynch & Maclean, 2000; 2001; Nation, 1989; Thai & Boers, 2016; Wang, 2014;), *procedural* repetition (repetition of the same *type* of task but with new content) (Garcia-Fuentes & McDonough, 2016; Pinter, 2005; 2007; Sample & Michel, 2014; van de Guchte et al., 2015) or *both* within a single study (Azkarai & Garcia Mayo, 2016; Bygate, 2001; Garcia Mayo & Imaz Agirre, 2016; Fukuta, 2016; Kim, 2013; Kim & Tracy Ventura, 2013; Patanasorn, 2010; Payant & Reagan, 2016; Takimoto, 2012).

Some studies have looked at what happens when the task is repeated *immediately* (Boers, 2014; Lambert et al., 2017; Lynch & Maclean, 2000; 2001; N. de Jong & Perfetti, 2011; Thai & Boers, 2016), while others have looked at repetition after a period of a few days (e.g. Bygate, 1996; Pinter, 2005), weeks (e.g. Fukuta, 2016; Hsu, 2017) or even months (Azkarai & Garcia Mayo, 2016). This may be argued to impact on the priming effect of TR (explained in the previous chapter) because learners may be more able to draw on content planning and priming benefits of TR if the initial iteration was relatively recent (Ahmadian, 2016; Ahmadian & Bui, in prep). As de Jong (2012) explains: "immediate repetition makes it more likely that benefits of conceptualization and formulation persist into the repeated deliveries" (de Jong, 2012, p. 44).

Using Ellis' (2001, p. 49) concept of 'reciprocity' in pedagogic tasks, the 'treatment' tasks used in TR studies have been either non-reciprocal (monologic) or reciprocal (dialogic) (Table 9). Ellis (2001) explains that reciprocity in task performance can be seen as a continuum rather than a dichotomy. At the extreme end of this continuum might be tasks as defined in N. de Jong and Perfetti (2011) where monologues are delivered at a computer and recorded via a microphone. At the other end of the continuum there are studies which have been highly interactive and involved active collaboration between participants (e.g. Pinter, 2005; Sample & Michel, 2014).

Table 9: Repetition studies and degree of reciprocity

Task type	**Studies**
Non-reciprocal task	Bygate, 1996; Bygate, 2001; Wang, 2014; Gass et al., 1999; N. N. de Jong & Perfetti, 2012; De Jong, 2012; Boers, 2014; Thai & Boers, 2016; Ahmadian, 2011; Ahmadian & Tavakoli, 2010; Hsu, 2017; Lambert et al., 2017
Reciprocal task	Lynch & Maclean, 2000; 2001; Sample & Michel, 2014; Pinter, 2005; Payant & Reagan, 2016; Kim, 2013;

It becomes clear that the majority of tasks used in TR studies have been towards the non-reciprocal end of the continuum. This is particularly true of the studies which have looked at TR effects on fluency and especially in larger scale studies. There are also clearly implications for the measurement of fluency in these more interactive tasks, with both Pinter (2005) and Sample and Michel (2014) providing fluency measurements on a 'per dyad' basis. The non-reciprocal task bias in TR studies is an issue because while monologic tasks are more stable and manageable from a research point of view (Witton-Davies, 2014), interactive tasks are more common in language classrooms (Foster, 2009; Willis & Willis, 2007). Furthermore, monologic tasks, while they are more amenable to fluency analysis (Hilton, 2014) are less 'interactionally' and 'situationally authentic' than interactive tasks because it is rare for individuals to launch into a monologue in real-life without any possibility of interjection from another person (Witton-Davies, 2014).

5.4 Dependent variables

In terms of dependent variables, while a small handful of studies have investigated TR effects on L2 written production (Amiryousefi, 2016; Indrarathne, 2016; van de Guchte et al., 2015), the vast majority of studies have focused on oral production. These studies can be broadly divided into those that are interested in interactional patterns of behaviour and those which look at L2 'performance'. Studies have looked at the impact of TR on interactive behaviour in communication such as 'amount of L1 use' (Askarai & Garcia Mayo, 2016) and 'language related episodes' (LREs) (Hawkes, 2012; Kim, 2013; Kim & Payant, 2014; Payant & Regan, 2016). A common finding of these studies is that procedural repetition encourages higher numbers of LREs than task repetition and that during peer-peer interaction, repetition results in many of these LREs being resolved correctly (which, they argue, is essential for language learning to take place).

Other studies have looked for TR effects on sociopragmatic features such as 'argumentation' (Nemeth & Kormos, 2001), use of politeness strategies (Garcia-Fuentes & McDonough, 2016) and 'request downgraders' (Takimoto, 2012), specific syntax such as past tense markers (Patanasorn, 2010), accuracy on particular forms (Gass et al., 1999), or discourse markers (Bygate & Samuda, 2005). Some have used holistic or evaluative measures of performance (Gass et al.l, 1999; Lynch & Maclean, 2000; 2001) while others have measured performance in (at least one of) the aspects of oral complexity, accuracy and fluency (CAF) (Ahmadian, 2011; Ahmadian & Tavakoli, 2010; Bygate, 1996; 2001; Boers, 2014; N. de Jong & Perfetti, 2011; Fukuta, 2016; Hsu, 2017; Kim & Tracy-Ventura, 2013; Lambert et al., 2017; Sample & Michel, 2015; Thai & Boers, 2016; Van de Guchte et al., 2015; Wang, 2014).

As explained in Chapter 2, although many studies explore performance using indices of complexity and accuracy along with fluency, the primary focus of this literature will be on repetition effects on *fluency* specifically and other aspects of performance are discussed in relation to fluency. The studies that have investigated repetition effects on fluency have differed in how they have measured these aspects operationally (see Table 10). As we saw in Chapter 2, the wide range of measures for fluency both within individual studies and among studies has been highlighted as problematic (Segalowitz, 2010; Palotti, 2009; Lambert & Kormos, 2014).

Table 10: Measures of fluency in TR studies

Study	Fluency measures
Ahmadian & Tavakoli (2010)	Raw speech rate Pruned speech rate
Ahmadian (2011)	Raw speech rate Pruned speech rate
Arevart & Nation (1991)	Speech rate (words per minute) Number of hesitations per 100 words
Boers (2014)	Speech rate (words and syllables per minute) Number of hesitations
Bygate (1996)	Number of repetitions
Bygate (2001)	Unfilled pauses per T-unit
Fukuta (2015)	Pruned speech rate (words per minute)
Hsu (2017)	Speech rate (pruned syllables per minute)

Kim & Tracy-Ventura (2013)	Speech rate (syllables per minute) Frequency of reformulations
Lambert et al. (2017)	Pruned speech rate Frequency of filled pauses Ratio of end -clause pause to number of syllables Ratio of mid-clause pause to number of syllables Frequency of overt self-repairs
N. de Jong & Perfetti (2011)	mean length of fluent run phonation time ratio mean length of pause articulation rate
Nation (1989)	Speech rate (words per minute) Number of repairs per minute
Pinter (2005)	Amount of silence (per dyad) Words per minute (per dyad)
Sample & Michel (2014)	Filled pauses per minute Words per minute (per pair) Time to complete the task (per pair)
Thai and Boers(2016)	Speech rate (words and syllables per minute)
Van de Guchte et al. (2015)	Speech rate Articulation rate Number of silent pauses
Wang (2014)	Speech rate (words per minute) Mean length of end-AS pauses* Mean length of mid-AS pauses* Number of reformulations

In addition to the observation that different studies have operationally defined fluency in different ways, what also emerges from reviewing the fluency measures used in these studies is that the majority have adopted only one or two measures and that these measures tend to focus on what might be termed 'global' or 'composite' fluency in that they incorporate speed, breakdown and repair aspects of fluency. There are few studies which have attempted to hone in on speed, breakdown and repair in isolation. Furthermore, all studies have favoured 'objective' means of capturing fluency with none adopting a fluency measure based on rater scoring, for example.

In the interests of ensuring this discussion of literature is relevant to the present study's *fluency* focus, what follows is a synthesis of the literature which has looked at TR and PR effects on short-term fluency and/or fluency transfer. Although, occasional references will be made to, for example, the methodology which has been used in TR studies with alternative foci, these eighteen studies (Table 10) will form the empirical base for this literature review. I will often refer to them as the 'core studies'.

5.5 Key questions in TR/fluency research

Broadly speaking, these core studies (Table 10) can be divided into three main areas of enquiry: 1) the impact that repeating a task has on fluency during the intervention (short-term fluency) 2) the impact that repeating a task has on fluency on a new task (fluency transfer) and 3) the attitudes of students and/or teachers towards TR and/or PR. Table 11 shows that while some studies have only sought to explore one of these areas, others have looked at two within a single study. To the best of my knowledge, no study has looked at all three within a single study. These three broad questions will be used as a framework for the literature review.

Table 11: Core studies and research focus

Study	**Type of repetition (TR/PR/Both)**	**Short-term fluency (on same task/procedure)**	**Fluency transfer (on different task)**	**Learner/Teacher perceptions**
Ahmadian & Tavakoli (2010)	TR with time pressure	✓	×	×
Ahmadian (2011)	TR	×	✓	×
Ahmadian et al. (2017)	TR	×	×	✓
Arevart & Nation (1991)	TR with time pressure	✓	×	×
Boers (2014)	TR (two conditions: +time pressure and –time pressure)	✓	×	×
Bygate (1996)	TR	✓	×	×
Bygate (2001)	BOTH	✓ (only for TR)	✓ (only for PR)	×
Fukuta (2016)	TR	✓	×	×
Hsu (2017)	TR	×	✓	×
Kim & Tracy-Ventura (2013)	BOTH	×	✓	✓ (in Kim, 2013)
Lambert et al. (2017)	TR	✓	×	✓
N. de Jong & Perfetti (2011)	BOTH with time pressure	✓	✓	×
Nation (1989)	TR with time pressure	✓	×	×

Pinter (2005)	PR	✓	×	✓(in Pinter, 2007)
Sample & Michel (2014)	PR	✓	×	×
Thai and Boers(2016)	TR (two conditions: +time pressure and –time pressure)	✓	×	×
Van Guchte et al. (2015)	PR	×	✓	×
Wang (2014)	TR	✓	×	×

5.5.1 Short-term fluency effects of task and procedural repetition

Overview

In almost all studies which have investigated TR effects on short-term fluency researchers have observed an increase (Wang, 2014; Boers, 2014; Thai & Boers, 2016; Lambert et al., 2017; Bygate, 1996; 2001). Of the thirteen[8] studies that have investigated short-term fluency effects for TR, the only study which found no fluency effect was Fukuta (2016). In other words, the finding that TR has an impact on additional enactments *of the same task* seems to be relatively robust to different contexts, participants, repetition conditions, task type, and proficiency level. As well as fluency, many studies have suggested that accuracy and/or complexity also increase during TR (e.g. Bygate, 2001; Gass et al., 1999; Lynch and Maclean, 2000; 2001; Wang, 2014;).

Although they are far fewer in number, studies which have investigated short-term fluency effects of *procedural* repetition have similarly had consistently positive findings with regards to fluency (N. de Jong & Perfetti, 2011; Pinter, 2005; Sample & Michel, 2014). Although there is growing academic interest in the pedagogic usefulness of PR, to date its effect has been

[8] Some of the core studies focus only on the long-term effects of TR and are therefore not included

gauged primarily on other dependent variables such as LREs (e.g. Kim, 2013; Kim & Payant, 2014; Payant & Reagan, 2016).

In what follows, I first look at key studies which have investigated the impact of task repetition on fluency. This includes those studies which consider 'inherently repetitive' tasks.

Exact Task Repetition and short-term fluency

It is Bygate (1996) who is most often cited as being the first to explore task repetition effects on L2 speech performance (including fluency). Indeed, his research does focus solely on task repetition as opposed to repetition and time pressure (as in Nation, 1989 and Arevart and Nation, 1991) and considers repetition of the exact same task as opposed to procedural repetition (as in Plough & Gass, 1993). In his exploratory study, a single language learner watched and then narrated a video clip on *two* occasions, *three* days apart. The video used in the study was a short extract from a Tom and Jerry cartoon. In terms of analysis, Bygate selected a wide range of features which were grouped into those that he argued revealed something about 'repertoire', 'accuracy' and 'fluency'. In terms of fluency, the measures he selected were 'verbatim' repetitions and 'substitutive' repetitions. He found that the repeated performance was characterised by a decrease in verbatim repetitions (where the exact same phrase or word was repeated) and a slight increase in substitutive repetitions (where the utterance was reformulated). There were also increases observed for complexity measures. Overall, then, the results led Bygate to assume that "performance of a task on a second occasion may well be better than the first" (p. 144). The study is limited in the fact that there was only a single case and there was no control. In other words, it is impossible to say exactly what the effect of *repetition* was over and above the effect of time spent studying. Nevertheless, research interest into task repetition was piqued by this study.

In a larger-scale and more complex study, Bygate (2001) returned to the subject of task repetition effects. This time, 48 ESL students were recruited to take part in the study. Participants performed 2 different tasks at Time 1 (a narrative and an interview task) and then repeated the same tasks at time 5 (10 weeks later) along with two different tasks of the same type. In order to establish the effects of task repetition, he compared all participants' performances at time 5 on the repeated task with performances of the new task (i.e. this was a cross-sectional study and only compared performances at time 5). He found that fluency (unfilled pauses per *t*-unit) and complexity (words per *t*-unit) were significantly higher on the repeated versions of the task than completely new tasks of the same type. He was confident that "this [could] only be an effect of highly contextualised cognitive rehearsal, releasing spare capacity on the part of the speaker to increase fluency or complexity" (Bygate, 2001, p. 42). However, the tasks in Bygate's (2001) study are not counterbalanced, meaning that the difference in fluency and complexity found between repeated and new versions of the task could be due to differences in the content of the tasks (i.e. more conceptually difficult, differing lexical demands). Another issue is that only a single measure of *breakdown* fluency is considered and I have already established (Chapter 2) that this may not be the most indicative of cognitive fluency.

Wang, 2014, similarly found that fluency (speech rate, frequency of end-clause pauses) increased during task repetition. In her study, repetition was defined as watching and *simultaneously* narrating a silent movie and then watching and simultaneously narrating the same movie a second time. There were 13 participants in the study who were Chinese speaking EFL students. The findings suggested that repeating a task resulted in improved complexity (mean length of AS-unit; subordination), accuracy (error-free clauses; error-free clause rate) and fluency (speech rate; end-clause pauses) on the subsequent retelling. She argues that an explanation for increased accuracy in the repeated performance is due to having the opportunity to monitor the output which is not provided by other types of planning. However, the nature of the task (simultaneous watch and narrate) is highly specific. Presumably, speed of speech will be restricted by the speed at which events unfold in the video and this may limit its generalisability to other tasks. While speech rate increased and the frequency of end-clause pauses decreased on the second re-enactment, Wang (2014) did not find any change in mid-clause pauses. Due to the collinearity between speech rate and amount of pausing, and the fact that no pure measure of 'speed' (e.g. articulation rate) was used we do not know, whether the increase in speech rate is *only* due to the reduction in end-clause pause frequency. End-clause pause frequency has been argued to be related only to content planning and is not indicative of L2 fluency (N.H. de Jong, 2016; Hilton, 2014; Lambert et al., 2017). Nonetheless, Wang's finding that complexity and accuracy increased during TR and with large effect sizes, does point to an interesting *holistic* performance effect for TR.

In a recent and very relevant study, Lambert et al. (2017) found that immediate ecological TR resulted in increased fluency, irrespective of task type or proficiency level of the learner. Their study is particularly relevant for the current research because they investigated immediate ecological repetition and they used a range of fluency measures that could tap different aspects of fluency. As I explained in Chapter 2, fluency is a multi-dimensional construct and if the aim of a study is to shed light on underlying cognitive processes, then a clearer distinction must be made between different aspects of fluency. In this study, 32 Japanese L1 learners of English at three different levels of proficiency were recruited. They performed three[9] tasks in quick succession. The tasks were not interactive and the participants were given the role of either "speaker" or "listener". They alternated roles until all participants had both listened and spoke on the three tasks. They then formed new dyads and repeated the procedure. This was repeated until all participants had spoken and listened about all of the tasks six times.

In order to analyse the speech which was produced they used a number of measures which represented overall (speech rate), breakdown (number of filled end-clause and mid-clause pauses) and repair (number of self-corrections) aspects of fluency (there was no pure speed measure and breakdown measures were limited to 'noticeable' filled pauses). In terms of speech rate, they found that repetition continued to have an effect over 5 performances of the same task but that the effect size was larger across the first three performances. They also found that filled pauses (ums and ers) in the end-clause position decreased considerably between Time 1 and Time 2 but there was no further reduction. The fact that this study looked only at filled pauses could be a possible explanation as it could be that filled pausing reached a 'floor' at the second performance. They notice that filled pauses in the mid-clause position continued to decrease until the fourth performance. Their repair measure: 'overt self-corrections' showed greater variation. They found that overt self-repairs remained relatively stable in early performances of the task and then gave way to greater variation after the fourth performance. They suggest that repair fluency may be related to some extent to proficiency level (or performance level). This adds support to the theory that different areas of performance benefit from differing numbers of repetition. Another possibility is that there is a priority hierarchy for performance gains through TR with certain areas of fluency (e.g. pausing) taking priority over others (e.g. repair). While this study sheds light on certain areas of fluency, others (e.g. speed) are not catered for. The study is also limited in that the time

[9] There was also a dialogic task which was carried out but this did not form part of the study

allotted for the learners' monologues changed between performance two and three. This may have had an impact on learners' orientation to the task and consequently their fluency.

In contrast to the findings of Bygate (1996, 2001), Wang (2014) and Lambert et al. (2017), Fukuta (2016) found that fluency (speech rate) did *not* increase during a second enactment of the task. In this study, 28 EFL students were split between an experimental (TR) and a control group. The TR group performed identical picture narrative tasks on two occasions, one week apart. The control group performed two different tasks over the same time period. Fluency, as measured by pruned words per minute, was not higher than the control group at time two. The researcher offered that one repetition may have been insufficient to have any impact on the fluency of the participants' speech. An alternative explanation is that the participants selected for this study may have found the task easy and were therefore operating at already maximally fluent levels (Fukuta, 2016). It is worth noting, however, that accuracy and lexical variety did increase for the TR group in this study. Drawing on 'protocol analysis' (stimulated recall interviews), the author observed that "participants in the experimental group focused more attention on syntax and less on conceptual aspects at the second enactment. This provides further support for Wang's (2014) finding that task repetition has form-focused effects. In other words, even though speech rate (i.e. *utterance* fluency) did not seem to increase during TR in this study, it does seem that cognitive processes underlying speech production were, in fact, more fluid, giving rise to increased attentional capacity to focus on form which, in turn, resulted in higher accuracy and complexity.

As explained in the previous chapter, language pedagogy has developed ways of providing opportunities to repeat (e.g. mingles, 'fairs', 'exhibitions' etc). These task sequences are inherently repetitive in that the task sequence necessitates a certain degree of repetition. Two particular examples were provided above: (1) the Poster Carousel (Lynch and Maclean, 1994) and (2) the Fluency Workshop (4-3-2) (Maurice, 1983). Both of these task sequences have been the subject of empirical research to investigate their performance-enhancing credentials. The poster carousel was investigated by Lynch and Maclean (2000; 2001) and the 4-3-2 technique has been investigated in a string of studies (Nation, 1989; Arevart & Nation, 1991; N. de Jong & Perfetti, 2011; de Jong, 2012; Boers, 2014; Thai & Boers, 2016)

Lynch and Maclean (2000; 2001) report on research that was conducted into the effectiveness of the Poster Carousel for heightening performance (in terms of a number of different features). In the study, fourteen learners took part in the Poster Carousel task sequence as described in Figure 9. Their qualitative analysis of the performances of 5 of these learners is presented in Lynch and Maclean (2000; 2001). Lynch and Maclean (2000) presents a case-study analysis of the performances of two students: Alicia, an upper-intermediate level student and Daniela, a low-intermediate level student while Lynch and Maclean (2001) presents a further three cases. In general, Lynch and Maclean (2000; 2001) show how the quality of the talk from all five participants improved over the six performances. It is difficult to say precisely in which ways the learners increased their fluency because fluency is not focused on specifically. Instead, the researchers provide a more holistic and individualised account of performance development in these learners. In general the note that performance was 'improved' on subsequent performances for all of the learners and that particular evidence for this was improvements in "syntactic accuracy" (2001, p. 148) "more precise and less awkward" performance (2001, p. 142), "greater precision in choice of words" (2001, p. 144) "less vagueness" (2001, p. 151) and "correct fluent use of some language forms after initial difficulty" (2001, p. 144). In other words, and as this range of observations attests, although CAF was not measured in a quantitative sense, it was still the basis for a lot of their observations. A particular finding was that the learners seemed to draw on cues and input provided by their interlocutor to improve their production on subsequent task performances.

They did not find that proficiency level affected performance gains but they do suggest that the two students used the task (and the repetition it involves) in different ways. Additionally, they note that there were differences in how the two students *perceived* they gained from the task. The higher level student was able to articulate how she had improved over the course of the session whereas the lower level student did not feel she had made any progress in her speaking. The researchers suggest that this is due to the fact that the lower level learner had to focus all her attention on getting the task done and was therefore not able to monitor her output and perceive the changes. They suggest that teachers have an important role to play in helping lower level learners 'see' the improvements that they make in such task sequences, perhaps by asking them to listen to their performance.

This study is described as "unique in the field" (Bygate & Samuda, 2008) because it looks at TR in the context of an authentic language classroom and with an inherently repetitive task. In other words, it investigates task repetition in a "pedagogically convincing" way. However, it fails to investigate the task sequence at a quantitative level which means that researchers may be unconvinced by its generalisability and validity and unable to compare their findings with their own.

Other studies which have investigated inherently repetitive tasks are those that have looked at Maurice's (1983) 4-3-2 technique (Nation, 1989; Arevart & Nation, 1991; N. de Jong & Perfetti, 2011; de Jong, 2012; Boers, 2014; Thai & Boers, 2016). This task sequence combines repetition and increasing time pressure as learners are required to perform a task (an oral presentation) in four minutes then three minutes and finally in two minutes. The 4-3-2 is therefore similar in many ways to the Poster Carousel (Lynch & Maclean, 1994), but it involves additional time pressure which can be interpreted as manipulating the amount of online planning time that the learners are given (Ellis, 2005). Maurice (1994) explains that "most speakers tend to pause a lot and use many fillers when they speak, especially when talking about a topic for the first time. As we speak more about a topic, these pauses…tend to decrease" (1994, p. 54) He further argues that the shrinking time frames push students to speak "more fluently and naturally toward the end of the activity" (1994, p. 54). The findings of the 4-3-2 studies have produced consistent results for learners' oral performance, especially for fluency.

First, Nation (1989), who was intrigued by the 4-3-2 technique, set out to examine the exact nature of its impact on oral performance. His subjects were 6 advanced adult learners of English. Participants were required to perform a speech on a topic which interested them first for 4 minutes, then 3 minutes and finally for 2 minutes. He measured the learners' performance on all three tasks in terms of fluency, accuracy and 'control of content'. This final dimension allowed Nation to explore "the ways in which speakers reduced the material in each section of their talk to fit the reduction in time available to deliver the talk" (Nation, 1989, p. 379). Fluency was operationalised as words per minute as well as the number of hesitations, repetitions and false-starts per 100 words. Accuracy was defined as errors per 100 words. Nation found that speech rate increased between performance one and performance 3 for all but one participant. The number of hesitations etc. decreased for each participant for each performance. He took these results as convincing support for the effectiveness of the 4-3-2 task sequence for pushing learners to increase their fluency. Results for accuracy, however, were less convincing with learners only showing small improvements in syntactic accuracy. In terms of their control of the content, Nation found that the participants omitted considerable amounts of (non-important) information from subsequent performances and, at times, embedded clauses within longer syntactic units. However, the study was small in scale and drew only on descriptive statistics.

In a follow-up study, Arevart and Nation (1991) further explored this "technique [which] allowed learners to perform at a higher than…normal level of fluency" (1991, p. 84). They suggested that it was the very particular nature of the task sequence which brought about this change in performance and specifically three main features: Firstly, the fact that the interlocutor changes for each performance meaning that there is fresh need to communicate each time and means that "the speaker's focus continues to remain on the message" (1991, p. 84) and that they don't feel obliged to add new details to the talk to keep it interesting for the listener. Secondly, they identify the inherent repetition as a key feature of the task. They note: "the repetition of the talk…has a major effect on fluency because it increases the speaker's familiarity with both the form and content of the material and thus increases the speed with which a speaker can access…forms" (Arevart & Nation, 1991, p. 84).

Thirdly, they note that the added time pressure of the 4-3-2 task sequence performs two 'fluency-enhancing' roles. Firstly, it encourages the learner to speak quicker (i.e. to explicitly focus on increasing speed of speech) and secondly, as available time decreases, it means that the speaker does not need to fill any extra time with new talk. Their results were, once again, very supportive of the 4-3-2 technique, particularly for fluency development.

The addition of time pressure might encourage learners to actively focus on increasing speed as they are aware that they have to complete the task in a shorter space of time.

> "Under time pressure…speakers may choose to rely on already automatic processes, e.g., using highly frequent vocabulary and simple syntactic constructions." (De Jong, 2012, p. 51)

This might come at the expense of complexity and accuracy (and perhaps other aspects of fluency). Indeed, Nation (1989) suggests that this is a key feature of the 4-3-2 task sequence: that learners are unable to add any new information, which might otherwise result in error. Instead they are driven to shorten and simplify their message, relying on previously used language. A criticism of the early 4-3-2 studies would be that they fail to explore the differential effects of the two defining elements of the 4-3-2 – task repetition and time pressure – on performance.

The only studies which have attempted to disentangle repetition and time pressure within a single study are de Jong (2012), Boers (2014) and Thai & Boers (2016). These three studies have very similar findings. Firstly, they agree that TR combined with time pressure results in higher fluency. Secondly, time pressure is associated with lower levels of accuracy and complexity when compared to the constant time condition. Repetition with constant time, on the other hand seems to also benefit fluency (although to a lesser degree) but also points to some improvement in either accuracy, fluency or both. Thai and Boers (2016) conclude that the 4-3-2 might not be the best fit for language teaching because it sacrifices accuracy and complexity. Instead they suggest that straight repetition might be more beneficial. The authors refer to the 4-3-2 technique as a "double-edged sword" because, while it may enhance fluency and potentially encourage learners to proceduralise knowledge, it may be that they proceduralise incorrect forms (see also Foster, 2001; Lennon, 2001).

Table 12: The 4-3-2 studies and impact on performance

	TR + Time pressure		**TR**	
	Fluency	Accuracy and Complexity	Fluency	Accuracy and Complexity
De Jong (2012)	Higher fluency	No impact on complexity. Accuracy = N/A	Higher fluency	Higher complexity. Accuracy = N/A
Boers (2014)	Significant speech rate gains esp. between second and third performance. No real impact for hesitations.	Lower accuracy. No change to complexity	Increased fluency	Increased Accuracy and complexity

Thai & Boers (2016)	Phonation time ratio increased between t1 and t2 and from t2 and t3 with large effect sizes; AR decreased between T1 and t2 only; number of reformulated words between t2 and t3	No change to complexity and accuracy	Increased fluency	Increased accuracy and complexity

Overall, then, it seems that exact task repetition, and in particular *immediate* task repetition, results in increased fluency, complexity and accuracy on subsequent performances. Some aspects of fluency are only affected after certain numbers of repetitions (e.g. repair, mid-clause pausing) and repetition *without* additional time pressure is probably most conducive to increasing all three aspects of the CAF triad simultaneously.

Procedural repetition

Fluency effects have also been observed in the short-term for procedural repetition although there is a paucity of research in this particular area. To the best of my knowledge, only three studies have investigated short-term fluency effects during PR.

Firstly, Pinter (2005) investigated PR effects on the oral performance of young Hungarian learners of English. In this study, learners engaged in a spot-the-difference and an instruction task. The twenty learners in her study repeated *similar* versions of each of these tasks twice over the course of three weeks. Each performance was with the same partner. Pinter explains that “one of the most noticeable changes of the performances was that children increased their *pace* on the tasks” (Pinter, 2005, p. 117 my emphasis). She goes on to explain that for one pair, the speech rate doubled. However, she gives no more indication of fluency development and the data is not analysed quantitatively.

A study which was similar in design but which *did* provide more detail in terms of fluency effects is Sample and Michel (2014). Although this, too, was a small, exploratory study, the researchers set out to understand what happens to the variables of complexity, accuracy and fluency over the course of three performances of the same type of task (procedural repetition). They conducted their research on 6 ESL learners who were 9-10 years old. Participants formed pairs and performed a spot the difference task three times over a period of three weeks. The researchers analysed the performances for complexity, accuracy and fluency. Their findings showed that, for this group of students, fluency increased on the second performance of the task but three performances were needed to give rise to gains in all three areas. They conclude “initial performances that benefit in one area come at the expense of another; by the third performance, however, trade-off effects disappear” and “with growing task familiarity, students are able to focus their attention on all three CAF dimensions simultaneously” (Sample & Michel, 2014). The highly interactive nature of the tasks in this study mean that fluency is calculated on a ‘per pair’ basis meaning that it is not possible to make any claims about the impact on individual fluency. It is worth noting, also, that both of these studies were conducted with young learners.

With adult learners, N. de Jong and Perfetti (2011) looked at the effects of two versions of the 4-3-2 task sequence: one which required participants to repeat the exact same task and another which required participants to perform slightly different tasks. While the TR group responded to the same question three times (e.g. *what do you think about pets?*), those in the PR group responded to three different questions. They found that fluency (mean length of run) increased for the PR group as well as the TR group although they note that it was in general more variable for the PR group.

In terms of PR effects on fluency, then, there is very little empirical research to draw on, and that which there is, is either exploratory (Pinter, 2005; Sample & Michel, 2014) or only looks at PR combined with time pressure (N. de Jong & Perfetti, 2011).

5.5.2 Fluency transfer effects of task and procedural repetition

Most TR studies only consider TR/PR effects on short-term fluency (Wang, 2014; Sample & Michel, 2014; Lambert et al., 2017; de Jong, 2012; Boers, 2014; Thai & Boers, 2016; Nation, 1989; Arevart & Nation, 1991; Fukuta, 2015; Kim, 2013). However, while increased performance in the short term might be expected due to reductions in conceptualisation and formulation time *on a particular task*, it doesn't, on its own, provide evidence that TR fosters SLA (Ellis, 2005; 2009). Ellis (2009) suggests that if any strong avocations are to be made for TR, there needs to be found some kind of long-term, enduring effect on performance. Fluency transfer effects for TR have typically been gauged by the extent to which performance *on a different task* is affected. Only seven of the core studies have sought to investigate the impact of TR and/or PR on long-term L2 fluency (Ahmadian, 2011; Bygate, 2001; N. de Jong & Perfetti, 2011; Hsu, 2017; Kim & Tracy-Ventura, 2013; Patanasorn, 2010; Van de Guchte et al., 2015)

These studies have attempted to gauge fluency transfer effects by analysing pre-test and post-test data either in addition to online task performance data (N. de Jong & Perfetti, 2011) or in place of online performance data[10] (Kim & Tracey-Ventura, 2013; Patanasorn, 2010). Arguably, the most interesting studies are those which show us what happened both during TR itself *and* on a different task as this allows us to 'peer into the box' to determine what the connection might be between the intervention and the long-term findings. In other words, *were there any improvements to transfer in the first place?* To my knowledge, only one study has combined online performance data during TR (albeit combined with time pressure) with a test which establishes if TR effects transfer to a new task (N. de Jong & Perfetti, 2011) and, as explained earlier, this study focussed solely on the 4-3-2 activity.

Task repetition

Only two studies looked only at the long-term impact of TR on fluency. These two studies (Ahmadian, 2011; Hsu, 2017) are discussed below.

Table 13: Task repetition and fluency transfer

	Online gains for TR?	**Increased fluency on new task?**
Ahmadian (2011)	N/A	Yes
Hsu (2017)	N/A	No

[10] Kim (2013) is a sister study for Kim & Tracy-Ventura. It *does* analyse online task performance data but not in terms of fluency.

Very recently, Hsu (2017) found no fluency transfer effect for TR. In her study, 39 Taiwanese university students who were studying English as a foreign language were randomly assigned to three groups, two experimental and one control. One of the experimental groups repeated a narrative (picture description) task once (TR group), while the other repeated the task and the participants listened to, transcribed and corrected[11] their speech on the first performance before the repeated task (TRPT group). The performances on the repeated task were compared between the two groups and no differences were found for fluency. However, the choice of cross-sectional (between-subjects) analysis means that we do not know whether there was an effect for repetition on TR and TRPT, only that the gains were not greater in either group. This study also looked for the effect of TR and TRPT on fluency transfer and accuracy. This was assessed by way of analysing performance *on a new task* for all three groups (inc. control). Due to the presence of a control group, coupled with the researcher's finding that one-way ANOVAs conducted on the 'pre-tests' revealed no significant difference between groups, Hsu argues that the lack of a significant difference in fluency between groups on the performance of the new task reveals that both TR and TRPT did not impact on fluency transfer. She did, however, find that accuracy was higher for the TRPT group. Hsu (2017) explains that likely reasons for the lack of a fluency transfer effect for TR are the 'boredom factor' and that possibly one repetition of the task was not sufficient to increase fluency in the long-term. She cites Ahmadian's (2011) study as evidence that a higher number of repetitions may be needed to drive fluency transfer.

Ahmadian (2011), on the other hand, did find that TR had an impact on fluency transfer. He investigated two intact classes of EFL students in an Iranian university (n=30). One class was assigned to the experimental condition and the other to the control condition. The participants of both groups performed a narrative task at Time 1 (watching short silent film then retelling the story) then a personal opinion task at time 12. The experimental group repeated the exact same narrative task a further 10 times between Time 1 and Time 12, with each performance coming 2 weeks after the last. Statistical analysis using a between subjects design (t-tests between groups at Time 1 and time 12) found that there were significant

[11] Corrections were not overseen by a teacher. Some mistakes were corrected while others were not. Some mistakes were 'corrected' incorrectly.

differences between the groups at time 12 in terms of fluency (pruned and unpruned speech rate). There were no statistical differences at time 1.

Ahmadian takes this finding as support for the theory that intensive repetition (11 performances) results in fluency transfer (and complexity) gains (e.g. Bygate, 2001). However, there are a number of methodological concerns about this study. Firstly, the mean speech rates which are reported for the groups are *very* low (e.g. 26.7 syllables a minute) with very low standard deviation. This suggests that the data probably featured extremely long pauses which were not discounted from analysis. A second concern, and one which is mentioned in the paper, is that the study took place over six months and although the researcher explains that the curriculum and teacher were the same for both classes, there is certainly potential for additional influences to have exerted an effect.

Procedural repetition

Two studies have investigated the long-term effects of PR on fluency (Table 14). These two studies (Bygate, 2001; van de Guchte et al., 2015) are explored in more detail below

Table 14: PR studies and fluency transfer

	Online gains for PR?	**Increased fluency on new task?**
Bygate (2001)	N/A	No
Van de Guchte et al. (2015)	N/A	No

To date, there has been no study which has found any fluency transfer effect for PR. Firstly, Bygate (2001) looked at the effect of, what he calls 'task type practice' on long-term CAF. Participants (n= 48) were tested during week 10 on their performance on a new task of the same type that they had practiced previously and also on a new task of a completely different type (Table 15). He was looking for what he called a 'task-type practice' effect (Bygate, 2001, p. 30).

Table 15: Design of Bygate's 2001 study

Group	Pre-test (week 1)	Treatment (week 3)	Treatment (week 5)	Treatment (week 7)	Post-test (week 10)
Narrative group	Nar. 1 Int. 1	Nar. 2 & 3	Nar. 3 & 4	Nar. 4 & 5	Nar. 1 &6 Int. 1 & 6
Interview group	Nar. 1 Int. 1	Int. 2 & 3	Int. 3 & 4	Int. 4 & 5	Nar. 1 & 6 Int. 1 & 6
Control group	Nar. 1 Int. 1	-	-	-	Nar. 1 & 6 Int. 1 & 6

He measured performance in terms of CAF and found that task-type practice had no any effect on performance of a completely different task type. He argues that "more – or more massed – task exposure might be needed" for any long-term effect to be seen (Bygate, 2001, p. 43). He also suggest that the participants may have already found the tasks quite easy meaning that task type practice (procedural repetition) did not have any effect because learners were already performing the task at high levels of fluency.

Although the original aim of Van de Guchte et al. (2015) was to look at the effect of TR on oral fluency (among other things), they ended up looking at PR because they found that during pilot studies, learners became disengaged when asked to repeat the exact same task. In their study, 48 L1 Dutch ninth-graders who were learning German as a foreign language were split between two conditions. One was an experimental condition which looked at the long-term impact of having repeated a task (there were two tasks, each targeting a different grammatical structure). They also had a control group. They found that there was no long-term effect for PR on oral fluency and they explain that this finding may be related to the fact that the repetition was not of exactly the same task and also that the repeated task was two weeks after the first.

Task repetition versus procedural repetition studies

Three studies have looked at the effects on fluency transfer of *both* task *and* procedural repetition (Table 16).

Table 16: TR and PR studies and fluency transfer

Study	Short-term fluency gains for TR	Short-term fluency gains for PR	Fluency transfer for TR	Fluency transfer for PR
Kim & Tracy-Ventura (2013)	N/A	N/A	No	No (but increase in grammatical complexity)
N. de Jong & Perfetti (2011)	Yes	Yes	Yes	No
Patanasorn (2010)	N/A	N/A	No	No (but increase in accurate use of simple past)

Firstly, Kim & Tracy-Ventura (2013) found that TR did not result in fluency transfer effects. They investigated the impact of task and procedural repetition on long term CAF gains. The participants were 36 Korean learners of English (age= 13-14 years old). The design of their study is illustrated in Table 17.

Table 17: Study design in Kim & Tracy-Ventura (2013)

	Day 1 (week 1)	Day 2	Day 3	Day 4	Day 5	Day 19 (week 4)
Task Repetition	Pretest	Hosting an American friend	Hosting an American friend	Hosting an American friend	Posttest 1	Posttest 2
Procedural Repetition	Pretest	Hosting an	Describing school events/activities	Discussing mayoral candidates	Posttest 1	Posttest 2

	American friend	

Participants were assigned to either a task repetition or procedural repetition group (no control). The task repetition group repeated the same interactive task on three occasions over three days while the procedural repetition group carried out three different tasks with a similar procedure. All participants performed monologic narrative tasks during scheduled class time (Kim, 2013). They also carried out pre- post- and delayed post-tests. In Kim & Tracy-Ventura (2013), only this pre- and post-test data Is reported on. In terms of the impact of TR on fluency transfer, Kim & Tracy-Ventura (2013) found that there was no significant effect for either time or condition on the participants' fluency at the posttest. An interesting finding, however was that that for the PR group there was a significant interaction between time and condition for syntactic complexity (clauses per AS-unit). Clauses per AS-unit was significantly higher at the post-test for the procedural repetition group than the TR group.

They suggest that a possible explanation for the lack of a transfer effect is the fact that they used interactive tasks for the treatment and monologic tasks for the pre- and post-tests. Segalowitz (2010) points out that there must be transfer appropriate processing involved if proceduralisation is to take place. In other words, learners must engage in the same cognitive processes in class that they need to engage in at a later point. They explain the increase in complexity for the PR group because "is possible that carrying out tasks with the same procedure but different content encouraged learners to produce more diverse clause types, which in turn promoted learners' production of more clauses per AS-unit as well as more complex AS-units [on the post-test]" (Kim & Tracy-Ventura, 2013, p. 838)

N. de Jong and Perfetti (2011) compared the impact of repeating same or similar tasks *with increasing time pressure* and they found that the fluency benefits *did* transfer to fluency transfer but only for the group of students that had repeated the exact same task (i.e. the TR group). 24 language learners were recruited to take part in a fluency training course. The course ran over two weeks, during which time two experimental groups took part in three fluency workshops (Table 18). In one of these groups, the workshops involved repeating monologic tasks with a computer with increasing time pressure (4-3-2). In the other experimental group, the students performed similar tasks which were not identical (procedural repetition). The tasks involved talking about a particular topic (e.g. pets; sports) without stopping. The researchers also had a control group which received no training. The researchers found that both task repetition and procedural repetition had a positive impact on fluency within the training workshops and that for the exact TR group, this carried over to a new task. The researchers explain this finding in terms of the proceduralisation of language which has been facilitated by repeated practice.

This is an important study which is particularly relevant to the present research. It is important because it has suggested that there may be some long-term benefits to task repetition. This finding could have huge implications for language teaching, in particular. However, there are a number of issues that need to be addressed. Firstly, this study was concerned with Maurice's (1983) 4-3-2 technique and therefore the TR was conflated with time pressure. It is impossible to say whether the findings of this study are the result of the TR, the time pressure or both. No other study before or since has looked at the transfer potential of the 4-3-2. Other recent 4-3-2 studies such as Boers (2014) and Thai and Boers (2016) have looked only at online task performance data. An additional idiosyncrasy of this study is that it involved a 'programme' of TR as opposed to a single session. Participants took part in three separate 4-3-2 workshops over the course of the training. One final point about this study was that the students received frequent and extensive feedback about their performances throughout the training. Elsewhere in the literature it has been suggested that this might be key for TR effects to transfer (Sheppard 2006; Hsu, 2017). Lynch and Maclean (2000) also suggest that there is:

> "a need for teachers to follow up task-based practice with 'noticing' activities, so that we can help learners consolidate for the longer term what may otherwise be fragile changes in their interlanguage" (Lynch & Maclean, 2000, p. 245)

The fact that there were so many additional factors involved in this study means that it is difficult to ascertain what impact repetition alone may have had on fluency transfer.

Table 18: Design of N. de Jong & Perfetti (2011)

Group	**Pre-test**	**Session 1**	**Session 2**	**Session 3**	**Post-test**				**Delayed post-test**
Repetition I	a	1,1,1	4,4,4	7,7,7	B	-	-	-	c
No Repetition (procedural repetition)	a	1,2,3	4,5,6	7,8,9	B	-	-	-	c
Control/Rep II	a	-	-	-	B	1,1,1	4,4,4	7,7,7	c

In his (unpublished) doctoral research, Patanasorn (2010) investigated the effects of TR and PR on fluency transfer and accuracy. EFL learners were divided among three treatment groups (no control): task repetition, procedural repetition and content repetition. The task repetition group took part in three identical decision-making tasks while the procedural repetition group took part in three *different* decision-making tasks. Similarly to Kim and Tracy-Ventura (2013) their fluency on these performances was not analysed. Instead, the focus of the study was on long-term effects of the intervention which was gauged by way of pre-, post-, and delayed post-tests. The pre- and post-tests were film retell tasks. Patanasorn (2010) found no increase in fluency between pre- and post-test for either the TR or PR groups although he did find that the PR group increased in past tense accuracy. He explains that this is possibly due to exposure and practice of a wider range of past tense verb types because "during treatment sessions, the procedural repetition group used more verb types than [the other groups]" (Patanasorn, 2010). Reasons for the lack of fluency gains for the TR group are explained in terms of a lack of engagement in the tasks because learners may have been bored by repetition.

Overall, then, it is clear that far fewer studies have investigated fluency transfer effects of TR and PR. Those that have, tended to focus *solely* on long-term effects meaning that we cannot make links between changes (or lack of changes) in fluency during the intervention itself and any long-term findings. The research which is available is contradictory. While some have found long-term gains for fluency as a result of TR (N. de Jong & Perfetti, 2011; Ahmadian, 2011) certain methodological issues make it difficult to be entirely convinced either way. Furthermore, these two studies were conducted in laboratory environments and with tasks which were specifically designed for research purposes (i.e. performing monologues with a computer). It is therefore difficult to ascribe any pedagogic significance to these findings because the studies were carried out in environments that are unlike a working language classroom (see also Chapter 7) and did not consider learners' and teachers' attitudes to the intervention.

5.5.3 Learners' and teachers' attitudes to task and procedural repetition

I move, now, away from looking at the impact of TR on performance in order to focus on learners' and teachers' *attitudes* to TR and PR. Studies which have attempted to understand learners' views about TR have been motivated by the widely discussed assumption that TR may be 'boring' for learners. This 'boredom factor', as I will call it, has motivated many researchers to elicit learner opinion about TR, presumably because one of the many aims of studies into SLA is to be able to make sound pedagogic recommendations and "a task can only be said to have worked if the students have found it enjoyable and/or useful" (Ellis, 1997, p. 39). It has also been suggested that, whatever the potential benefits of TR, teachers will be unlikely to use TR if their students find it boring (Kim, 2013; Ahmadian et al., 2017). Indeed, Nassaji (2012) found that keeping learners engaged and motivated was a primary concern (see also Burns, 1992; Samuda & Bygate, 2008). In addition to such practical concerns, this research also has a theoretical dimension, as the 'boredom factor' has been suggested as a possible explanation for the lack of fluency transfer effect which is commonly observed in TR studies (e.g. Gass et al., 1999; Kim and Tracy-Ventura, 2013; Hsu, 2017).

Table 19: Studies which have explored learner attitudes towards TR/PR

	Type of repetition	**How elicited?**	**Learner attitudes**
Ahmadian et al. (2017)	Delayed task repetition	Semi-structured interviews	Generally positive
Kim (2013)	Task repetition and procedural repetition	Questionnaire	Learners respond more favourably to procedural

			repetition than task repetition
Lambert et al. (2017)	Immediate task repetition (ecological)	Questionnaire	Generally good
Lynch & Maclean (2000; 2001)	Immediate task repetition (ecological)	Questionnaire	Generally good but higher levels saw benefit more than lower levels
Payant & Reagan (2016)	Task repetition and procedural repetition	Interview	Positive towards both types of repetition
Pinter (2007)	Procedural repetition	Interviews	Generally positive
Plough and Gass (1993)	Delayed procedural repetition	Interpretation of results (lower incidence of interruption)	Repetition results in 'disengaged' students
Takimoto (2012)	Task repetition and procedural repetition	Questionnaire	Generally good
Van de Guchte et al. (2015)	Immediate task repetition	Pilot study	Students found immediate task repetition 'boring'

Many TR studies refer to Plough and Gass (1993) as providing evidence that learners find TR "boring" (e.g. Takimoto, 2012). The study reported in Plough and Gass (1993) involved two groups of students, one of which was 'familiar' with the task type in that they had carried out similar tasks in their class previously. The other was completely new to the school and had therefore not carried out those sort of tasks in class because they had not yet attended *any classes.* The researchers found that, overall, there was more negotiation of meaning in the 'familiar' group but they also point out that the 'unfamiliar' pairs had a higher tendency to interrupt each other. They suggest that a tendency to interrupt one's partner is linked to task involvement: "unfamiliar pairs were more interested in the task itself which is exhibited by more instances of interruptions" (1993, p. 49) and that an unfamiliar task meant "greater conversational commitment or involvement" (1993, p. 50) while the 'familiar' students became "somewhat disinterested" (1993, p. 50). Plough and Gass (1993) offer the following two examples of the interaction between a dyad in the familiar and unfamiliar groups respectively.

A dyad from the 'familiar' group:

T: so…if we choose number 9 our work finished

M: hm-m

A dyad from the 'unfamiliar' group:

A: Good, OK. This is, I think this is perfect, this is great

M: We are creating a new civilization with scientists

A: Yea

M: And teachers

A: hm-m it's going to be great

There are a number of points which are of note here. Firstly, it appears that in these two extracts, the students are discussing different things. In the first they appear to be concentrating on the task in hand, whereas in the second the students are talking about the intervention in general (and are clearly very excited to be taking part). Secondly, and perhaps, crucially, a look at how the two groups of participants were recruited shows that the 'unfamiliar' group were at the very beginning of their studies. Presumably, some of the interruption behaviour demonstrated by these students could be attributed to their eagerness and enthusiasm. What I am suggesting, therefore, is that the eagerness to interrupt might be an attempt on the part of the student to demonstrate his or her English skills and to show that they are keen and willing students, in contrast with the students who have been studying at the school for some time and no longer feel the need to present their 'best' self. A final limitation of this study is that they seem to interpret similar findings in varying ways.

They suggest that higher incidence of interruptions in the unfamiliar group is due to greater involvement with the task but that higher incidence of overlaps (which arguably perform a similar function) in the familiar group is due to an “eagerness for task completion” (p. 51) because the participants had “little interest in the task”. I would argue that this study alone does not provide compelling evidence for the claim that TR is therefore ‘boring’ for students.

However, more recently, other studies have also reported that students found TR boring. For example, Van de Guchte et al. (2015) conducted a pilot study before their main investigation into the effects of task repetition in young learners. They note that “learners found it ‘boring’ to repeat the task and were not motivated to perform it” (2015, p. 2). Their way around this was to present the task at a later time (i.e. not immediately afterwards) and also to change the content of the task slightly. We are not given any indication of how this conclusion was arrived at.

In a recent exploration of task repetition effects on L2 fluency, Lambert et al. (2017) also elicited students’ views about a) the value of repeating the task, b) the optimal number of task repetition for a number of different task types and c) the value of repeating the tasks with different partners. Their study involved 32 students who repeated three different task types (instruction, narrative, opinion) on six occasions. The repetition was immediate and ecological. The learners’ opinions were elicited via a questionnaire which was administered after the intervention. The researchers report that their participants’ feelings towards TR were, on the whole, positive. They note that “nearly all” of the participants responded favourably to the opinion and narrative tasks and 75% of the participants in the instruction task found it beneficial. It is not clear, however, exactly how this information was elicited as no example of the questionnaire questions is provided. On the other hand, the majority of participants felt that six performances were unnecessary and that three or four performances would be adequate. The researchers also note that few participants mentioned that they became bored although, once again, we do not have information about exactly how this information was elicited.

It has been suggested that “using slightly different content for similar tasks [might] sustain students’ motivation and interest over multiple repetition” (Michel, 2017, p. 59) In a study which compared the two types of repetition, Kim (2013) investigated the differential impact of TR and PR on learner and teacher attitudes. She found that “learners who repeated the same task three times rated that kind of task repetition less interesting and participated in fewer language related episodes overall in comparison to learners who completed three versions of the same procedure but with different content" (2013). She also found that the teacher who was involved in the study displayed apprehension about using TR with a class of students.

Differences between student and teacher attitudes to TR

What is interesting is that some studies which have elicited learner feedback on TR have found that while learners may be happy to repeat the task, teachers may have an entirely different impression. I have already explained how the teacher in Kim's (2013) study was apprehensive about using TR. Ahmadian et al. (2017) set out to explore differences between learner and teacher perceptions of TR explicitly. In their study, 21 English language students at a private language centre in Iran carried out the same picture description task on two occasions with a one week interval. Following the second performance, all students took part in individual, semi-structured interviews with the researchers in which they were asked about their views on TR, their understanding of the goal of TR and their beliefs about which aspects of their performance might be affected by TR. Overall, the findings were positive in that 18 out of the 21 participants responded favourably to TR and 'boredom' was not mentioned. The remaining three participants felt that it might have been better to repeat 'similar' tasks instead of identical ones. It became clear to the researchers that a few students did not understand the underlying reasons for repeating the task and so they suggest explaining this to students at the outset. The teachers in this study, on the other hand, reported feeling very unsure about using task repetition as a teaching device as they believed that students would find it boring.

Given these mixed findings for learner and teacher attitudes to task repetition, an important question is: *Why do some learners find TR boring while others find it acceptable?* Possible explanations are that the diversity of learners across studies as well as differences between task procedures have meant that the overall learner 'experience' of TR is different in these studies. Based on the diverse findings of studies, it is possible to suggest a number of potential sources of influence on learner perceptions of task repetition.

1) **Age of learner** - It is possible that younger learners have a lower tolerance to task repetition (Van de Guchte et al., 2015 although see also Pinter, 2007 and DeKeyser, 2010)
2) **Motivation** – Motivation may play a part in the extent to which learners are prepared to repeat tasks. The participants in Lynch and Maclean (2000; 2001) for example were highly motivated and the tasks that they were repeating were directly relevant to their careers. Nunan (1991) has suggested that learners may be more willing to perform certain tasks if they are being reimbursed (as is commonly the case for research studies). In other words, if a student has been paid in order to take part in a study they may respond more favourably to the activity as opposed to if they were not offered payment.
3) **Nature of the repetition task** - It is possible that the task itself may factor into learners attitudes toward repetition. For example, Lynch and Maclean found that students didn't really perceive the task as repetitious, presumably because there was fresh incentive to communicate each time (i.e. the repetition was ecological). Lambert et al. (2017) similarly found that ecological task repetition was received positively by the students. As suggested in Kim (2013), *exact* task repetition may be seen as less tolerable than *procedural* repetition.
4) **Framing** - Although no studies have yet tackled this subject *empirically,* a number of researchers have noted that the key to TR's acceptance by learners might be how it is 'sold' to them. In other words it may be, as Lynch and Maclean (2000) point out, that students need to be explained the reasons for repeating the task (see also Maurice, 1994; Foster & Hunter, 2016). It may also be that the teacher needs to have confidence in using TR (Ahmadian et al., 2017)
5) **Proficiency** - Although Lambert et al. (2017) found that proficiency didn't interact with *performance* and TR, Lynch and Maclean (2000) suggested that their lower proficiency learner was not able to appreciate the gains that she had made in her performance. This could arguably impact on attitudes toward task repetition. The researchers argued that lower learners need to be guided towards noticing improvements that they make.

6) **Cultural differences** – DeKeyser (2010) suggests that cultural background can have an impact on the extent to which learners are tolerant of repetitiveness in language learning.

5.5.4 Summary of key findings

To provide an interim summary of research findings in the area of TR, we can say with some degree of confidence that exact task repetition increases short-term fluency (and accuracy and complexity). Very recent research which has looked at fluency specifically, suggests that increases in specific aspects of fluency (i.e speed, breakdown, repair) may come after different numbers of repetitions, suggesting that specific cognitive processes are affected differently during TR. In terms of whether other aspects of performance are affected by TR (i.e. accuracy and complexity) it seems that a crucial factor is whether the performance is 'pressured' or not. When performance is pressured, it seems that learners are unable to make use of increased attentional capacity to ameliorate performance in terms of complexity and accuracy. When the repetition is unpressured, there is support for the belief that accuracy and complexity increase. PR also seems to result in increased fluency (and accuracy and complexity)

The picture with regards to the effect that TR and PR has on fluency transfer is mixed. It may be that TR has long-term benefits for fluency especially with massive practice (i.e. both Ahmadian (2011) and N. de Jong and Perfetti (2011) featured substantial amounts of repetition). No longer-term fluency gains have yet been found for PR. However, there have been a transfer effect found in other areas of performance such as complexity (Kim & Tracy-Ventura, 2013) and accuracy (Patanasorn, 2010) and there is still a dearth of research which has investigated PR and longer-term fluency or *fluency transfer*.

Student attitudes to TR and PR are mixed. In studies that have elicited students' opinions via interview/questionnaire, it seems that students' attitudes are, on the whole, positive towards TR/PR. In other studies where attitudes have been *inferred* by teachers or researchers based on task performance (Van de Guchte, 2015; Plough & Gass, 1993) it seems that TR is less favourable. In a study which compared attitudes towards *both* TR and PR, PR came out as being favoured by students. Studies which have elicited teacher opinion about TR suggest that they are apprehensive about using exact TR with a class.

5.6 Chapter summary

In this chapter, I have provided a comprehensive synthesis of the research in the area of TR. I then presented the key empirical base that I would be drawing on in the literature review (essentially, the studies that have investigated TR and/or PR effects on L2 short- and/or fluency transfer). Using these studies as a base, I explained the main findings of the research as they related to three key areas of academic debate: (1) the effects of TR/PR on short-term fluency, (2) the effects of TR/PR on fluency transfer and (3) learners' and teachers' attitudes towards TR/PR. It was explained that there is plenty of evidence to suggest that both TR and PR result in increased *short-term* fluency. However, there is a dearth of research which has looked at the transfer effects and that which has, is contradictory and inconclusive. PR, in particular is under-researched in the area of L2 fluency but has thrown up some interesting results in the sense that transfer *has* been observed for PR in certain areas of performance. In a particularly relevant study, N. de Jong & Perfetti (2011) found both short-term and fluency transfer for TR. This is a hugely important finding and yet, as discussed, certain practical and methodological choices mean that the study is limited in terms of its generalisability, particularly to an authentic classroom context.

In terms of learners' and teachers' perceptions of fluency it seems that, on the whole learners are more accepting of TR than teachers, who seems to be concerned that their students will find TR boring. There is also support for the argument that learners may find repeating the task procedure but with different content more engaging than repeating the exact same task.

Overall, then, TR is a rich and interesting area of study because it provides an opportunity to observe the complex process of speech production as well as offering a potential pedagogic tool with which to develop L2 fluency. However, empirical and methodological gaps exist in the literature which this study aims to fill. In the next chapter, I explain how the findings of this ever-growing body of research have informed the research questions and hypotheses which have guided the present research.

6 – Rationale, RQs and Hypotheses

6.1 Introduction

In this chapter, I provide a brief rationale for this research, before presenting the research questions that have guided the study and the corresponding hypotheses.

6.2 Rationale

In Chapter 2, I showed that fluency and its analysis for L2 research purposes has made great strides over the past two decades. Researchers have identified more reliable and robust ways of capturing and measuring L2 fluency and new technological solutions have been offered to assist with the analysis of large quantities of L2 data. Most importantly, we are closer than ever to being able to identify certain characteristics of speech that are related to particular cognitive processes which underlie L2 speech production. Furthermore, research in SLA has offered certain pedagogic interventions which may foster fluency development, a key skill which learners are keen to improve.

Chapter 3 focused on one of these 'fluency-fostering' practices: *task repetition*. Here, we saw that research in this area has grown and that, certainly in the short-term, performance seems be enhanced by TR. In terms of fluency specifically, there is growing support for the theory that task repetition might assist learners in terms of planning and priming meaning that learners are able to divert more of their attentional resources to form, to noticing gaps in their interlanguage and taking on board (corrective) feedback from their interlocutor (Kim & Payant, 2014; Kim, 2013; Payant & Reagan, 2016). What is lacking in TR studies is a sufficiently fine-grained conceptualisation and operationalisation of fluency and particularly that which considers the impact of TR on different aspects of fluency.

TR has also been shown to have the potential to impact on longer-term language development by providing the extensive practice needed to push students towards the proceduralisation of language knowledge although the picture with regards to fluency transfer effects of TR is still not entirely clear. Similarly, it is unclear what the differential effects of repeating exactly the same tasks or the task *procedure* might be, although it seems that PR may have particular long-term benefits as well as being more popular with students and teachers (Kim, 2013).

As explained in Chapter 1, a cause for concern is that, despite growing support for the fluency benefits of TR, neither materials writers (Rossiter et al., 2010) or language teachers (Tavakoli & Hunter 2017) appear to be incorporating TR into their work. It is suggested, then, that Lynch and Maclean's (1994) poster carousel may be a useful tool for researching TR effects because it provides learners with an opportunity to repeat tasks in a *pedagogically convincing* way (Bygate & Samuda, 2005) and therefore might have more appeal to language teachers who adopt a task-based or task-supported approach to language teaching. However, we know very little about the effects of the poster carousel task sequence on fluency specifically because it has only ever been investigated on a qualitative, case-study basis. Another 'inherently repetitive' pedagogic task sequence, the 4-3-2 technique, has enjoyed considerable academic interest (Nation, 1989; Arevart & Nation, 1991; N. de Jong & Perfetti, 2011; Boers, 2014; Thai & Boers, 2016; De Jong, 2012), but this line of enquiry has reached something of a pedagogic dead-end because it has been found that this sort of repetition favours fluency but sacrifices accuracy and complexity.

6.3 Research Questions

Using the Poster Carousel task sequence as a tool for exploring task and procedural repetition effects on fluency, the overall questions which guided this research were:

1. (a) What is the impact of immediate, narrative task repetition (TR) and procedural repetition (PR) on adult ESL students' oral fluency? (b) How are complexity and accuracy affected by this form of TR and PR?
2. What is the impact of immediate, narrative task repetition (TR) and procedural repetition (PR) on adult ESL students' performance on a new task of a different type (fluency transfer)?

3. What are adult ESL students' opinions of the two different versions of the carousel task sequence?

6.4 Hypotheses

It is possible to make a number of hypotheses based on research which has explored TR and PR in similar ways. The following hypotheses are grouped into those which have to do with short-term fluency and fluency transfer.

6.4.1 Short-term fluency

1a) Short-term fluency will increase during both TR and PR carousel task sequences but fluency will increase to a greater extent for the TR group (N. de Jong & Perfetti, 2011).

1b) For the TR group, different aspects of fluency (i.e. speed, breakdown, repair) will develop at different points in the carousel cycle. Speech rate will increase steadily over the course of the three performances. Subtler changes relating to breakdown fluency (e.g. mid-clause pausing) and repair will take place only on the third performance, if at all (Lambert et al., 2017).

1c) Fluency will increase for both the TR and PR groups in terms of holistic rater scoring of fluency (Gass et al., 1999).

1d) Speech rate and mean length of run will correlate most strongly with the NS rater scores (Kormos & Denes, 2004; Derwing et al., 2004).

1e) Complexity (Bygate, 2001) and accuracy (Wang, 2014; Lynch & Maclean, 2000; 2001) will increase for both the TR group and PR group (Sample & Michel, 2014).

1f) For the TR group, complexity and accuracy will increase on the first repetition (Wang, 2014). For the PR group, fluency will increase on the first repetition while complexity and accuracy will not increase until the third performance (Sample & Michel, 2014).

6.4.2 Fluency transfer

2) There will be an increase in fluency between pre- and post-test only for those students that took part in TR carousel training (N. de Jong & Perfetti, 2011). There will be no increase in fluency for the PR group.

6.4.3 Learners' perceptions of TR and PR

3) Learners will respond positively to both task sequences but will display a preference for procedural repetition (Kim, 2013)

Having outlined the theoretical and pedagogic rationale for the current study as well as introducing some research questions and hypotheses, I turn now to a discussion of the methodological approach of the present study.

7 - Discussion of Methodology

7.1 Introduction

In this chapter, I present a discussion of the methodological approach of the current study. I begin with a review of the methodology and methods of the core TR studies and discuss these in terms of design, data collection and analysis. Next, I discuss the rationale for mixed methods research in task-based research and in the area of TR, more specifically. Following the advice offered in Dörnyei (2007), and Riazi and Candlin (2014), I set out my philosophy and reasons for mixing methods. I explain how philosophy and purpose translate into the specific ways in which the present study 'mixes' quantitative and qualitative methods. I follow this with the rationale for conducting the study in an ESL classroom. I conclude this chapter with a brief chapter summary.

7.2 Methodology in TR/PR fluency research

Ellis and Barkhuizen (2005) explain that research in SLA can be divided into that which is 'normative' and 'interpretive' while Riazi and Candlin (2014) prefer the terms 'deductive' and 'inductive'. They explain:

> it has been common practice to differentiate between deductive (top-down) and inductive (bottom-up) approaches to research in language teaching and learning. In a deductive or theory-driven approach, we begin from a theory or theoretical framework and derive a hypothesis from it; we may then be able to provide evidence for or against the hypothesis by observing the phenomena under review and by collecting and analysing appropriate data. The outcome of such deductive research will either strengthen the theory by verifying the hypothesis or weaken its explanatory power if the evidence and analysis do not support the hypothesis. In contrast, inductive or data-driven approaches to research begin from inspection of the data,

seeking meaningful patterns and generating hypotheses which may then, in the case of well-designed large-scale research projects, generate further theory. (Riazi & Candlin, 2014, p. 136)

Both Ellis and Barkhuizen (2005), Riazi and Candlin (2014) and Samuda and Bygate (2008), explain that the researcher's adherence to a particular paradigm will influence other aspects of the research design, data collection and analysis. For example, Riazi and Candlin (2014) explain that deductive research studies tend to also adopt a quantitative approach. Inductive studies, on the other hand, often draw on qualitatively analysed case study data. Samuda and Bygate (2008) further mention that quantitative research tends to draw on group data (macro) and qualitative research tends to be carried out with case study data. I will therefore explore the methodological approach of the core TR studies along these two different dimensions: 1) quantitative/qualitative and 2) macro/micro.

7.2.1 Quantitative/qualitative

The labels 'quantitative' and 'qualitative' can refer to the *type* of data which is collected and also to the *way* in which the data is analysed (Dörnyei, 2007). Examples of quantitative *data* can obtained through the administering of language tests and questionnaires. Qualitative data can be obtained through, for example, interviews with participants or ethnographic study. In the core TR studies, much of the data has been quantitatively collected using oral language tests of some kind. This is because the main research questions in TR literature have been related to TR effects on task performance. However, a number of studies in the field also collected qualitative data in the form of interviews and/or questionnaires with teachers (e.g. Kim, 2013), learners (Pinter, 2007; Lambert et al., 2017) or students *and* teachers (e.g. Ahmadian et al., 2017).

'Quantitative' and 'qualitative' can also refer to the way in which the data is analysed (i.e. the extent to which data is 'counted'). In general, the TR studies which have elicited quantitative data tend to analyse that data quantitatively (although see Lynch & Maclean for an exception). Likewise, data which is collected via qualitative methods tends to be analysed qualitatively (although questionnaire data can be analysed both quantitatively *and* qualitatively e.g. Lambert et al. (2017)).

Table 20: Core TR studies and research methodology

Study	Quantitative/qualitative
Ahmadian & Tavakoli, 2010	Quantitative
Ahmadian, 2011	Quantitative
Boers, 2014	Quantitative
Bygate, 1996	Quantitative
Bygate, 2001	Quantitative
Fukuta, 2016	QUANT + qual
Hsu, 2017	Quantitative
Lambert et al., 2016	QUANT + qual
Lynch & Maclean, 2000; 2001	Qualitative
N. de Jong & Perfetti, 2011	Quantitative
Sample & Michel, 2014	Quantitative
Thai & Boers, 2016	Quantitative
Wang, 2014	Quantitative

The picture that emerges is that, while some studies collect complementary qualitative *data* in the form of post-intervention questionnaires (see Table 20), few have adopted a qualitative approach to data *analysis*.

7.2.2 Macro/micro studies

The 'Macro/Micro' dimension, as outlined in Samuda and Bygate (2008) refers to the divide in TBLT research between that which adopts a group design (macro) and that which adopts a case study design (micro). Although this tends to be linked to the deductive/inductive dimension, outlined above, in that deductive studies *tend* to adopt a macro design and inductive studies *tend* to adopt a micro design, Samuda and Bygate (2008) also note that some deductive[12] studies can be micro in design and *vice versa*. There are clearly benefits and drawbacks to using both research designs.

Table 21: Core studies and design (micro/macro)

Study	Macro/Micro
Ahmadian & Tavakoli, 2010	Macro
Ahmadian, 2016	Macro
Ahmadian, 2011	Macro
Arevart & Nation, 1991	Macro
Boers, 2014	Macro
Bygate, 1996	Micro
Bygate, 2001	Macro
Fukuta, 2015	Macro
Hsu, 2017	Macro
Lambert et al., 2016	Macro
Lynch & Maclean, 2000; 2001	Micro
N. de Jong & Perfetti, 2011	Macro
Nation, 1989	Macro
Sample & Michel, 2014	Macro
Thai & Boers, 2016	Macro
Wang, 2014	Macro

Table 21 shows that the vast majority of TR studies adopt a macro design. Bygate (1996) and Lynch and Maclean (2000; 2001) are the only two exceptions. What this means for the overall profile of TR research is discussed below.

[12] Although they use the term “systemic”

7.2.2.1 The case for more case studies

In second language research, there have been numerous calls for more case study research for a number of reasons. Firstly, it is widely acknowledged that studies which rely on group data alone can miss important individual variation in the data (Samuda & Bygate, 2008). Secondly, case studies often support group study findings by attempting to answer the “why?” or “how?” questions that researchers may have about a particular topic (Dörnyei, 2007). A final point is in relation to accessibility. It has been suggested (Nassaji, 2012) that case studies are more relevant for a non-academic audience (i.e. language teachers) because they are easier to interpret than statistical analysis. This final point is particularly salient to the current study, given that L2 fluency development is an area where the ‘gap’ between research and practice is noticeable. Previous research into TR effects on L2 performance has tended to be based on (small) group studies with only Bygate’s (1996) and Lynch and Maclean’s (2000; 2001) studies offering case study analysis.

Bygate’s (1996) ‘exploratory’ case study of TR effects on a single learner of Spanish attempted to shed light on the TR process. This study investigated exact task repetition and its impact was measured in terms of CAF. Despite the fact that this research was conducted with a single learner, this was a hugely influential study and generated a great deal of interest in the topic of task repetition effects on L2 performance.

Lynch and Maclean (2000; 2001) similarly adopted a case study design in order to explore the benefits of the poster carousel task sequence. They studied a total of five cases (out of an intact class of 14) across 6 performances of the task. Their findings shone light on what different speakers do when they repeat a task. Repetition in this study was ecological task repetition whereby the input material was the same but the interlocutor changes. The impact of performance was examined in terms of performance (loosely related to CAF). The main findings of this study were that students used the task sequence to improve their performance in qualitatively different ways. The researchers focused on how learners seemed to be taking cues from their interlocutors’ input to ameliorate their subsequent performances. Once again, although this study centred on a handful of specific cases, its contribution to the field was considerable, in part because of the context and nature of the task, but also, I believe, because its case study design lent it a relatable air.

There are two additional TR case studies which have focused on TR effects on other discursive phenomena. Firstly, Bygate and Samuda (2005) presented three case studies to exemplify and triangulate the associated group findings in their 2005 study. The type of repetition in this study was exact task repetition and the performances were analysed in terms of instances of 'framing' of the narratives. Their case studies allowed them to present their group findings (i.e. that TR led to more 'framing' in the learners' narratives). Secondly, Pinter (2007) also adopted a case study approach in her analysis of the task performances of two 10-year-old boys on original and repeated versions of a spot-the-difference task. Repetition in this study was defined as similar versions of the task and would therefore come under the heading of procedural repetition as defined in the current study. The performances were analysed for instances of 'peer assistance'. Although the focus of these two studies was not L2 performance, *per se*, they do contribute a great deal to our understanding of the TR *process* and, in particular what happens when *communicative* tasks are repeated.

Over the last decade, however, there have been, to the best of my knowledge, no published case studies which have compared TR and PR effects. Furthermore, despite the considerable interest in TR effects on L2 utterance fluency (e.g. N. de Jong & Perfetti, 2011; Lambert et al., 2017; Thai & Boers, 2016), there have been no attempts to carry out case study analysis on data in order to investigate fluency effects specifically. The current study will therefore aim to cover these lacunae.

7.2.3 Summary of methodological approach in TR/fluency studies

In general, the theoretical approach of research in TR can be classed as primarily *deductive*. Analysis has been largely quantitative and carried out at the group level (although not necessarily large-scale). Only a handful of TR studies have deviated from this (Lynch and Maclean, 2000; 2001; Pinter, 2005). Unfortunately, though, the fact that the poster carousel has *only* been investigated in a qualitative way and through a handful of studies means that it is difficult to connect this study up with the rest of the TR research and it is also difficult to be entirely convinced that the findings are generalisable. So, while Lynch and Maclean (2000; 2001) provide a counterpoint in terms of methodology, it is 'out on a limb', disconnected from the growing body of research in TR. Arguably, what is needed, then, and certainly if the aims of the present research are to be met, is research which is able to combine ecological validity with internal validity (DeKeyser, 2010; 2017). One way of achieving this is through the use of mixed methodology. This is the subject to which I will now turn.

7.3 A mixed-methods approach to task repetition in the ESL classroom

7.3.1 What is mixed-methods research and what can it achieve?

Riazi and Candlin (2014) explain that research approaches can be classed as deductive or inductive and they illustrate this in terms of Fishman's (2010) German expressions Erklarung (explanation) and Verstehen (understanding). Riazi and Candlin go on to explain that "to achieve harmony in the theory and practice of research, the specific methods selected need to fit the general purposes and specific objectives of the investigation". This means that research which seeks to either 'Explain' or 'Understand' will be characterised differently in terms of the study design and scale as well as other factors such as the location for the research (Samuda & Bygate, 2008).

As explained above, research which has investigated task repetition and fluency has tended to be largely *deductive* or that which seeks to 'Explain'. At the other end of the spectrum, *inductive* research in TR has been thin on the ground. A response to this observation might be to simply cover this lacuna by conducting inductive research into TR and fluency. However, if I return to the motivation for the study which was set out in Chapter 1, I am interested in contributing to two fields of knowledge and each of these fields has certain expectations in terms of research. If I conducted a study which was purely *inductive,* drawing on a handful of interesting cases and interpreting the data based on qualitative analysis, it may well speak to language teachers but it might not hold water from a psycholinguistic standpoint. The reverse is also true. There is, however, a "third method" in research: that which involves a *mixture* of both quantitative and qualitative components within a single research project. This provides a means of conducting research which is interested in both 'explaining' and 'understanding' what happens when learners repeat tasks. This third methodology can be seen as an "emerging design of considerable scope and value" (Riazi & Candlin, 2014, p. 138) whereby the relative strengths of both a quantitative and qualitative approach to research are combined within a single research design.

In simple terms, a mixed methods study comprises "the collection or analysis of both quantitative and qualitative data in a single study with some attempts to integrate the two approaches at one or more stages of the research process" (Dörnyei, 2007, p. 163). However, this view may be seen as simplistic and there is "a degree of confusion among language teaching and learning researchers as to what precisely constitutes MMR" (Riazi & Candlin, 2014, p. 140).

Tashakkori and Teddlie (2003) and Riazi and Candlin (2014) explain that researchers need to be clear about *why* they are using MMR. This is because an understanding will help researchers conceptualise and design the MMR study (Tashakkori & Teddlie, 2003) and aid researchers in interpreting their findings (Riazi & Candlin, 2014). Riazi and Candlin (2014) present *five* main purposes for conducting MMR. These are 1) *Triangulation*, which is the dominant purpose for employing MMR in language teaching research and which involves the use of different methodologies to 'cross-check' each other (see also Dörnyei, 2007), 2) *Complementarity*, which requires MMR so that different methods can be used to look at different levels or layers of a phenomenon, 3) *Development*, whereby a MMR is preferred because one part of the study leads to or informs another, 4) *Initiation,* which refers to the purpose of using MMR to uncover contradiction and paradox between the two methodologies and thereby generate further data collection and analysis and 5) *Expansion* which seeks to extend the breadth and depth of the enquiry by adding an additional research phase. This is considered the most flexible and features less integration of quantitative and qualitative methods.

7.3.2 Mixed methods in Applied Linguistics and language teaching

Hashemi (2012) reviewed 273 articles that had appeared in applied linguistics journals between 1995 and 2008. It was found that 205 (75%) incorporated both quantitative and qualitative elements to some extent. This suggests that MMR is a popular methodological approach in applied linguistics generally. Riazi and Candlin (2014) narrowed the scope to language teaching and selected 40 papers published in language teaching journals (e.g. *TESOL Quarterly*; *Language Teaching Research*). They selected papers based on key search terms such as "mixed-methods", "quantitative" "qualitative" and "triangulation". In other words, studies were selected based on the extent to which they described themselves as combining quantitative and qualitative methods. The researchers found that most studies utilised MMR with the purpose of *triangulation* and *expansion.* They found that *complementary* studies – those studies which adopted mixed methods in order to look in more detail at a particular phenomenon - in their words: "fully-fledged and unequivocal MMR studies", were very thin on the ground.

7.3.3 Mixed methods in task repetition studies

However, despite the gaining popularity of MMR in language teaching research (Hashemi, 2012) there have been few mixed methods studies in TR/CAF research and even fewer of the sort that Riazi and Candlin call 'true' MMR. Lambert et al. (2017) and Kim and Tracy-Ventura (2013) both include a qualitative phase of data collection and analysis alongside their quantitative study in order to *expand* on their findings. Both use post-intervention questionnaires/interviews to gauge learner opinion about TR. In other words, these two studies both employ MMR with the purpose of *expansion* and, to the best of my knowledge, there have been no MMR studies in this area which have mixed methods for the purpose of *complementarity.*

7.3.4 Rationale for mixed-methods in the current research

I have shown, then, that there are five purposes for employing MMR. The present study mixes-methods in two different ways with two distinct purposes: *expansion* and *complementarity.* Firstly, I have explained that one of my overall aims is to identify ways in which TR can be operationalised in the classroom (Chapter 1). This was motivated by the awareness of a 'gap' between research and practice. In this sense, the purpose of using MMR is for *expansion* purposes; to reach a wider pedagogic audience (Dörnyei, 2007). A qualitative phase of research will consider the ways that TR can be implemented in real classrooms by considering learners' views of task repetition. This purpose for mixing methods can be described as one which seeks *expansion.*

Secondly, given that I have described the construct of fluency as a multi-dimensional and complex phenomena (Chapter 2), understanding how it interacts with TR will involve examining these different layers (e.g. cognitive, utterance, perceived fluency). L2 fluency development is a 'complex' phenomenon in terms of conceptualisation and measurement

(Tavakoli, 2016; Segalowitz, 2010; Tavakoli & Hunter, 2017). While utterance fluency can be measured quantitatively using measures such as speech rate, making connections to specific cognitive processes (i.e. *automaticity*) will require analysis at a deeper (qualitative) level. The qualitative analysis will *complement* the quantitative analysis in that it will provide a more nuanced picture of L2 fluency. This purpose for employing a mixed-methodology can therefore be defined as one of *complementarity.*

7.3.5 How does this study mix methods?

Following on from the two separate reasons for mixing methods given the current research questions and hypotheses, I will now give some indication of precisely the way that quantitative and qualitative phases of the research are implemented to achieve *expansion* and *complementarity*. Firstly, and similarly to Lambert et al. (2017) and Kim (2013), the study will involve a quantitative phase which establishes the impact of the intervention followed by *qualitative* focus group phases which elicits learners' feelings and attitudes towards the intervention. The study aims to offer working solutions for teachers who may wish to implement TR in their own classrooms, as such the research seeks to identify key factors which may be important in that endeavour.

Secondly, the focus of the study, L2 fluency, has been defined as a complex and multi-dimensional phenomenon. As such, the intervention data (performances) will be analysed both quantitatively (utterance fluency, perceived fluency) *and* qualitatively. Both methodologies are employed concurrently and the case studies which are selected are *nestled* within the group data.

7.4 Research environment in TR studies

Another methodological dimension which divides research in TR is the research environment and, specifically, whether research is carried out in a laboratory or in a classroom (Ellis & Barkhuizen, 2005). Nunan (1991) and Foster (1998) explained that while many SLA studies purported to have implications for language teaching, they were often not carried out in classroom environments. In 1991, Nunan wrote:

> "As the language classroom is specifically constituted to facilitate language development, this should constitute sufficient justification for studying what goes on

> there. Despite this seemingly uncontroversial observation, it is evident from this review that little second language research is actually carried out in language classrooms, and that we know comparatively little about what does or does not go on there. The existence, and indeed persistence, of this state of ignorance may seem surprising given the frequency with which attempts are made to import into second language classrooms insights from research conducted outside the classroom." (Nunan, 1991, p. 265)

Foster (1998, p. 21) similarly explains that if "language acquisition research wants to feed into teaching methodology, the research environment has to be willing to move out of the laboratory and into the classroom". More recently, however, Dörnyei (2007, p. 176) suggested that things have changed somewhat and that "the classroom – and most often the foreign/second language classroom – is a primary research site in applied linguistic investigations". However, the review of the methodology in TR studies (see Table 22) revealed that the vast majority of research has been conducted within a laboratory setting. While studies that have looked at TR effects in relation to alternative dependent variables such as LREs are sometimes carried out in classroom environments (e.g. Pinter, 2007; Kim, 2013), those that seek to analyse learner speech in terms of CAF are almost invariably carried out in a laboratory environment (although exceptions are Lambert et al., 2017 and Hsu 2017.

The reason for this is perhaps because the laboratory environment means that elicited speech data will be clearer and easier to analyse, allowing researchers, for example, to make use of automated speech analysis (Hilton, 2014). Does this mean that CAF analysis is simply not possible when carrying out TR in an authentic classroom? One very recent study (Lambert et al., 2017) did attempt to investigate fluency effects of TR in a classroom environment. However, they report that the choice of location meant that they couldn't analyse silent pauses automatically and were therefore unable to calculate certain measures which are commonly used in the field. Furthermore, their study took place in a room that was used for delivering language classes but the study did not use intact classes, nor did it take place during scheduled class time.

Table 22: Core studies and research environment

Study	Type
Ahmadian & Tavakoli 2010	Laboratory
Ahmadian 2011	Laboratory
Boers 2014	Laboratory
Bygate 1996	Laboratory
Bygate 2001	Laboratory
Fukuta 2015	Laboratory
Hsu 2010	Classroom
Lambert et al. 2016	Classroom
Lynch & Maclean 2000; 2001	Classroom
N. de Jong & Perfetti 2011	Laboratory
Sample & Michel 2014	Laboratory
Thai & Boers 2016	Classroom
Wang 2014	Laboratory

One of the reasons for carrying out research in a classroom is because research environment may impact on learners' task performance (Foster, 1998). This means that if conclusions about the nature of, for example, interactional patterns between learners are *only* investigated in a laboratory environment, they may not be generalisable to language classrooms because the interaction may not happen in the same way. In reaction to Foster's (1998) observations, Gass, Mackey & Ross-Feldman (2011) conducted research which explored the two environments and the impact on interaction. They found that there was not a significant difference between the two environments leading them to suggest that the findings of laboratory-based studies may also be generalised to classrooms. However, it has been suggested that it is not necessarily the location itself that makes classroom and laboratory task performance distinct but rather all the other factors that go with it such as whether or not the class is during scheduled lesson time, whether the classes are 'intact', whether the class teacher is present, whether students are reimbursed for taking part in the study, the nature of the tasks themselves and the pre- and post-task activities or lack of them (e.g. Nunan, 1991; Nassaji, 2012). In the Gass et al. study, it was physical *location* which was investigated. Arguably it is less likely that physical location is *the* factor which might characterise the differences between 'laboratory' and 'classroom' environments.

In terms of the core TR studies, then, while the actual location for some of the TR studies were classrooms, we can see that some did not involve intact classes (Thai & Boers, 2016; Lambert et al. , 2016) or used non-reciprocal tasks (Hsu, 2017; Thai & Boers, 2016; Lambert et al., 2017). In other words, many of these studies may well have been situated in a room which is also used for teaching purposes, but they may not represent the ways in which an authentic class would behave and respond to TR (Nassaji, 2012). An exception is Lynch and Maclean (2000; 2001). In this research, although the focus was not on fluency specifically, an intact class of students carried out interactive tasks in an authentic classroom environment with their class teachers. Van den Branden (2016, p. 247) argues that it is exactly this sort of research which is sorely needed in in TBLT because "this kind of research could generate practice-based recommendations on how to work with tasks in specific kinds of classrooms"

The result is that, to the best of my knowledge, no TR study has attempted to conduct detailed utterance fluency analysis within an authentic, intact classroom. This study will therefore attempt to address this imbalance. Of course, attempting to do so requires facing a number of particular challenges. Schachter and Gass (1996, p. viii, in Dörnyei, 2007) explain:

> Reports of [classroom] research projects make it all look so simple… There is no indication of the blood, sweat, and tears that go into getting permission to undertake the project, that go into actual data collection, that go into transcription, and so forth.

These challenges may be further exaggerated if one wants to conduct L2 *fluency* research in a noisy classroom and need to analyse fluency features *manually*, as Hilton (2014, p. 45) explains: “The transcription and tagging of an oral corpus is… a long and difficult task” (see also the discussion of fluency analysis in Chapter 2). Nonetheless, it is felt that if this particular classroom intervention is going to be properly assessed for its usefulness for teaching and learning purposes, it should be examined in an actual classroom. A discussion of the feasibility of carrying out fluency research in the classroom is taken up in the final chapter.

7.5 Chapter summary

In this chapter, I have reviewed the methodological approach of core TR research, and established that the majority of studies have been motivated by a desire to ‘explain’ the effects of TR on oral fluency but that few have taken the necessary methodological steps to ‘understand’ exactly *why* these effects may occur. I suggested that there was very little qualitative research into the impact of TR on oral fluency and that, in fact, *purely* qualitative research may not hold enough water from a psycholinguistic standpoint. Instead, I argued that there was a strong rationale for a mixed-methods approach to TR research, given that MMR provides a framework to allow researchers to draw on different methodologies to *triangulate, expand,* and provide balance or *complementarity* in their studies. I went on to explain that MMR would be used in the present study with the purposes of expansion and complementarity and explained how the present study mixes methods. I also explained the dearth of TR studies which have been conducted in a working language classroom and outlined the need for more classroom-based research in this area if the ‘gap’ between TR research and pedagogy is to be lessened.

8 – Design and execution of the present study

8.1 Introduction

In this chapter, I explain the design and execution of the study. I describe its key elements: the participants, the context, the study design and the task; and explain my rationale for each. I also set out the independent and dependent variables that I use for the analysis of data collected and describe how case studies were identified.

Pilot study

In addition to being guided by the research methodologies of other similar studies, the design and procedure of the current study was also informed by a pilot study which I carried out prior to undertaking the present research. The primary aims of the pilot study were:

- to assess the feasibility of analysing temporal fluency in an authentic classroom, including whether it would be possible to take advantage of automated methods of speech analysis, as *per*, for example De Jong and Wempe (2009).
- to test the logistics of carrying out the proposed tasks, including whether the materials were appropriate for learners, what issues are associated with recording a number of individuals simultaneously in a classroom and whether it was possible for the students to complete the task successfully within the time allowed by the school's timetable.

The regular class teachers that were involved in piloting the tasks were consulted on the validity and usefulness of the tasks and were asked for their ideas on how to improve the carousel task sequence. Participants from two intact classes of ESL learners were recruited on a voluntary basis for the pilot. In total, 14 participants with a range of L1 backgrounds were involved in the study.

The poster carousel was used as described in Chapter 3. Students were recorded performing three performances of the task on voice recorders. Dependent measures used in this study were: mean length of run, articulation rate and mean length of pause. These measures of fluency were selected because they represented 'global', 'speed' and 'breakdown' measures and also to allow for a comparison with N. de Jong and Perfetti (2011), in which these variables also featured. Mean length of run was calculated as the average number of syllables between filled or silent pauses greater than 250ms. 250ms was used as a minimum pause threshold following many other researchers who have used this same threshold. Articulation rate was calculated as the number of (unpruned) syllables per minute (excluding pauses over 250ms). Mean length of silent pause was calculated as the average length of silent pause over 250ms.

It was found that the materials were appropriate for learners, and that timings, although tight, allowed for the necessary data collection to take place within the language school's timetable. Feedback from teachers indicated that the tasks were appropriate for the classroom, and also suggested possible follow-up tasks which would ensure that both participants had clearly defined roles. The findings pointed to clear gains for students over the course of the three performances in terms of mean length of run and articulation rate and stable length of pause. A number of additional observations were made that were relevant for the final design and procedure of the present research:

First, it became clear early on in the data coding process that it would not be possible to analyse temporal fluency 'automatically' with this sort of data. Attempts were made to run both the 'textgrid to silences' function in PRAAT and customised scripts for fluency analysis (e.g. de Jong & Wempe, 2009) but even with high threshold settings for silences, they were simply not compatible. As a result, it was clear that I would have to use PRAAT manually in order to annotate the soundfiles, individually marking pause boundaries, and that I was required to develop programming code that would allow me to compute frequency counts and durations of the features I had highlighted. As other researchers (e.g. Hilton, 2014; Witton-Davies, 2014) have noted, this is an intense and time-consuming endeavour. However, I found that my speed increased with practice and I concluded that this would be a feasible way to code and analyse the data for the main study which would obviously be much larger in scale.

Second, it was observed that the highly interactive nature of the carousel would make it difficult to make specific claims about the fluency of individuals. As McCarthy (2010) pointed out, fluency in interaction is complicated to analyse because there are additional factors such as turn taking which need to be factored in. Although the analysis of fluency in interaction is a crucial direction for future research (Tavakoli, 2016), there is not yet any established method for the systematic analysis of L2 fluency in this mode (Kim & Tracy-Ventura, 2013). Following this observation, in the main study, instructions were given to participants to allow the 'presenters' to tell their stories before asking any questions that they might have. This would ensure that a certain proportion of the speech that was elicited would be monologic (and more amenable to utterance fluency analysis) but students would still be given the chance to discuss and negotiate for meaning.

A final observation from the pilot study was that additional measures would be necessary to counter loss of data through human error. Specifically, several recordings were lost because students had accidentally switched off their voice recorders at the wrong time. In order to ensure that as much of the data in the main study was recoverable, it would be necessary to 'lock' voice recorders so that they could not be interfered with during recording.

8.2 The main study design

The study design is a quasi-experimental, mixed within- and between-subjects design with three phases of data analysis (pre-test/post-test; intervention; post-study focus groups) which address the three broad research questions. Intact classes of learners were randomly assigned to the three conditions:

1) Task repetition group (TR) (n= 24);

2) Procedural repetition group (PR) (n=22);

3) Control group (C) (n=18)

The study therefore involves two experimental groups and one control group, with multiple classes of students assigned to each group (10 classes in total: 3 in the TR group, 4 in the PR group and 3 in the C group.) For each class of students involved in the study, the procedure took one school week (5 days). There were a total of 10 classes involved in the study so the entire data collection took 10 weeks.

As illustrated in Table 23, learners in all three groups completed oral pre- and post-tests on the first (Monday) and fifth (Friday) day of the training. On the fifth day, after the post-tests were carried out, the students took part in semi-structured focus groups with the researcher, their own teacher and their fellow classmates.

The experimental TR and PR groups took part in carousel training sessions which took place on the Tuesday, Wednesday and Thursday of the same week[13]. The two experimental groups differed in that the TR group took part in a carousel which involved task repetition while the PR group took part in a carousel which involved procedural repetition. Those students that were assigned to the control condition carried out the pre- and post-tests but received no specific training beyond their usual scheduled lessons. However, in order to be fair to those students who were in the control group, carousel training was provided in the week following the intervention although this data was not collected for analysis.

Table 23: Overview of study

Group	Day 1	Day 2 Intervention	Day 3* Intervention	Day 4* Intervention	Day 5
Task repetition (TR)	Pre-test	TR carousel 1	TR carousel 2	TR carousel 3	Post-test
Procedural repetition (PR)		PR carousel 1	PR carousel 2	PR carousel 3	
Control (C)		-	-	-	

Data from day 3 and day 4 were collected but not analysed for this study

8.3 Context and classes

The study was carried out at a private English language school in central London during the summer and autumn of 2013. The school provided learners with 'General' English language courses (min. 20 hours per week). This course consisted of English classes in the morning

[13] An analysis of all three sessions was not required to answer the research questions posed by the present study. For this reason, I only refer to the training session data that was elicited during the first session (day 2).

which followed a coursebook and which covered a range of practical English vocabulary and grammar as well as all-round skills development, as well as a ‘Speaking and Listening’ class which focuses on developing oral skills in particular. Afternoon sessions were elected by the student and included Business English, ‘Word Power’ - a vocabulary-based class, ‘Grammar and Writing’, and exam preparation (e.g. TOEIC, TOEFL, IELTS).

The research was carried out during the students’ ‘Speaking and Listening’ class. In this class, which was 50 minutes in length, teachers generally tended to use role-play, debates and discussions, or what Tavakoli and Hunter (2017) refer to as “free-communication” activities, as well as creative ‘project-type’ activities such as making a short documentary or news bulletin. The class was chosen because the training in this study was communicative in nature and would therefore be the most natural fit with the school’s curriculum and the students’ expectations.

The research was concerned with 10 intact classes. Each class had a maximum of 14 students (see appendix for a breakdown of the exact numbers in each class). The proficiency level of these classes was Intermediate Plus (B1-B2). The school allocated new students to a level based on their performance on an in-house grammar test, a short oral interview and a writing sample. Due to the nature of the ‘continuous enrolment’ of students at this language school, some students were at the beginning of their course of study while others were in the middle or approaching the end of their studies. This was unavoidable and a natural consequence of using intact classes. This was expected to have limited impact on the overall proficiency level of the group, however, as students were often ‘moved up’ to the next level when teachers considered it appropriate, meaning that the overall proficiency level of each class remained more or less the same.

8.4 Participants

A total of 93 students agreed to take part in the study on an entirely voluntary basis. However, the fact that the data were collected over the course of an entire week meant that, due to student absences, only 64 complete datasets were available for analysis. Of these remaining students 30 were male, 34 female. The age range of these students was 18 - 42, and the average age was 20.1 years. First languages of all participants can be found in Table 24 below. French *(n*=13) and German (*n*=13) were the most commonly reported L1s.

Length of study for these students varied a great deal, with some students planning to stay in the school for up to one year and others for only two weeks. While they were studying, these students took advantage of a range of accommodation options with some choosing to stay with a family in London, others living in shared student accommodation arranged by the school, and others who made their own arrangements for accommodation in London. The implications of these additional variables as a potential source of influence was beyond the scope of the current study.

Ethics approval was obtained for the research (see Appendix 2) and all learner participants signed a consent form to take part in the study (see Appendix 3). Participants were not reimbursed for taking part. Students were aware that they were taking part in a study to investigate the effectiveness of the carousel technique for teaching and learning purposes but they were not provided with an explanation of 'fluency' or given any other information about how their data would be analysed. It was explained that they could withdraw from the study at any time. Arrangements were made with other classes to take any students who did not wish to take part in the study but, ultimately, this was not needed as all students from the target classes agreed to take part.

Table 24: L1 backgrounds of all participants

L1 background	Number of participants
Arabic	3
Chinese	2
Dutch	1
French	13
German	13
Italian	8
Japanese	4
Kazakh	1
Korean	5
Portuguese	2
Serbian	1
Slovenian	1

Spanish	7
Swedish	1
Thai	1
Ukrainian	1

Ten ESL teachers were also involved in the study as teachers of the classes and facilitators of the research. They were all minimum CELTA qualified with 2 years' teaching experience. A high proportion also held a DELTA qualification. The researcher was an active teacher at the school at the time of data collection and was therefore known to the teachers involved in the study and also to some of the participants. For reasons of anonymity, the teachers and students involved in this study are given pseudonyms throughout.

8.5 Session tasks and procedure

8.5.1 The Poster Carousel

The session tasks used for both the TR and PR groups were loosely based on Lynch and Maclean's 'Poster Carousel' (1994) (see Chapter 4), which requires students to create posters based on academic journal articles and then to respond to visitors' questions about their posters in very much the same way that poster presentations operate in academic conferences. As their learners were medical students, who were on a highly specialised '*English for Cancer Conferences'* course, the stimulus material was entirely relevant and appropriate. However, as the learners in the present study were enrolled on 'General English' courses, the stimulus material needed to reflect this.

As trialled in the pilot study, the 6 stimulus texts were written first-person descriptions of extreme life experiences (see Appendix). The material was sourced from the 'Experience' column of the Guardian newspaper (www.theguardian.com/experience). These first-person accounts had titles such as 'I was crushed by a cow', 'My cat saved my life' and 'I was trapped in a ravine for eight days'. They were of equivalent length and were similar in many ways (e.g. tense, point of view) although the particular lexis varied greatly from story to story in reflection of the different contexts and nature of the story being told.

Participants in this study were asked to create storyboards to aid their description of the stimulus material. The use of storyboards for speech elicitation has a rich history in SLA studies, meaning that certain comparisons can be drawn among studies (e.g. Fukuta, 2016; Hsu, 2017). As noted in Hsu (2017) this type of "picture narrative" task elicits a certain amount of speech which is monologic in nature, and is therefore more stable from a temporal fluency research point of view.

Given that the purpose of the study was to investigate the impact of both task and procedural repetition, two versions of the carousel were designed and implemented; one which meant that the inherent repetition was of the exact same task (TR carousel) and another which meant that the repetition was of the same type of task with different content (PR carousel). The details of each of these tasks are outlined below.

Task repetition carousel

Preparation/planning stage

1) Students are put into pairs
2) Each pair is given an envelope containing two copies of a written text. Each pair has a different envelope with different texts inside
3) Each pair works together to understand and summarise their text
4) The students create a poster-sized, 6-frame storyboard based on their written text (see Appendix for sample storyboards)
5) Students' posters are displayed around the room and students stand next to the poster that they have created

Dummy stage[14]

6) Students are split into 'As' and 'Bs', so each pair is made up of an A and a B.
7) The Bs stay with the poster while the As move clockwise to the next poster. The As carry a voice recorder for data collection.
8) Each B presents their story to the visiting A (duration: 2 minutes). The As are instructed to ask questions about the narrative only when the speaker (Bs) had finished explaining the narrative.
9) The As and Bs swap places

[14]Reasons for the 'dummy stage' are given in Section 8.5.2 below

10) The Bs leave their original poster and move clockwise to the next poster

Performance stage

11) Performance 1: The As present the poster to the visiting Bs, answering questions where appropriate (duration: 2 minutes)
12) The Bs continue to move clockwise around the room, to the next poster
13) Performance 2: The As present the same poster to a new visitor (duration: 2 minutes)
14) The Bs again move clockwise around the room, to the next poster
15) Performance 3: The As present the same poster to a new visitor (duration: 2 minutes)
16) Visitors return to their original poster
17) As and Bs now swap roles and the process is repeated (this time moving anticlockwise) until the Bs have presented the same poster three times to three different visitors

Post-task stage

18) In their original pairs, students discuss the different stories that they have heard and make a decision about their favourite.
19) There is a plenary discussion about the posters.
20) The teacher provides individual and general feedback on performances and there may be a language focus stage
21) For homework, students are given the task of writing up their favourite story for the school newspaper.

Procedural repetition carousel

While the procedure for the TR carousel could be modelled on Lynch and Maclean's (1994) 'poster carousel', there was no obvious pedagogic model to draw on for the PR group. It was therefore necessary to come up with a new version of the carousel task sequence which required students to present three different posters rather than the same poster three times. The result was the following procedure:

Preparation/planning stage

1) Students are put into pairs
2) Each pair is given an envelope containing two copies of a written text. Each pair has a different envelope with different texts inside
3) Each pair works together to understand and summarise their text
4) The students create a poster-sized, 6-frame storyboard based on their written text

5) Students’ posters are displayed around the room and students stand next to the poster that they have created

Dummy stage

6) Students are split into ‘As’ and ‘Bs’, so each pair is made up of an ‘A’ and a ‘B’.
7) The ‘As’ stay with the poster while the ‘Bs’ move in a *clockwise* direction to the next storyboard
8) The As present and answer questions about their poster with their ‘B’ interlocutor.
9) After 2 minutes, the Bs (who have been listening to the presentation) are instructed to stay with that poster while the As (who have been presenting their storyboard) move around the room in an *anticlockwise* direction to the next poster

Performance stage

10) Performance 1 (for Bs): the Bs present and answer questions about their new poster with the A interlocutor (duration: 2 minutes)
11) The As stay with the poster and the Bs move around the room in a *clockwise* direction, to the next poster
12) Performance 1 (for As): The As present the poster to their B interlocutor (duration: 2 minutes)
13) The Bs stay with the poster and the As move around the room in an *anticlockwise* direction, to the next poster
14) Performance 2 (for Bs): the Bs present and answer questions about their new poster with the A interlocutor (duration: 2 minutes)
15) The As stay with the poster and the Bs move around the room in a *clockwise* direction, to the next poster
16) Performance 2 (for As): The As present the poster to their B interlocutor (duration: 2 minutes)
17) The Bs stay with the poster and the As move around the room in an *anticlockwise* direction, to the next poster
18) This process (step 14-17) is repeated for a third time.
19) All students return to their original storyboard

Post-task stage

20) In their original pairs, students discuss the different stories that they have heard and make a decision about their favourite.
21) There is a plenary discussion about the posters.
22) The teacher provides individual and general feedback on performances and there may be a language focus stage
23) For homework, students are given the task of writing up their favourite story for the school newspaper.

8.5.2 Controlling for topic familiarity

This study is rare insofar as it looks at the impact of both TR and PR techniques in the classroom. This was essential to meeting this study's aims, but it provided certain challenges from a practical perspective. So while Lynch and Maclean's poster carousel (1994) provided the template for task repetition included in this study, certain modifications were needed to ensure internal validity.

In the 'traditional' TR carousel (as per Lynch & Maclean, 1994), students present the story that they had been working on (i.e. their *own* storyboard) three times. Doing the same thing for the PR carousel, on the other hand, would entail presenting the students' own story followed by two stories which were unfamiliar. Following research on the impact of topic familiarity on performance (e.g. Bui & Huang, 2016), and discussions with researchers in the field (de Jong, 2013, personal communication; Bygate, 2013, personal communication) it was felt that this provided a potentially large source of influence which might impact on results. Specifically, any reduction in fluency (and complexity and accuracy) on the second and third performances for the PR group could be attributed to a lack of topic familiarity and not the nature of the repetition.

In order to overcome this difficulty, it was therefore necessary to analyse students' performances in *both* groups only on stories with which they were not familiar with beforehand (i.e. not using the storyboard that they had created themselves). This meant that the procedure for the TR carousel had to be modified slightly from Lynch and Maclean's (1994) template, to include a 'dummy' stage. Specifically, this meant that, for the TR group, the ‘dummy’ stage involved the original authors of each poster performing the task to a visitor. Each visitor would then ‘take charge’ of the new poster (which they had not authored themselves) and would present it three times to three different visitors.

The PR procedure also included a ‘dummy’ stage, which also involved the original authors of each poster performing the task to a visitor.

For both TR and PR groups, the very first performance, in which the original authors presented their own poster, was purposely not analysed. This meant that all analysed performances were based on stories that students had originally heard from another learner, rather than from the written material provided. This modification has two methodological implications:

Firstly, the result of the ‘Dummy’ stage is that half of the participants in both the TR and PR groups told their own story before they told the three stories which were analysed. Of course this could be argued to have some impact on performance in terms of inadvertent procedural repetition. However, the proportion of students that presented their own poster before the three which were analysed in each group was the same and this potential source of influence is constant between groups. Unfortunately, this was an inevitable compromise which had to be made in order to introduce this element of control into the task sequences and allow for a comparison between the two types of repetition to be made.

Secondly, all analysed performances were based on oral performances that were preceded by listening to a fellow student perform the same story. It can therefore be considered ‘aural/oral’ task repetition (Lambert et al., 2017) and is also similar to studies such as Arslanyilmaz and Pedersen (2010), in which students are presented with a video of native speakers performing a similar task before they perform the task themselves. In other words, the input for this task comes in the form of an interlocutor’s narrative. Since this is a characteristic of the procedure in both experimental groups, it was not considered problematic.

8.5.3 Role of the teacher

The stimulus material was not abridged in any way but the researcher and class teacher were on hand to offer explanations and pronunciation of new vocabulary during the reading of the input materials and creation of the storyboards. English language dictionaries were also provided to the learners. Training sessions were facilitated by the researcher along with the class teacher. Facilitation involved giving instructions for tasks, providing input during creative stages (but not during the performances themselves), providing clarification when necessary and ensuring that students completed the tasks (i.e. ensuring voice recorders were switched on etc.)

Following the presentations there was some general group feedback which involved deciding upon the most interesting story and best storyboard. In contrast with N. de Jong and Perfetti (2011), no explicit feedback was given on individual performance and students did not listen to their recordings. This was to eliminate additional sources of influence.

8.5.4 Number of repetitions

The analysis of the session data was based on three task performances. Although TR studies have varied considerably in respect to the number of repetitions from a single repetition (e.g. Wang, 2014) to eleven repetitions of the same task (Ahmadian, 2011), it was felt that even the most keen students would eventually tire of repeating a task (Bygate, 2001). Through careful piloting, it was established that three performances struck a balance between potential repetition effects and what learners would tolerate before becoming bored, as well as what was allowed within the timeframe afforded by the school's timetable. Three to four performances was also reported to be sufficient for the students in Lambert et al. (2017) to experience gains in different aspects of oral fluency and was described as the optimum number of repetitions by the participants in that study. In addition, using three iterations would allow for comparisons to be drawn between other repetition studies which similarly looked at three performances (e.g. N. de Jong and Perfetti, 2011; Sample & Michel, 2014; Boers, 2014).

8.6 Pre- and post-tests and procedure

For the pre- and post- tests, two IELTS oral exam practice questions were selected from the IELTS 5 student workbook (CUP, 2006). There are a number of reasons why IELTS oral exam questions were selected. Firstly, using IELTS exam questions was useful because this type of question is designed to elicit spontaneous speech. Speaking spontaneously is what students are required to do in their lives and therefore it is an authentic and ecologically valid test of their speaking ability (Segalowitz, 2010). Secondly, this approach would allow for comparisons to be made between this study and N. de Jong and Perfetti (2011), who used similar pre- and post-test questions. Finally, and although the training sessions did not take place during an exam class, the majority of the students involved in the research were taking IELTS classes in the afternoon and were planning to take the IELTS exam at some point during their studies. It was therefore felt that this type of exercise would be of greatest practical use for the learners themselves. The two questions used in the study can be found in Figure 10 below.

Question 1 - Describe a wedding or party you have been to and which you enjoyed. Talk about where it was, what you wore and why you enjoyed it

Question 2 - Describe your favourite holiday. Talk about where you went, what you did and why it was your favourite holiday.

Figure 10: Pre- and post-test questions

These specific questions were chosen because they would elicit the same form (narrative tenses) as the session tasks and might therefore be more amenable to transfer (N. de Jong and Perfetti, 2011)

The procedure for the pre- and post-test was as follows: First, students were grouped into pairs and assigned a letter (A/B). The ‘A’s were given the slip of paper with the question 1 and then had two minutes to prepare. They were allowed to make basic notes during this time, although most chose not to. When they were ready their partners switched on the recording device and the ‘A’s answered the question while the ‘B’s recorded them. The ‘B’s were instructed to listen actively but not to ask any additional questions until the end of the recording. Students were given as much time as they liked to answer the questions in order that the performance was not ‘pressured’ which may have introduced an additional source of influence. Most students had finished speaking after two minutes. The process was then repeated for the ‘B’s who were given question 2. Pre- and post-test questions were therefore naturally counterbalanced because half of the participants (i.e. the ‘A’s) responded to question 1 during the pre-test and the other half responded to question 2 (i.e. the ‘B’s). This pattern was then reversed for the post-test.

8.7 Student focus groups

The purpose of conducting focus groups with the learners was to elicit their opinions about the two task procedures. Previous studies which have elicited students' opinions about task repetition have employed a range of methods. Most recently, Lambert et al. (2017), Kim (2013) and Hunter (2011) used a questionnaire in order to gauge student opinion about task repetition (see Chapter 4 for an overview of data collection methods) while Ahmadian et al. (2017) used semi-structured interviews.

An oral discussion (focus group) was felt most appropriate, partly due to the fact that the research was taking place during a 'Speaking and Listening' class and was therefore the best fit with the expectations of both the school and learners, and partly because it was felt that the focus group data would support the existing questionnaire data already elicited by Richards (2011) and offer a more qualitative perspective. A focus group setting was selected partly for reasons of timing because feedback had to be collected during a single 50 minute session that also included the post-test data collection and also because it was felt that students might feel more confident to give their true opinion with the support of their peers as opposed to talking individually with the researcher.

Focus groups took place after the post-tests on the final day of the study. The focus groups involved each intact class and the class teacher. Prompt questions were provided by the researcher to stimulate discussion. The questions provided were: *What did you think about the activities you took part in this week? Was there anything you liked about the activity? Was there anything you didn't like about the activity? Is there anything you want to change about the activity? Why?* In addition, the researcher responded naturally to the students' responses but tried not to lead the discussion in any particular direction. It was explained to students that they could speak freely, and that all their feedback (positive and negative) would be useful.

8.8 Analysis of data

In this section, I explain how the design of my study meets the methodological requirements of the research questions. The present study addresses three main research questions. The first is the *short-term impact* of the two different carousel task sequences (RQ 1). The second investigates the extent to which the two different carousel sequences move learners to speak with higher fluency on a completely new task (RQ 2). The third concern is with *L2 learners' perceptions* of the two carousel task sequences (RQ 3).

Research Question 1: The short-term effects of TR and PR carousels

In order to investigate the short-term impact of the TR and PR carousels, data collected during training session 1 (day 2) was analysed. The aim was to investigate whether L2 learner's oral fluency changed significantly over the course of three performances elicited by the TR and PR carousels respectively, as well as to investigate the impact on complexity and accuracy over the same period. As such, the focus of analysis was the oral data collected during the first (day 2) intervention[15]. This included data from the TR (n=24) group and PR (n=22) group (see Table 25 below).

Table 25: Data involved in short-term fluency analysis

	Pre-test (day 1)	**Session 1 (day 2)**			**Post-test (day 5)**
TR (*n*=24)	1	P1	P2	P3	2
PR (*n*=22)	1	P1	P2	P3	2
C (*n*=18)	1	-			2

Note: P = performance

For this Research Question, there were two independent variables. The first independent variable is the within-subjects variable *time*, for which there are three levels for each subject, representing the three performances of the task. The second independent variable is the between-subjects variable *group,* which separates those students who repeated the same task (TR carousel group) and those who performed three similar tasks (PR carousel group) (Table 26).

Table 26: Independent variables for short-term analysis

	Type of variable	**Number of levels**

[15] As discussed above, data collected during the second (day 3) and third (day 4) iterations were purposely not analysed as part of this study

Time	Within-subjects	3 (Performances 1, 2 and 3)
Group	Between-subjects	2 (Task Repetition (TR) and Procedural repetition (PR))

Research Question 2: The fluency transfer effects of TR and PR carousels

The aim of the second research question was to investigate what effect, if any, the TR and PR carousels have on fluency on an entirely different task, relative to a control group. In order to do this, the oral data which was collected during the pre- and post-tests (i.e. day 1 and day 5) were analysed (Table 27).

Table 27: Table highlighting data involved in fluency transfer analysis

	Pre-test (day 1)	**Session 1 (day 2)**	**Post-test (day 5)**
TR (n=24)	1	P1 P2 P3	2
PR (n=22)	1	P1 P2 P3	2
Control (n=18)	1	-	2

Note: Pre- and post-test questions were counter-balanced

For RQ2 there are two independent variables. The first independent variable is the within-subjects variable *time*, which relates to the performances at pre-test and again at post-test. The second independent variable is the between-subjects variable *group,* which separates those students who repeated the same task during training sessions (TR group), those who performed three similar tasks during training sessions (PR group) and those who did not take part in any training sessions at all (Control group) (see Table 27).

Table 27: Independent variables for fluency transfer analysis

	Type of variable	**Number of levels**
Time	Within-subjects	2 (Pre-test and Post-test performances)

Group	Between-subjects	3 (Task Repetition (TR); Procedural repetition (PR); Control (C))

Research Question 3: Learners' perceptions of the TR and PR carousel techniques

The aim of the third research question is to ascertain what learners thought of the task sequences undertaken as part of this study, in order to guide pedagogic implications. To address this research question, data from the focus group sessions (with learners from TR and PR groups) were analysed. All focus group data was audio-recorded and then transcribed by the researcher. A thematic analysis was conducted on both the TR and PR focus group data, following Ahmadian et al. (2017). Following the procedure set out in Ellis and Barkhuizen (2005), the thematic analysis involved detailed note-taking by the researcher during the focus-groups themselves and during careful listening to the recordings. From these initial listenings, a list of relevant codes emerged. The recordings were then transcribed, and the written text read multiple times to identify additional codes. These were then collapsed into related themes based on frequency. For example, codes related to 'speaking better' and 'improving in English' were collapsed into a theme that was to with 'improvements'. Themes therefore emerged if a particular code was activated multiple times or if a number of different codes collapsed into a theme.

8.9 Dependent variables

In order to answer Research Questions 1 and 2 (i.e. what are both the *short-term* and transfer fluency effects of TR and PR?), a number of dependent variables are used to measure L2 learners' *fluency*. In addition, a small number of dependent variables were used to measure learners' *complexity* and *accuracy* in order to investigate the possible short-term impact of the carousels on these dimensions of performance. An overview of these measures is provided in Table 28 below.

In the following section each of the dependent variables is introduced and I give a rationale for its inclusion in the current study. In addition, and in response to numerous calls in the literature (such as Ellis, 2009; Norris & Ortega, 2009) for more precise explanations of performance measures, I provide detailed information on how these measures will be calculated.

Table 28: Overview of dependent variables used in this study

Dimension	Aspect	Measure	Calculation
Fluency	**Global**	Speech rate	Total number of syllables produced divided by total time taken to produce the sample (incl. pauses)
		Mean length of run	Total number of syllables divided by total number of 'runs' (i.e. stretches of speech uninterrupted by pauses >250ms)
	Speed	Articulation rate	Total number of syllables produced divided by total 'speaking time' (i.e. total sample time minus all pauses >250ms)
	Breakdown	Phonation time ratio	Total speaking time divided by time taken to produce the sample (incl. pauses) multiplied by 100
		Frequency of mid-clause pause (filled, silent and 'composite')	Number of mid-clause pauses >250ms divided by time taken to produce the sample, multiplied by 60
	Repair	Frequency of reformulations and self-correction	Total number of reformulations and self-corrections divided by time taken to produce the sample multiplied by 60
	Holistic	Holistic native-speaker rating	Mean score of the two NS raters
Complexity (RQ1 only)	**Syntactic**	Amount of subordination	Total number of clauses divided by total number of AS-units
		Length of clause	Total number of syllables divided by total number of clauses
		Length of AS-unit	Total number of syllables divided by total number of AS-units
Accuracy		Weighted clause ratio	Mean accuracy-per-clause score

8.9.1 Fluency

Accurately measuring learners' fluency is far from straightforward. Rather than relying on one or two measures, this study uses a total of seven dependent variables in order to build up a comprehensive picture of fluency changes in L2 Learners' performance. The measures which have been identified as the most reliable indicators of perceived fluency (see Chapter 2) do not always fit neatly into one single category. Instead, commonly-used measures of fluency, and those that correlate highly with perceived fluency, such as *speech rate* and *mean length of run* are better described as 'composite' or 'global' measures because they encompass all three aspects of oral fluency (Witton-Davies, 2014; Lambert, 2017, personal communication). A measure like speech rate, for example, will be affected by speed of speech, by amount of pausing and also by tendency to repair.

At the same time, and as has previously been explored in Chapter 2, there is a general agreement among SLA scholars that the most important aspects of fluency are *speed, breakdown* and *repair* (Skehan, 2003). In addition, a fine-grained analysis is essential to understanding particular cognitive processes involved in speech production. In other words, linking specific utterance phenomena to stages of speech production (Lambert et al., 2017). As such, it is imperative to use measures that reflect each of these.

A final point is that, in light of the second overarching aim of this study (to provide pedagogic guidance), it was felt that the impact of the TR and PR carousels should also be measured in a way that might be replicable in language *classrooms* and by language *teachers.* For that reason, a holistic rater score of fluency was also obtained which would reflect the way that language teachers might perceive the fluency of these samples of speech.

Table 29: Fluency measures used in task repetition studies to date

TR Study	Fluency Measures adopted
Ahmadian & Tavakoli (2010)	Raw speech rate Pruned speech rate
Ahmadian (2011)	Raw speech rate Pruned speech rate
Boers (2014)	Speech rate (words and syllables per minute)

	Number of hesitations
Bygate (1996)	Number of repetitions
Bygate (2001)	Unfilled pauses per T-unit
Hsu (2017)	Speech rate (pruned syllables per minute)
Kim & Tracy-Ventura (2013)	Speech rate (syllables per minute) Frequency of reformulations
Lambert et al. (2017)	Pruned speech rate Frequency of filled pauses Ratio of end -clause pause to number of syllables Ratio of mid-clause pause to number of syllables Frequency of overt self-repairs
N. de Jong & Perfetti (2011)	mean length of fluent run phonation time ratio mean length of pause articulation rate
Sample & Michel (2014)	Filled pauses per minute Words per minute (per pair) Time to complete the task (per pair)
Thai & Boers (2016)	Speech rate (words and syllables per minute)
Wang (2014)	Speech rate (words per minute) Mean length of end-AS pauses Mean length of mid-AS pauses Number of reformulations

In sum, in order to develop an in-depth account for changes in L2 fluency that is useful and relevant for both researchers and teachers, it was deemed necessary to look at both *global* measures of fluency and more *specific* measures which capture particular aspects of fluency (speed, breakdown, repair), as well as *holistic* measures.

Global fluency measures

For this study, global utterance fluency was calculated in terms of both *Speech rate* and *Mean length of run.* Speech rate (e.g. Lambert et al., 2017) and mean length of run (e.g. N. de Jong & Perfetti, 2011) are commonly used measures in TR studies (see Table 29 above) which correlate highly with subjective ratings of fluency (e.g. Kormos & Denes, 2004; Derwing et al., 2004) and were therefore considered appropriate for this study as global measures of fluency.

Speech Rate

Speech rate tells us about both a person's speed of delivery and the extent to which they pause. In the TR literature, studies are divided in terms of whether they calculate speech rate as *syllables* per minute or *words* per minute. In a few CAF studies (e.g. Fukuta, 2016) pruned *words* per minute have been used as an overall fluency measure. The present study opted for a *syllables* per minute rate of speech because, as TR is believed to increase complexity (and therefore, possibly, word length), counting number of words might mask increased fluency due to a higher number of multi-syllabic words being uttered during second and third task performances. Syllables per minute, therefore, provided a more stable measure that would be unaffected by word length.

Studies of TR are also divided in terms of whether or not they rely on a 'pruned' syllable (or word) count or an 'unpruned' count. Some studies do not report whether their calculations are based on pruned or unpruned speech. 'Pruning' in the L2 speech analysis sense involves excluding from analysis words or syllables which are associated with repair mechanisms (i.e. repeating words/sounds, reformulating utterances, self-correcting). By pruning L2 speech, the researcher is able to perform calculations which reveal the extent to which a speaker communicates a particular message in real time. In order to distinguish between speech that is quick and speech which is quick *and* conveys the intended message effectively, *pruning* is a useful tool. Furthermore, it is *pruned* speech rate which seems to correspond most often with subjective scores of fluency (Kormos & Denes, 2004), presumably because raters are able to differentiate between speech which is fast and speech which is fast *and* coherent.

In the present study, speech rate was calculated as the total number of pruned syllables produced in the speech sample[16] *divided by total sample time and multiplied by 60.*

[16] Although students were given two minutes for each task performance, only the first minute was analysed. This was because some students stopped talking before the two minutes were up. In order to have a consistent amount of data across participants, it was therefore necessary to reduce the amount of the sample that was analysed.

Mean length of run

Mean length of run is another commonly used indicator of utterance fluency. It is typically measured in terms of the average number of syllables (or words) that are uttered between two pauses (i.e. a 'run' of speech). As well as providing insight into the speed of speech, mean length of run also gives an insight into the patterns of speaking; how much content can be packed into a run between pauses. It also provides a means of incorporating pause position into a global measure because longer runs will tend to mean that pausing is not happening as frequently in the middle of clauses.

It is important to be aware that both speech rate and mean length of run will be directly affected by the way that a study defines 'pauses'. This is because pauses mark the beginnings and ends of runs. Therefore, if a pause is defined as being any silence which is longer than 200ms, this will result in shorter runs than a study which defines pauses as being a minimum of 250ms. Similarly, some studies will define pauses as being silent *or* filled pauses but others will focus only on silence. Again, this will impact on mean length of run. This is crucial to bear in mind when comparing research findings across studies. Pauses as short as .2 of a second have been used in some L2 fluency studies (Hilton, 2014; N. de Jong & Perfetti, 2011) while much longer pauses have also been defined in studies. In the current study, .25s was chosen following N.H. de Jong et al. (2012) who treated minimum pause length as an empirical question. They found that 'fluency' defining pauses are typically between .25 and .3 of a second.

In this study, mean length of run was calculated by dividing the total number of pruned syllables produced in the speech sample by the total number of runs produced.

Specific fluency measures

This study used four specific measures of utterance fluency, namely *articulation rate* (to measure speed), *phonation time ratio* and *frequency of mid-clause pause* (to measure breakdown) and *frequency of reformulations and self-correction* (to measure repair). The reasoning behind the choice of these particular measures is set out below.

Speed fluency

Following N. de Jong and Perfetti (2011), speed fluency was measured by calculating the *articulation rate* of participants' speech. Because the purpose of using a specific measure of 'speed' is to isolate speed as a variable, it is necessary to use a raw syllable count, in order to avoid conflation with 'repair' fluency and also to use only speaking time rather than the total sample time to avoid any overlap with breakdown fluency. Some have argued that articulation rate only represents the speed of articulatory processes (N. de Jong & Perfetti, 2011), however, it also provides a gauge of the extent to which a speaker 'buys time' while speaking through lengthening syllables (i.e. 'drawls', Hilton, 2008). In other words, speakers may 'pause' by using sound elongation in a similar way to how they might use a filled pause such as 'um' or 'er'. Some studies (e.g. Lambert et al., 2017) include these sorts of sound elongations in their frequency counts of filled pause, although they do not provide any details of a minimum length of drawl requirement.

A final point about articulation rate is that a few L2 fluency studies prefer to use the term: *inverse articulation rate* (e.g. N.H. de Jong, 2016). This measures the same phenomena as articulation rate but inverses the formula for calculation, providing a 'mean length of syllable' measure as opposed to a 'syllables per minute' measure. However, in order to remain consistent with N. de Jong and Perfetti (2011), articulation rate was used in the present study.

Articulation rate was calculated as the total number of raw syllables produced in one minute divide by the total *speaking* time (excluding all pauses >250ms) and multiplied by 60 to provide a syllables per minute measure. Raw syllables includes all uttered syllables including partially uttered words, repaired utterances, short filled pauses and epenthesis divided by total speaking time.

Breakdown fluency

In line with Hilton (2014), breakdown fluency measures selected for this study were: *phonation time ratio*, and *frequency of mid-clause pauses*. In contrast to speed fluency, for which there is arguably only one potential measure, there is a huge selection of potential measures for breakdown fluency. It is therefore of paramount importance that breakdown measures are selected for inclusion in a study in a systematic and informed way.

Many TR studies include some kind of 'breakdown' fluency measure but they vary enormously among studies and often there is little rationale provided for a particular breakdown fluency measure over another. Potential candidates for breakdown fluency include, for example: phonation time ratio, frequency and/or length of filled pauses, frequency and/or length of silent pauses. As well as using different measures to tap breakdown fluency, studies often differ in terms of how they define pauses (e.g. minimum length; pause type)

A frequently-used measure of breakdown fluency is 'phonation time ratio' (PTR). PTR gives a gauge of the proportion of time which is spent speaking as opposed to pausing. It is often presented as a percentage of phonation time. Kormos and Dénes (2004) found that phonation time ratio was a good measure of fluency in terms of reflecting native speaker judgements. It has also been found to correlate strongly with proficiency (Ginther et al., 2010; Tavakoli et al., 2017). PTR does not discriminate between pause *length* and pause *frequency.* In effect, it simply gives a measure of *amount of pausing.* Although this may seem rather blunt an instrument, it is inclusive and therefore useful for capturing changes in pausing over time for groups of learners who may adjust their pausing behaviour in different ways. Phonation time ratio was therefore selected for use as a breakdown fluency measure in this study because it would provide a useful 'overview' to pausing behaviour.

Phonation time ratio was calculated by dividing total speaking time by total sample time and multiplying the result by 100.

Frequency of mid-clause pause was selected for this study because it was felt that, based on current understanding of pauseology in L2, this was the measure most likely to be indicative of L2-specific breakdown fluency. In terms of establishing more specific indicators of breakdown in L2 speech, a number of distinctions are often made. The first pause-related distinction which divides the field of L2 fluency is whether to consider pause *length* or pause *frequency*. Some researchers (e.g. Wang, 2014) consider mean *length* of pause to be most reflective of breakdown fluency while others (e.g. Lambert et al., 2017) consider pause *frequency.* Bosker et al. (2014) and Hilton (2014) have suggested that it is pause *frequency* that is more indicative of cognitive fluency. This is likely related to the fact that pause *position* also plays a part in fluency perception therefore increased frequency of pause means that pauses are more likely to come in the middle of clauses which might be considered more disruptive. Furthermore, Préfontaine et al. (2015; 2016) found that listeners to L2 French were tolerant of even very long pauses, provided they came before a suitably long run of fluent speech. In other words, it seems that long pauses can be 'earned'.

Another question is whether filled or silent pauses are more important in determining L2 breakdown fluency. Some TR studies only consider filled pauses (e.g. Lambert et al., 2017) while others only consider silence, others combine both silent and filled pauses (e.g. N. de Jong & Perfetti, 2011). Arguably, both filled and silent pauses perform a similar function, they allow the speaker time to deal with processing bottlenecks (Kormos, 2006). The main difference, however, is that a filled pause demonstrates a move to 'maintain the floor' when speaking (i.e. it is more difficult to interrupt a person who is saying 'er' than it is to interrupt someone who has stopped speaking altogether). Filled pausing therefore possibly reflects greater awareness of cultural and societal norms of oral interaction. As such, it may be related to perceived fluency. A final point in relation to filled and unfilled pauses is that, often, both filled and unfilled pauses will cluster together in both L1 and L2 speech (Hilton, 2014). However, no TR studies have discussed how they deal with 'clusters' of pauses, whether they count them as separate pauses or whether they amalgamate them into a 'hesitation group' (e.g. Hilton, 2009; 2014; Roberts & Kirsner, 2000; Campione & Veronis, 2005). If we believe that both filled and unfilled pauses perform similar functions in speech (i.e. allowing the speaker time to plan content/language) then pauses between runs which *combine* silence and filled pauses should be added together as a hesitation cluster. Composite pauses of this kind make up a large percentage of the pauses in the current data.

L2 fluency studies are increasingly making a distinction between pauses that come in the middle of a clause/speech unit and pauses that come at the end of a clause/speech unit. This is in response to research by a number of researchers who have identified that it is not just *amount* of pausing that is indicative of breakdown fluency but also the *distribution* of pausing behaviour (Davies, 2003; Towell, 2002). Swerts (1998) found that, in L1 speech, speakers are more likely to pause at syntactic boundaries which supports the theory that native speakers construct speech one clause at a time, planning the upcoming message, without having to worry about the structure of the current message (Pawley & Snyder, 1983). Rianzantseva (2001), Skehan and Foster (2007), Tavakoli (2010) and more recently de Jong (2016) have investigated the difference between L1 and L2 pause distribution. All these studies found that L2 speakers are more likely to pause within AS units (de Jong, 2016; Skehan & Foster, 2007), clauses (Tavakoli, 2010) and constituents (Riazantseva, 2001). De Jong (2016) suggests that “(w)ithin ASU’s, L2 speakers more often run into trouble while formulating the linguistic message than L1 speakers do and are therefore more likely to pause, for longer durations than L1 speakers do. The pauses within ASU’s that L2 speakers use..are due to less L2 knowledge and lower L2 skills.” (de Jong, 2016, p. 18). Given the compelling evidence, the current study takes the view that “pauses between utterances are not informative because they reflect conceptual planning. It is only the pauses that occur within utterances that are indicative of L2 proficiency and are therefore informative for distinguishing on the basis of aspects of fluency.” (de Jong, 2016, p. 18).

A final point in relation to pause measurement in L2 is that, very recently, de Jong (2016) and de Jong and Bosker (2014) provide evidence that L1 and L2 speakers differ in terms of what their pauses *precede*, with L1 speakers pausing before less frequent words, presumably due to lexical retrieval speeds. While it might very well be interesting to explore the impact of task repetition on what pauses *precede,* it was considered beyond the scope of the current study to analyse this particular feature.

Frequency of mid-clause pause was calculated by dividing the total number of mid-clause pauses (filled, silent and composite) by the total sample time and then multiplying the result by 60. Pauses were defined in the present study as either filled, silent, or composite[17] *(clusters of filled and unfilled pauses) pauses which were longer than 250ms in duration.*

Repair fluency

Repair fluency was measured in terms of the *Frequency of reformulations and overt self-corrections.* While there is a certain degree of agreement about the reliability of common predictors of global, speed and breakdown fluency, the picture seems less clear when it comes to repair. In most TR studies, overt repairs are counted and a frequency is calculated (Bygate, 1996; Kim & Tracy-Ventura, 2013; Boers, 2014; Wang, 2014) however a number of studies have found either that repair measures do not necessarily correlate with, or that they have a relatively small impact upon (Bosker et al. 2012), subjective ratings of fluency (Cucchiarini et al., 2002; Tavakoli et al., 2017). It may be that different types of repair impact

[17] Following Hilton (2014)

differently on overall perceptions of fluency (Olynyk et al., 1990) or that there is a need to distinguish between error repairs, that is, repairing errors of linguistic form; and appropriateness repairs (i.e. presenting a new or rephrased message) (Levelt, 1983; Kormos, 1999; Kormos, 2006, chapter 6).

Another important distinction is between overt and covert repair (Kormos, 2006). While the former can be measured by analysing speech which is produced, the latter can be analysed through the use of 'retrospective comments' on the speech (Kormos, 2000a; 2000b) but is often too involved a process to be carried out in TR studies (Lambert et al., 2017).

In Lambert et al. (2017), the researchers worked with a particular type of repair, overt self corrections, which is an attractive option as it allows the researcher to look at a particular type of repair in detail and has more theoretical validity (i.e. there are clear links to cognitive fluency – monitoring processes, Kormos, 1999). The problem with this measure, however, is the low incidences of such a specific phenomena. The result is a dataset with many zeros or perhaps one observation in the entire sample. An exploratory look at self-corrections in the current data set revealed an average of 2 self-corrections in the first performance of the task with almost a quarter of the participants producing no overt self-corrections whatsoever. Certainly, it is debateable how much such an infrequent indicator can tell us about a group's fluency.

Very often, when researchers want to capture 'repair' fluency in speech analysis, they bundle a number of so-called 'repair behaviours' together and calculate the total frequency. These studies also often include verbatim repetitions. For example, if a speaker says "the woman goes to goes to the shop", this would be counted as an incidence of repair. However, as Dörnyei & Kormos (1998) explain, this verbatim repetition is probably better thought of as an example of a 'stalling behaviour' and would therefore be more indicative of *breakdown fluency*. For this reason, verbatim repetitions were not included in the repair frequency measure which is used in this study (see also Witton-Davies, 2014 for a discussion of the use of repetition as a repair measure).

In the present research, frequency of reformulations and overt self-corrections (combined) was calculated as the total number of reformulations plus total number of overt self-corrections divided by total sample time and multiplied by 60.

Holistic fluency judgement

As well as the global and specific measures of fluency, and in order to triangulate fluency measurement and explore how far utterance fluency represented human judgement of fluency, fluency for all participants on all performances was also assessed *holistically* by two native speaker raters.

There is little TR research available on which to model holistic fluency rating as the only study which has used holistic judgement on TR performance was Gass et al. (1999). In their study, the raters were asked to judge samples of speech using a 'magnitude estimation' technique. This allowed raters to give not only an indication of how 'good' a particular performance was but also *how much better* it was when compared to other performances. Although, there is clearly a strong rationale for this approach to fluency judgement, it was not considered for the current study on the grounds that it would be too involved and too labour-intensive for the busy teachers who had agreed to rate the samples.

The two raters were experienced English language teachers who were recruited to take part in the study on a voluntary basis. Both teachers held certificates in teaching English to adults (CELTA) and had more than 5 years teaching experience, which involved working with learners from a very diverse range of L1 backgrounds. One was male and one was female.

In the current study, the two raters were asked to judge each sample on a 9 point scale which drew on the IELTS oral marking rubric (see appendix). The reason for using such a rubric was to limit the raters' judgements to a narrower definition of fluency and to increase generalisability of the findings. They rated samples over a period of a few days and were given regular breaks in order to reduce the possible effects of rater fatigue. Samples were presented to the raters in an entirely randomised order. Interrater agreement was 80% which, considering the limited instructions provided to raters, as well as the subjective nature of fluency judgement, can be considered adequate. For example, this degree of agreement is the same as that which was found between raters in Kormos and Dénes (2004). Scores awarded by each rater were then combined to provide a combined rater score for each participant for each performance and it was this adjusted score which was used for the statistical analyses which are presented in the following chapter.

8.9.2 Complexity and Accuracy

As part of the first research question, this study is concerned with what interaction exists, if any, between short-term fluency gains and complexity and accuracy scores. Similarly to fluency, complexity and accuracy are increasingly being viewed as multi-dimensional constructs (Housen, Kuiken & Vedder, 2012). This means that TR studies often adopt multiple measures in an attempt to 'tap' complexity and accuracy (see Table 30 below). It becomes evident that almost all complexity and accuracy measures depend on some kind of syntactic unit. Foster et al. (2000, p. 354) explain: "The analysis of spoken language requires a principled way of dividing transcribed data into units in order to assess features such as accuracy and complexity. If such analyses are to be comparable across different studies, there must be agreement on the nature of the unit, and it must be possible to apply this unit reliably to a range of different types of speech data".

In most cases the unit used for measurement is the *AS-unit.* The AS-unit was developed by Foster et al. (2000) in order to respond to the lack of agreement on how to divide L2 production into analysable units. The AS-unit is defined as "a single speaker's utterance consisting of *an independent clause or sub-clausal unit,* together with any *subordinate clause(s)* associated with either" (Foster et al., 2000, p. 365). See Chapter 11 for examples of orthographic transcriptions divided into AS-units.

Table 30: Complexity and accuracy measures used in TR studies to date

Study	Complexity Measures	Accuracy measures
Ahmadian & Tavakoli (2010)	Ratio of clauses to AS-unit Total number of grammatical verb forms	% of error-free clauses % of accurate verb usage
Ahmadian (2011)	Ratio of clauses to AS-unit Total number of grammatical verb forms	% of error-free clauses % of accurate verb usage
Boers (2014)	Number of subordinate clauses	Number of Errors

	Number of lemmas > 2k greatest frequency	
Bygate (2001)	Words per *t*-unit	Errors per *t*-unit
Hsu (2017)	Clauses per AS-unit Mean length of AS-unit (pruned words) Type-token ratio – D	% error-free AS-units % error-free clauses Accurate use of verb forms
Kim & Tracy-Ventura (2013)	Clauses per AS-unit Proportion of 'complex' AS-units Type-token ratio – D	Error-free AS-units Error-free clauses
Lambert et al. (2017)	N/A	N/A
N. de Jong & Perfetti (2011)	N/A	N/A
Sample & Michel (2014)	Clauses per AS-unit Words per AS-unit Type-token ratio – D	Error-free clauses per AS-unit Agreement errors per AS-unit Article errors per AS-unit Other errors per AS-unit
Thai & Boers (2016)	Mean clauses per AS-unit Lexical sophistication (beyond 2000)	Ratio of error-free clauses to total number of clauses
Wang (2014)	Total words Mean length of AS-unit Amount of subordination Type-token ratio – D	Total number of error-free clauses Error-free clause rate

Complexity

Complexity is most often discussed in SLA studies in terms of lexical complexity and/or syntactic complexity (see Skehan, 2009 and Ellis & Barkhuizen, 2005). Lexical sophistication was not measured in the current dataset due to the fact that the comparison group spoke about three different stories while the TR spoke about the same story three times. It was therefore probable that any change in lexical sophistication that the TR group could make above and beyond that of the PR would be inextricable from difference which were purely due to the different linguistic content of the tasks.

Instead, it was considered appropriate to focus solely on syntactic (i.e. grammatical) complexity, as even though vocabulary was expected to differ between performances for the PR group, the syntactic requirements (narrative) were stable across performances.

This study uses three basic measures of syntactic complexity, *amount of subordination, length of clause* and *length of AS-unit*. This follows Norris and Ortega (2009), who argue convincingly that, as a minimum, researchers should consider these three complementary measures of syntactic complexity. Firstly, they argue that there should be a measure which taps overall or 'global' complexity in terms of length of utterance (e.g. words/syllables per AS-unit). Secondly, they suggest the use of a measure of subordination (e.g. clauses per AS-unit). Thirdly they suggest a measure which taps complexity at the intra-clause level (e.g. length of clause).

Ultimately, they argue that developing complexity is a more complex picture that can be explained using such measures as amount of subordination (see also Lambert & Kormos, 2014) as certain types of subordination, for example, adverbial clause coordination appears sooner than relative clause coordination. Failing to discriminate between *type* of subordination might therefore mask rather than reveal language development. Nonetheless, given that complexity was not the main focus of this study, the three complementary measures chosen here were considered sufficient to provide an exploratory picture of complexity shifts during the carousel task sequence.

Syntactic complexity measures are calculated as follows:

Amount of subordination (clauses per AS-unit) is the total number of clauses divided by the total number of AS-units.

Length of clause (syllables per clause) is the total number of syllables divided by the total number of clauses.

Length of AS-unit (syllables per AS-unit) is the total number of syllables divided by the total number of AS-units.

Accuracy
The accuracy measure used in this study was *a weighted clause ratio (*Foster & Wigglesworth, 2016). As Table 30 above demonstrates, most TR studies have operationalised accuracy as either the number or percentage of error-free clauses (e.g. Bygate, 1996) or the number of errors per clause/AS (e.g. Boers, 2014; Bygate, 2001). While this approach to accuracy measurement is relatively intuitive, these have been highlighted as problematic for a number of reasons (Lambert & Kormos, 2014; Pallotti, 2009). Firstly, accuracy and complexity are linked and as complexity increases, accuracy might decrease. Secondly, a clause may be inaccurate to differing degrees and one clause may only have a simple error which does not impede on comprehension while another may have multiple errors which renders the listener completely unable to extract the intended meaning (Foster & Wigglesworth, 2016). The current study, therefore, follows Foster and Wigglesworth (2016), by using a weighted-clause ratio to measure accuracy.

This is calculated by individually scoring each clause for accuracy (see following chapter), then calculating the mean score per learner per performance.

8.10 Case studies

To supplement and help to better understand the quantitative analysis undertaken, a small number (2) of case studies were also undertaken. The intention was to identify examples that could be considered typical of the wider sample, and to use data from analysis of each case study to explore possible explanations for results observed. The limited number of case studies is clearly a limitation, however, nonetheless, the intention is to provide illustration, where possible, of wider themes that emerge through analysis elsewhere. The cases were selected on a 'typical sampling' basis (Dörnyei, 2007), meaning that they were felt to be representative of the overall group scores.

8.11 Chapter summary

In this chapter I have outlined the design and procedure of the present study. I began with a brief explanation of the pilot study that was undertaken beforehand and which informed this work. In light of the observations of the pilot study, lessons from relevant literature, and the particular context in which this study was carried out, I set out details about the participants, materials, tasks and task procedures. I set out the design of the pre- and post-tests and the design of and rationale for the focus groups with learners. In terms of the design of the analysis process, independent and dependent variables were described, along with a rationale for the choices made and an explanation of how each measure is calculated. Finally, I described how a small number of case studies were identified. I turn now to a discussion of how the data were coded and analysed.

9 - Data Coding and Analysis

9.1 Introduction

In this chapter, I discuss the ways in which the oral data elicited during the training session and at pre-test and post-test were coded and analysed. I explain the precise ways in which the oral data was inspected, annotated and analysed using computer software and how the data were also transcribed *orthographically,* allowing for the analysis of accuracy and complexity measures. I illustrate particular techniques and procedures with actual examples from the data.

9.2 Recording

All pre- and post- tests and training session performances were audio-recorded using Sony ICDBX140 4GB Digital Voice recorders. Following issues with recording in the pilot group, recorders were locked into the record function before being handed out to students. This prevented accidental turning off, pausing etc. Students were also told to keep the recorders in front of their bodies and at chest height. The fact that students changed partners each time they performed, meant that detailed notes had to be kept by the researcher in order to accurately associate participants with their speech samples for the analysis stage. The recording and retrieval process would have been infinitely more straightforward had the same partner been used for each iteration (as in Hawkes, 2012) but the task sequence would no longer be genuinely communicative and any lack of gains in CAF could be linked to a disinterest in communicating (as was the case in Gass et al., 1999).

9.3 Temporal speech analysis and coding – PRAAT

I explained in Chapter 2 that a number of researchers use computer software to assist in analysing L2 speech. I also explained that automatic analysis of certain measures was growing in popularity. However, there were a number of reasons that automatic fluency analysis could not be carried out with the current data set. Chief among these, was the fact that the research was carried out in a language classroom with many learners speaking at the same time. This means that background levels of noise are too high for automatic detection of silence (which is a prerequisite in calculating fluency automatically). A second issue was that automatic fluency analysis is limiting in the sense that it can only calculate *certain* measures. For example, 'speech rate' can be calculated (de Jong & Wempe, 2009) but not *pruned* speech rate and not even a version of 'speech rate' that considers filled pauses to be pauses. To the best of my knowledge, there has been no PRAAT script developed which is able to automatically identify and measure the length of filled pauses (Hilton, 2014). It was therefore made clear that fluency measures would need to be calculated *manually.*

PRAAT also allows for semi-automatic or *automated* fluency analysis which involves detailed *manual* investigation and annotation of speech samples and automatic measurement of the *duration* of annotated phenomena. The drawback of this sort of analysis is that it involves a substantial investment of time for the researcher and also requires the researcher to develop their own computer script which will analyse the newly-annotated speech in the way that they require. However, if the study's aim of conducting highly scientific quantitative analysis within a normally-functioning language classroom was to be met, this was the only possible route to take.

In order to answer the research questions, pre-tests and post-tests as well as all the performances from the training session were analysed. These 254 task performances, totalling over 4 hours of speech were converted to .WAV format which is compatible with PRAAT. One by one, these recordings were opened in PRAAT and were listened to at the same time as the spectrogram was studied (Figure 11). The spectrogram is accompanied by a 'textgrid' which allows the researcher to annotate the speech sample. The analysis began at the first syllable uttered by the test-taker, be this of lexical content (e.g. 'My'), a filler (e.g. 'OK') or a non-verbal filler (e.g. 'um'). When identifying the beginnings and ends of runs of speech and pauses, the screen view was zoomed in to at most .2 of a second resulting in very precise measurement (see Figure 11). As the screen view zooms in, so the spectrogram provides additional information about speech phenomena such as pitch and intensity and this additional information makes identifying beginnings and ends of pauses easier.

Silent, filled and composite pauses were identified through repeated listening to small stretches of the recording accompanied by visual inspection of the spectrogram. Only silences, non-verbal fillers or combinations of both which totalled .25 of a second or longer were marked as pauses.

Pauses were manually marked as either silent, filled or composite pauses on the textgrid.

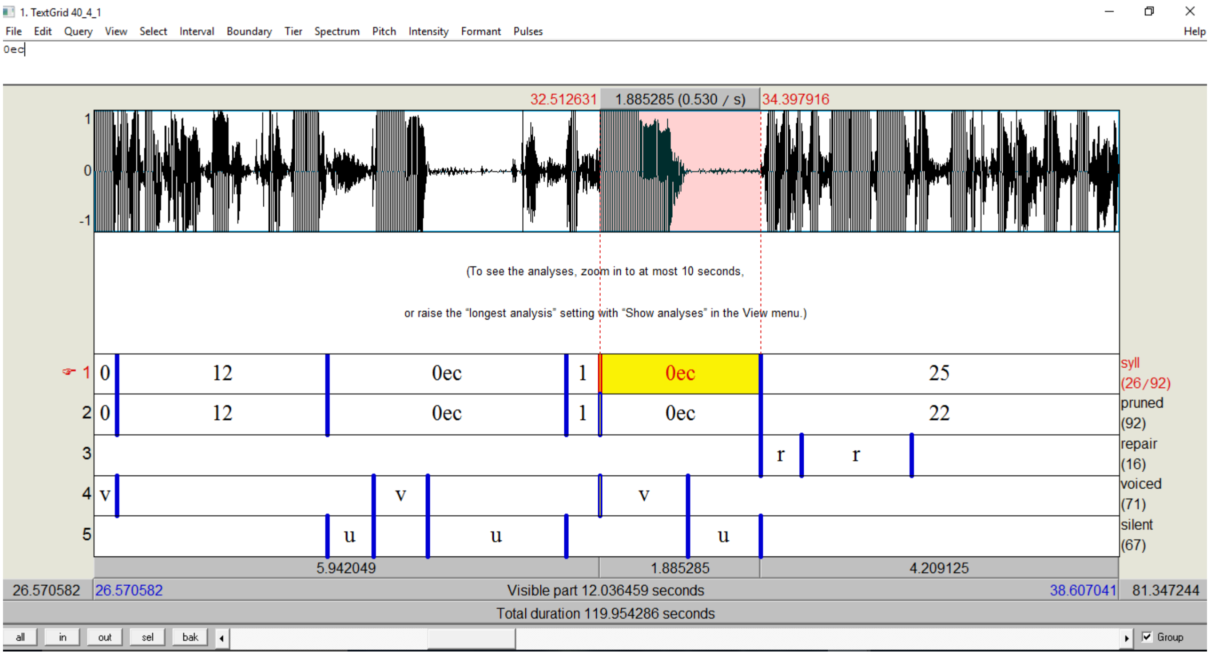


Figure 11: PRAAT soundfile and textgrid which illustrates 'composite' pausing

The beginnings and ends of all three types of pauses were marked against the spectrogram for the entire speech sample. This created alternating 'intervals' of speech and pause. Each pause interval was marked as either a filled, silent or composite pause.

Each pause was then studied again, this time to ascertain the pause position (mid-clause versus end-clause). This was done by careful listening to the recordings and examination of orthographic transcription which had been marked with clause boundaries. Information about clause position was also added to the pause intervals on the textgrid. Between these pauses are the stretches or 'runs' of speech generated by the test-taker. These were listened to and studied visually in order to manually count the number of syllables produced.

In most studies of L2 fluency, syllables are counted from orthographic transcriptions of the speech. It could be argued, however, that in spontaneous speech, especially that produced by language learners, syllables *uttered* does not conform with syllables *expected.* For example, in standard English ‘student’ is expected to have two syllables but a language learner, with epenthesis, may produce ‘*estudent’* which totals three syllables. N. de Jong and Perfetti (2011) state that “where there was doubt about the number of syllables pronounced (e.g., “every” can be pronounced as /ɛvri/ or /ɛv´ri/), the original recording was consulted” (2011, p. 545). My own experience of working with the data in the current study, however, revealed a huge number of discrepancies between anticipated and uttered syllables and these were not always *predictable*. A future study might address this issue empirically. Manual counting of syllables using the original recording therefore constitutes a more accurate (though by no means quick and simple) approach when compared to syllable counts using orthographic transcription. Single runs may be listened to multiple times in order to ascertain number of syllables which were then added to the textgrid.

Any non-verbal filler shorter than .25 sec was counted as a syllable along with partially uttered words, repetitions etc. Non-verbal phenomena such as laughter, coughing and throat-clearing was discounted from analysis altogether (i.e. it was not counted as a pause or part of a run). Any time spent laughing or coughing etc was also removed from sample time calculation which is used in the calculation of speech rate. However, these phenomena did mark the ends of runs. The number of syllables in each run was manually entered into the textgrid.

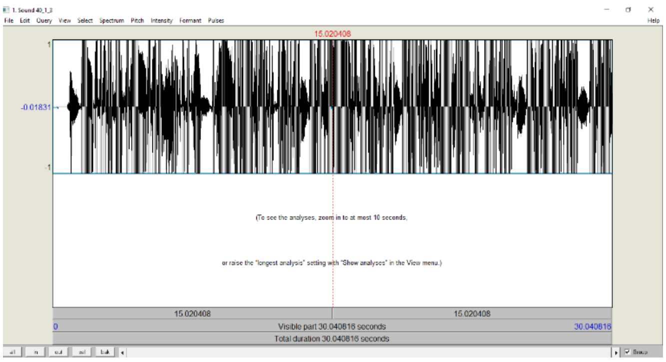


Figure 12: Spectrogram showing limited information

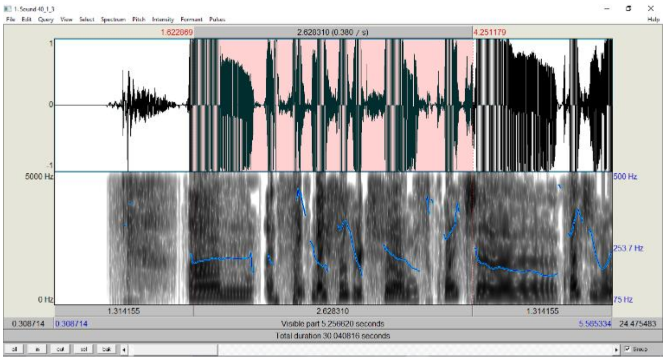


Figure 13: Zoomed-in version showing intensity and pitch contours

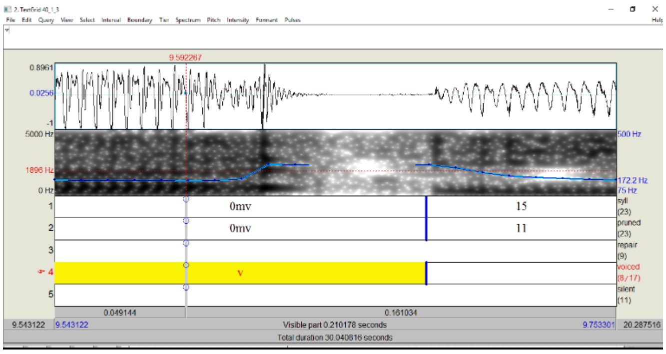


Figure 14: Spectrogram with accompanying 5-tier text-grid zoomed to 250ms

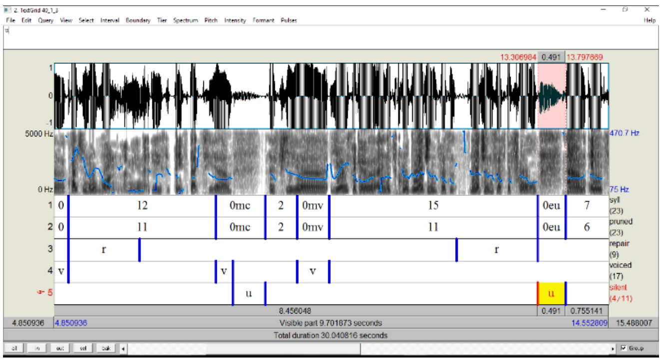


Figure 15: Spectrogram and 5-tier textgrid showing segments

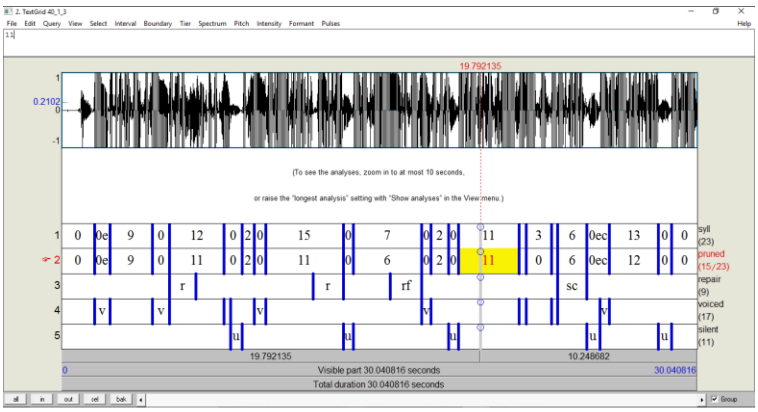


Figure 16: 30 seconds of annotated speech

As explained, the recordings in this study were cut off at one minute because this was the minimum length of speech which was produced by all participants on all tasks. This meant that often a test-taker's speech was cut off in the middle of a run of speech. Where this was the case, the analysis stopped at the previous run boundary and any subsequent pause or interrupted run was removed from analysis.

All recordings were analysed a second time. This time, reformulations and self-corrections were identified and marked on the textgrid. All recordings were analysed a third time. This time *pruned* syllables were counted and marked on the textgrid. Pruning involved discounting syllables which were any of the following:

1) Non-lexical fillers (*um; er*) shorter than 250ms
2) Syllables involved in repair (outlined above)
3) Lexical fillers (*well; you know*)
4) Epenthesis (e.g the word studio pronounced *estudio* – in this example the schwa would be pruned)

The 'pruned' syllable count was then added to the textgrid below the unpruned or 'raw' syllable count. The final result was a file which pairs the speech data with manually-annotated information about number of raw and pruned syllables, beginnings and ends of pauses, type of pause and position of pauses as well as incidences of repair (Figure 16).

When all 254 speech samples had been analysed in this way, a PRAAT script (see Appendix) was written which would generate the output necessary to calculate the various measures of overall, speed, breakdown and repair fluency. This script was run with each speech sample/annotated textgrid, generating 254 individual Microsoft excel spreadsheets which provided duration and frequency information about all the speech phenomena that had been annotated. These frequency counts and averages were then used to calculate the dependent variables of fluency as outlined in the previous chapter.

9.4 Orthographic transcription and coding

Having established that AS-units provided a good base for orthographic coding of accuracy and complexity measures (Chapter 8), the data were accordingly transcribed and then segmented into AS-units for analysis.

9.4.1 Accuracy

Global accuracy was measured using a weighted clause ratio. Accordingly, each clause produced by the participants was given an accuracy score based on the framework provided in Foster and Wigglesworth, 2016 (Figure 17).

Score	Definition	Example scoring in Foster & Wigglesworth (2016)	Example scoring from current data set
1.0	The clause is accurately constructed.	He is in the shop	that it's gonna be a hamburger
0.8	The clause has only minor errors (e.g., in morphosyntax) that do not compromise meaning.	It's man And when he open the bottle	and then the rescue the rescuer was swallowed by hippo but she only can move one arm and she bought some Christmas present
0.5	The clause contains serious errors (e.g., verb tense, word choice, or word order), but the meaning is recoverable, though not always obvious.	And he surprising Because the cork go down	and the baby opened the the eyes and suddenly s- a baby the baby fell down from building of window
0.1	The clause has very serious errors that make the intended meaning far from obvious and only partly recoverable.	And when he putting out of for the glass	so the car fly away and but first she can alive

Figure 17: Weighted clause ratio: definition and examples, from Foster and Wigglesworth, 2016

As narration often switches between past and present tense even in L1 speech, a decision was taken not to penalise ‘incorrect’ use of tense in this way unless a shift happened within an AS-unit. Following scoring, the mean accuracy score for each participant for each performance was calculated by dividing total accuracy score by number of clauses produced.

9.4.2 Complexity

Following Norris and Ortega (2009), complexity was measured in three ways. Firstly a ‘global’ measure of complexity was selected as the length of AS-unit. In this study, this measure was calculated by dividing the total number of pruned syllables produced by the number of AS-units in the sample. Secondly, the amount of subordination was assessed by dividing the total number of clauses by the total number of AS-units in the speech sample. Finally, the mean length of clause was calculated by dividing the total number of pruned syllables by the total number of clauses. Due to the narrative nature of the elicited speech, there were a large number of co-ordinated clauses. As noted in Foster and Wigglesworth (2016), it is oftentimes difficult to identify whether these clauses are coordinated and they suggest that prosodic features of the speech be taken into consideration when making decisions about this sort of clause. Due to the very high number of these clauses in this particular data set, a decision was taken to count any clause of this nature as a fresh AS-unit rather than co-ordinated clause. This would allow for continuity within the study.

9.6 Inter-Rater Reliability

Due to the detailed and time-consuming nature of the analysis, it was not possible to obtain an inter-rater coding of pause measurement. Instead, a 10% sample of the data was re-examined for syllable count by a trained researcher. Because syllable counts formed the base for most fluency and complexity measures, it was felt that this was important to obtain a second judgement on. The second rater coded twenty-two samples of speech (roughly 10% of the total data) which meant 396 discrete syllable judgements (a judgement was made on each individual runs of speech using the PRAAT file). Spearman’s correlations revealed an exceptionally high reliability score (.993) between scorers based on 396 syllable count judgements (Table 31).

Table 31: Interrater reliability for syllable count

	Rater 1	**Rater 2**
Rater 1 – Correlation co-efficient	1.0	.993
Sig. 2-tailed	-	.000
N	396	396
Rater 2 – Correlation co-efficient	.993	1.0
Sig, 2-tailed	.000	-
N	396	396

9.7 Chapter summary

In this chapter, I presented details of the speech data coding and analysis. First, I focused on the procedures I used to calculate *fluency* measures, using a computer software called PRAAT and using it *semi-automatically* to provide data which could then be used to calculate the measures that I had previously identified. I also briefly explained the procedure for calculating the accuracy and complexity measures. Finally, I provided the results of the inter-rater reliability exploration that was conducted.

10 – Quantitative Results

10. 1 Introduction

This chapter presents the results of the quantitative analyses which were conducted in the present study. The chapter is divided into two main sections. The first section (10.2) deals with the analyses related to short-term fluency and is therefore concerned only with the performances of the TR and PR groups during the training session. The second section (10.3) deals with the analyses related to fluency transfer and is therefore concerned with the performances of all three groups (TR, PR and Control) on the pre- and post-tests. Each section begins with an explanation of the preliminary analyses which were carried out in order to check for outliers, homogeneity and sphericity of the group data. The results of the research are then presented. This chapter concludes with an overview of the quantitative results of the research.

10.2 Task repetition, procedural repetition and short-term fluency

All analyses relating to short-term fluency, were conducted on the data elicited during the training session. It therefore involves only two groups: the task repetition (TR) group and the procedural repetition (PR) group.

Table 32: Fluency measures used in the study and their abbreviations

Fluency Measure	**Abbreviation**
Speech Rate	SR
Mean Length of Run	MLR
Articulation Rate	AR
Phonation Time Ratio	PTR
Frequency of Mid-Clause Pauses	FreqMid

Frequency of reformulations and self-corrections	FreqRepair
Holistic rater score	Hol

10.2.1 Preliminary screening of training session data

First of all, outliers were assessed by examining studentised residual values. There were 2 outliers for the dependent variable of mean length of run. Participant 47 from the PR group had a studentised residual value of 3.48 for performance 3. Participant 46 also from the PR group had a studentised residual of 3.66 for Time 1 and 3.13 for time 2. The 2-way mixed ANOVA was run with and without the inclusion of the outlier data but as there was no dramatic change in the outcome therefore a decision was taken to include the data in the analysis. There were no other outliers in any cell of the design as assessed by examination of studentised residuals for values greater than ±3.

In general, dependent variables were found to be approximately normally distributed as assessed by inspection of Q-Q plots and Levene's test. Some very slight positive skewedness was observed for frequency of mid-clause pauses. However, given that the skewedness was slight and ANOVA is considered robust to this sort of violation (e.g. Lansing, 2004; Bachman, 2004), a decision was taken not to transform the data.

There was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices for speech rate (p= .374), mean length of run (p= .112), articulation rate (p= .007), phonation time ratio (p= .099) frequency of mid-clause pause (p = .177), frequency of repair (p= .121) and holistic score (p= .84).

Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction for the measures of speech rate, $\chi^2(2) = 2.051$, $p = .359$; mean length of run, $\chi^2(2) = 2.634$, $p = .268$; articulation rate, $\chi^2(2) = 2.736$, $p = .255$; Phonation time ratio $\chi^2(2) = 2.960$, $p = .228$; mean length of pause; frequency of repair and the holistic score. The assumption of sphericity was violated, however, for the measure of frequency of mid-clause pause: $\chi^2(2) = 6.428$, $p = .04$. For that reason, the Greenhouse-Geisser adjusted values are reported for this measure.

To establish the homogeneity of the two groups in terms of their starting levels of fluency on the training session tasks, one-way ANOVAs were carried out on the participants' first performance. The results of the ANOVAs revealed no significant differences across the three groups in either the global fluency measures (SR $F(1, 45) = .933$, $p = .339$; MLR $F(1,45) = .395$, $p = .533$), speed fluency (AR $F(1,45) = .109$, $p = .743$), breakdown fluency (PTR $F(1,45) = 3.708$, $p = .061$; FreqMid $F(1,45) = .000$, $p = .989$) or repair fluency measure (freqRepair $F(1,45) = 3.180$, $p = .081$). However, the one-way ANOVA for the holistic fluency measure did reveal a significant difference between the two groups for the first performance of the carousel task ($F(1,45) = 7.902$ $p = .007$) with the TR group out-performing the PR group on the very first iteration. Therefore, any two-way interactions for the effects of type of repetition and time on the holistic measure of fluency must be interpreted with caution. For all other measures, it can be concluded that the two groups were roughly equivalent in their levels of fluency at the outset of the training.

10.2.2 One-way ANOVAs

Hypothesis 1a:

Fluency will increase during both TR and PR carousel task sequences (i.e. over the three iterations of the task) but fluency will increase to a greater extent for the TR group (N. de Jong & Perfetti, 2011).

In order to test Hypothesis 1a, univariate repeated measures (RM) ANOVAs (Analysis of Variance) were first conducted to establish the impact of each type of carousel training independently. This was essential because the aim of this study was to ascertain the usefulness of each form of repetition for language teaching purposes. As there was no 'control' task sequence for the session data, relying solely on two-way interactions could obscure the effects of each form of repetition independently (e.g. if both forms of repetition impacted on fluency but to different degrees, the result may be non-significant even though the impact of one form of repetition might reach significance independently). Effect sizes are interpreted following Cohen (1988) who suggests that partial values of .01, .09, and .25 be considered as small, medium and large effect sizes, respectively, and the alpha level for all statistical tests was set at .05.

One-way RM ANOVAs were conducted for TR and PR separately revealed the following results for each of the seven measures of fluency:

Pruned Speech Rate

There was a statistically significant effect of time on SR for the TR group, $F(2, 46) = 12.516$, $p < .000$, partial $\eta^2 = .352$, a very large effect size. There was no statistically significant effect of time on SR for the PR group, $F(2, 42) = .016$, $p = .9.84$, partial $\eta^2 = .001$.

Mean Length of Run

There was a statistically significant effect of time on PMLR for the TR group, $F(2, 46) = 6.667$, $p = .003$, partial $\eta^2 = .225$, a large effect size. On the other hand, there was no statistically significant effect of time on PMLR for the PR group, $F(2, 42) = .547$, $p = .583$, partial $\eta^2 = .025$.

Articulation Rate

There was a statistically significant effect of time on AR for the TR group, $F(2, 46) = 13.036$, $p < .000$, partial $\eta^2 = .362$, a very large effect size. There was no statistically significant effect of time on AR for the PR group, $F(2, 42) = .229$, $p = .796$, partial $\eta^2 = .011$.

Phonation Time Ratio

There was a statistically significant effect of time for the TR group ($F(2,46) = 3.104$ $p = .05$ partial $\eta^2 = .119$. There was no significant effect of time for the PR group ($F(2, 42) = .321$ $p =.727$ partial $\eta^2 = .015$

Frequency of Mid-clause pauses

There was a borderline statistically significant effect of time on FreqMid for the repetition group, $F(2, 46) = 2.999$, $p = .06$, partial $\eta^2 = .115$. There was no statistically significant impact on performance over time for the PR group $F(2, 42) = 1.080$, $p = .349$, partial $\eta^2 = .049$.

Frequency of Repair

RM ANOVAs revealed that there was no statistically significant effect of time on FreqRepair for the repetition group, $F(2, 46) = .639$, $p = .532$, partial $\eta^2 = .027$. Likewise, there was no statistically significant impact on performance over time for the PR group *F(*2, 42) = 2.212, *p* = .122, partial η^2 = .095.

Holistic judgement

There was a significant effect for time on holistic rater scoring for the TR group (*F* (2,46) = 5.09 *p* = .01 partial η^2 = .181. There was no significant effect of time for the PR group (F (2, 42) = 1.174 *p* =.319 partial η^2 = .053.

10.2.3 Mixed 2-way ANOVAs

The univariate ANOVAs were followed up with 2-way mixed ANOVAs to establish the extent to which *type* of repetition (TR versus PR) affected fluency differently across performances. The 2-way ANOVAs were conducted with 'time' as the within-subjects variable at three levels - one for each performance of the task. The between-subjects factor was 'group', which separated those participants that had repeated the same story (TR) and those that had presented three different stories (PR). Following Kim and Tracy-Ventura (2013) multiple mixed ANOVAs were used in place of MANOVAs because MANOVA does not have as much power as ANOVA especially for small sample sizes (Howell, 2002). Furthermore, MANOVA is not appropriate when there is likely high multicollinearity among dependent variables (as there commonly are in L2 fluency research) (Tabachnick & Fidell, 1996). Individual 2-way ANOVAs were conducted for each of the eight dependent fluency variables. An additional advantage of using individual ANOVAs is that the results are easier to interpret and interactions are more clearly observable. For each dependent variable, first a two-way interaction was sought. Where no two-way interaction was found, the main effect of time is reported.

Follow-up mixed 2-way ANOVAs revealed that there was a significant interaction between time and group for the fluency measures of speech rate (F(2, 88) = 4.702, *p* = .011, partial η^2 = .097), mean length of run (F(2, 88) = 3.593, *p* = .032, partial η^2 = .075), articulation rate (F(2, 88) = 3.997, *p* = .022, partial η^2 = .083) and frequency of mid-length pauses (F(2, 88) = 3.53, *p* = .03, partial η^2 = .074). Frequency of repair and holistic rater scoring (F 2,88 = .800 *p* =.453, partial η^2 = .101) were not significantly affected differently over time depending on type of repetition. This indicates that speech rate, mean length of run, articulation rate and frequency of mid-clause pause were all impacted differently over the course of the intervention depending on group membership. In other words, *type* of repetition was a crucial factor in terms of fluency gains over time. Referring back to the one-way ANOVAs reveals that fluency was *only* found to increase for the TR group with no increase observed for the PR group on any fluency measure.

Table 33: Results of one-way and two-way ANOVAs for the eight fluency measures

		Significant effect for time?	**Significant 2-way interaction?**
Speech rate	*Task repetition*	YES	YES
	Procedural repetition	NO	
Mean length of run	*Task repetition*	YES	YES
	Procedural repetition	NO	
Articulation rate	*Task repetition*	YES	YES
	Procedural repetition	NO	
Phonation Time Ratio	*Task repetition*	YES	NO
	Procedural repetition	NO	
Frequency of mid-clause pause	*Task repetition*	YES	YES
	Procedural repetition	NO	
Mean length of mid-clause pause	*Task repetition*	NO	NO
	Procedural repetition	NO	
Frequency of repair	*Task repetition*	NO	NO
	Procedural repetition	NO	
Holistic judgement	*Task repetition*	YES	NO
	Procedural repetition	NO	

10.2.4 Pairwise comparisons

Hypothesis 1b

Different aspects of fluency will develop at different points in the carousel cycle. Speech rate will increase steadily over the course of the three performances. Subtler changes relating to breakdown fluency (e.g. mid-clause pausing) and repair will take place only on the third performance, if at all (Lambert et al., 2017).

In the following sections, the pairwise comparisons with Fischer's LSD tests (following Lambert et al., 2017) for significant interactions are presented in order to establish the specific nature of the task repetition effect on short-term L2 fluency, thereby testing hypothesis 1b.

Pruned Speech Rate

Table 34: Means and standard deviations for SR

	Performance 1	**Performance 2**	**Performance 3**
TR	129.5 (24.21)	146 (26.41)	147.86 (27.68)
PR	120.49 (38.06)	121.15 (30.41)	121.31 (32.36)

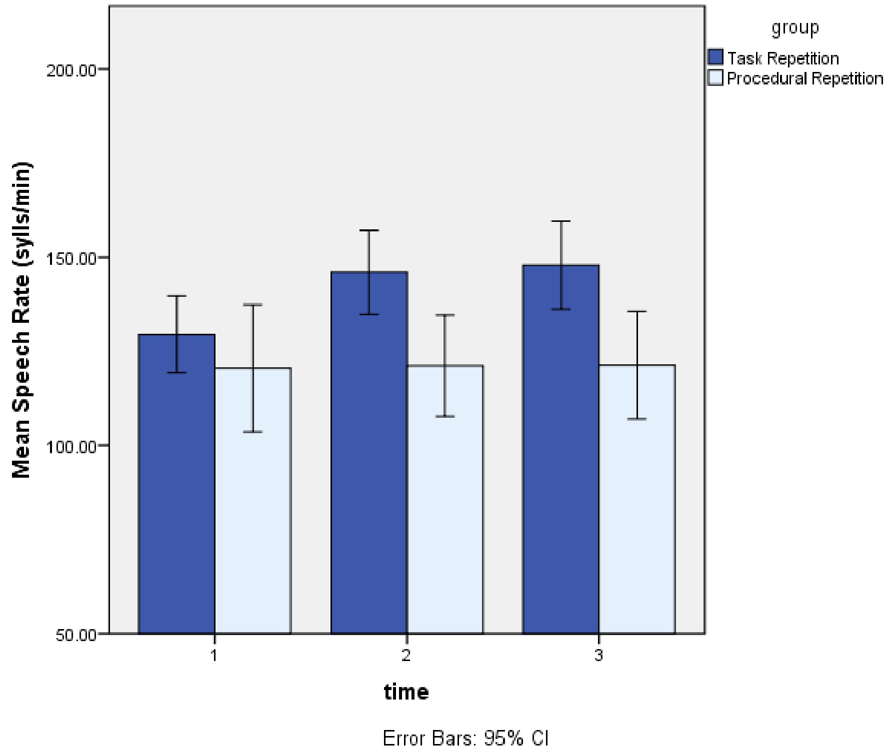


Figure 18: Training session: SR

Pairwise comparisons revealed that SR was statistically significantly higher at Time 2 when compared to Time 1 (M = 16.499, SE = 3.322 syll/sec, $p < .000$). There was also a significant increase between Time 1 and Time 3 (M = 18.365, SE = 4.479 syll/sec, $p = .001$), but not between Time 2 and Time 3 (M = 1.866, SE = 4.229 syll/sec, $p = 1.0$).

These results show that SR increases considerably and significantly on the first repetition for the TR group and is maintained for the second repetition.

Mean Length of Run

Table 35: Means and standard deviations for MLR

	Performance 1	**Performance 2**	**Performance 3**
TR	5.03 (1.34)	5.99 (1.73)	6.04 (1.57)
PR	4.68 (2.37)	4.84 (2.13)	4.50 (2.08)

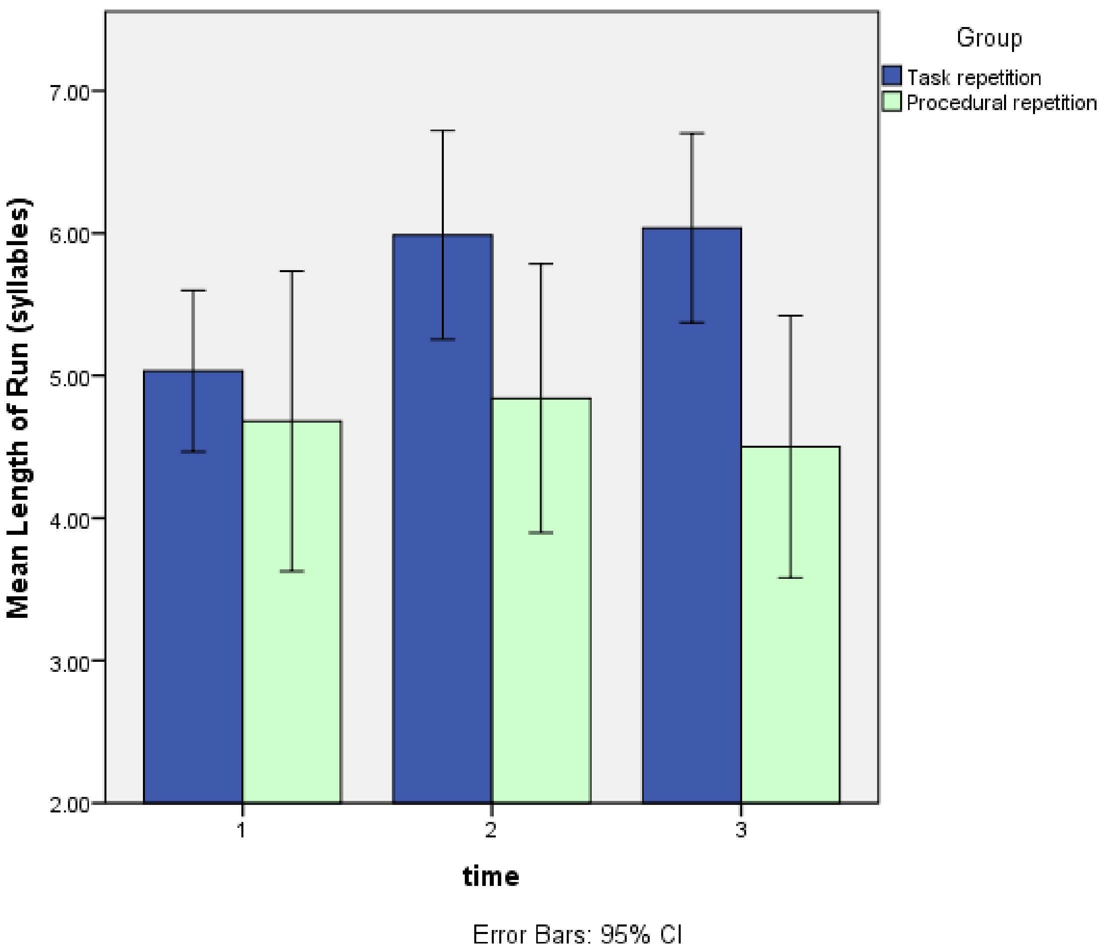


Figure 19: Training session: MLR

Pairwise comparisons revealed that PMLR was statistically significantly higher at Time 2 when compared to Time 1 (M = 0.954, SE = 0.295 sylls, *p* = .011). There was also a significant increase between Time 1 and Time 3 (M = 1.002, SE = 0.358 sylls, *p* = .03), but not between Time 2 and Time 3 (M = 0.048, SE = 0.270 sylls, *p* = 1.0).

Similarly to SR, the impact of immediate TR on MLR is considerably and significantly different. MLR increases considerably and significantly on the first repetition for the TR group and is maintained for the second repetition.

Articulation Rate

Table 36: Means and standard deviations for AR

	Performance 1	Performance 2	Performance 3
TR	214.43 (6.7)	228.78 (27.65)	227.65 (28.29)
PR	211.63 (34.52)	211.95 (25.80)	208.54 (26.40)

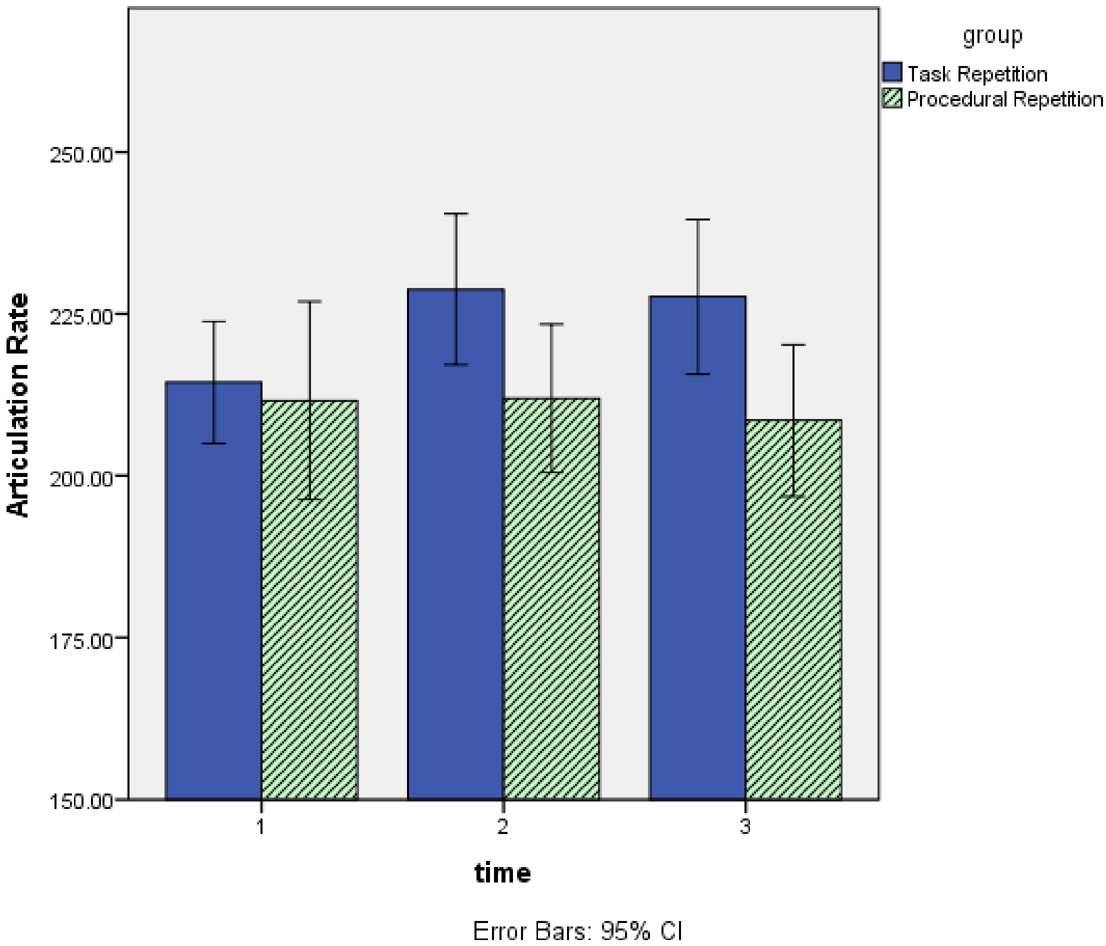


Figure 20: Training session: AR

Pairwise comparisons revealed that AR was statistically significantly higher at Time 2 when compared to Time 1 (M = 14.349, SE = 2.986 syll/sec, $p < .000$). There was also a significant increase between Time 1 and Time 3 (M = 13.223, SE = 4.479 syll/sec, $p = .001$), but not between Time 2 and Time 3 (M = 1.127, SE = 3.211 syll/sec, $p = 1.0$).

Phonation Time Ratio

Table 37: Means and standard deviations for PTR

	Performance 1	Performance 2	Performance 3
TR	73.98 (8.28)	76.33 (7.87)	76.86 (8.42)
PR	68.72 (10.20)	69.62 (11.6)	68.20 (9.64)

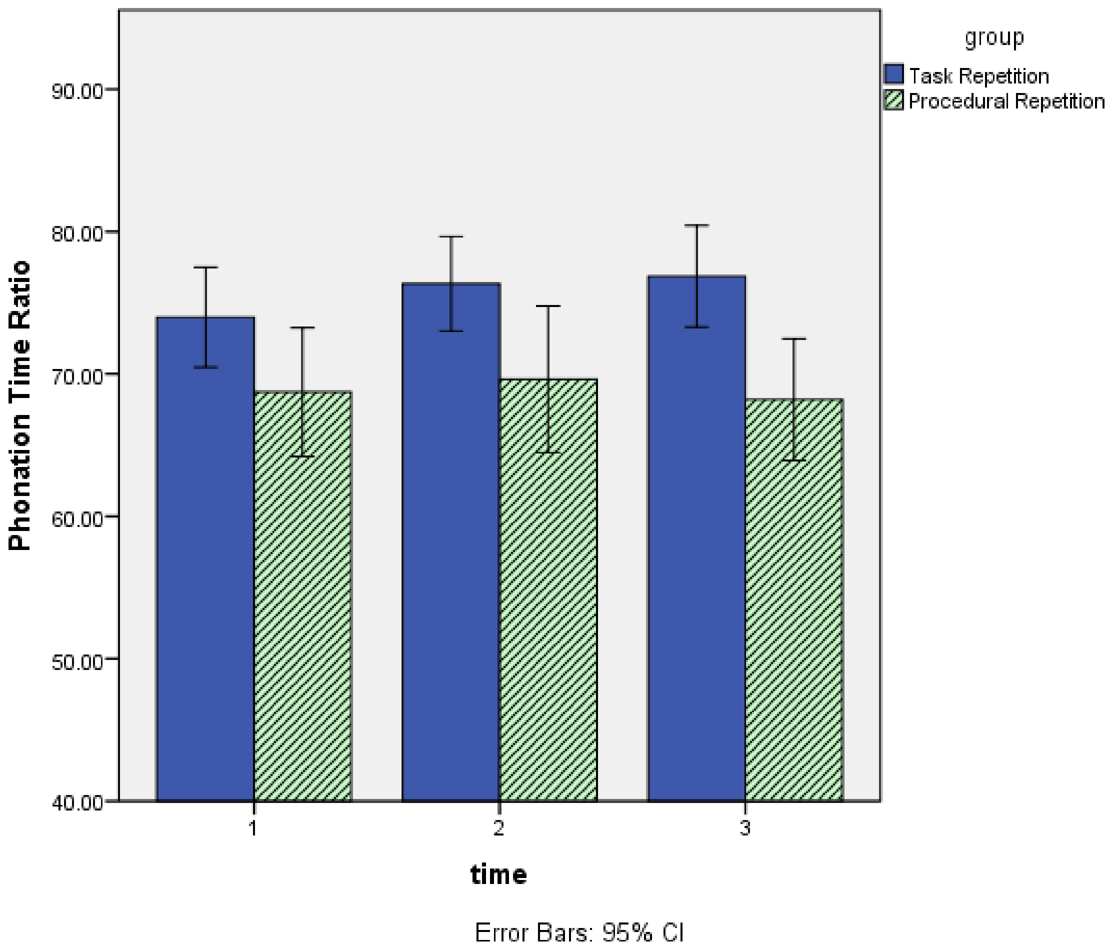


Figure 21: Training session: PTR

Pairwise comparisons revealed that PTR was statistically significantly higher at Time 3 when compared to Time 1 (M = 2.878, SE = 1.396%, *p* = .05). There was no significant increase between Time 1 and Time 2 (M = 2.353, SE = 1.206%, *p* = .063), or between Time 2 and Time 3 (M = .525, SE = 1.067%, *p* = .627).

Frequency of mid-clause pause (FreqMid)

Table 38: Means and standard deviations for FreqMid

	Performance 1	Performance 2	Performance 3
TR	7.67 (3.2)	6.75 (2.67)	6.04 (2.54)
PR	7.68 (3.88)	8.41 (4.27)	8.73 (4.32)

Pairwise comparisons revealed a significant decrease between first and third performances (M = 1.625, SE = .761, p = .044). There was no significant difference between Time 1 and Time 2 (M = .917, SE = .656, p = .176), or between Time 2 and Time 3 (M = .708, SE = .480, p = .153).

Holistic judgement

Hypotheses 1c and 1d

1c: An increase in holistic rater scoring will be observed during training for TR and PR groups (Gass et al., 1999). Gains will be greater for the TR group.

1d: Holistic judgements of fluency will only be partially explained by temporal fluency measures but speech rate and mean length of run will correlate most strongly (Kormos & Denes, 2004; Derwing et al., 2004).

In order to test Hypothesis 1c, One-way and two-way ANOVAs were carried out on the session data with the combined rater scores for each performance as the dependent, within-subjects variable and group as the between-subjects variable (for the two-way ANOVA).

Table 39: Means and standard deviations for holistic rater scores

	Performance 1	Performance 2	Performance 3
TR	5.46 (.97)	5.94 (1.07)	5.9 (1.29)
PR	4.66 (1.12)	4.82 (1.2)	4.95 (1.22)

Pairwise comparisons revealed that holistic rater scoring was statistically significantly higher at Time 2 when compared to Time 1 (M = .479, SE = .133, *p* = .001). The difference in scores between Time 1 and Time 3 was lower than between Time 1 and Time 2 but was still significantly higher (M = .438, SE = .174, *p* = .019), there was no significant difference for the reduction in score between Time 2 and Time 3 (M = .042, SE = .188, *p* = .826). Mean scores also increased steadily for the PR group but this did not reach significance.

In order to test Hypothesis 1d, Spearman's rank order correlations were used to compare holistic rating scores for students' performances and the temporal fluency measures as well as correlations between variables. Table 40 shows the results of this analysis. Hypothesis 1d was supported in that speech rate emerged as being most strongly correlated with rater scores, both individually (.749; .820) and as a combined score (.830) Mean length of run was also highly correlated with rater scores. All correlations are highly significant ($p < .000$) and based on 138 observations.

Table 40: Spearman correlations for individual and mean rater scores and fluency measures

	SR	**MLR**	**AR**	**PTR**	**FreqMid**	**FreqRepair**
Mean Rater	.830	.765	.571	.497	.457	.333
Rater 1	.749	.692	.513	.445	.442	.281
Rater 2	.820	.759	.571	.497	.432	.355

Summary of short-term fluency results

The results of the analyses carried out on the fluency measures for the training session data (i.e. short-term) revealed an effect for task repetition (TR group) on speech rate, mean length of run, articulation rate, frequency of mid-clause pauses and holistic scores. Effect sizes were considered large (Cohen, 1988) for speech rate and articulation rate. Procedural repetition was not found to impact on fluency in the short-term as measured by any of the DVs of fluency used in this study.

10.2.5 Task repetition, procedural repetition and short-term complexity and accuracy

Hypotheses 1e and 1f

Complexity (Bygate, 2001) and accuracy (Wang, 2014; Lynch & Maclean, 2000; 2001) will increase for both the TR group and PR group (Sample & Michel, 2014) during the training session. Gains in both areas will be larger for the TR group.

For the TR group all three aspects will increase on the first repetition (Wang, 2014). For the PR group, fluency will increase immediately (2nd performance) while complexity and accuracy will not increase until the third performance (Sample & Michel, 2014).

Table 41: Complexity and accuracy measures used in the study

	Operationally defined as	**How calculated?**
Complexity	Clauses per AS-unit (amount of subordination)	Total number of clauses produced divided by total number of AS-units
	Mean length of AS-unit	Total number of syllables produced divided by total number of AS-units
	Mean length of clause	Total number of syllables divided by total number of clauses
Accuracy	Weighted clause ratio	Total accuracy score divided by total number of clauses

In order to test Hypotheses 1e, one-way RM ANOVAs and two-way mixed ANOVAs were carried out to establish the impact of both training procedures on complexity and accuracy individually and in contrast with each other. In order to test hypothesis 1f, pairwise comparisons with Fischer's LSD tests (Lambert et al., 2017) are conducted wherever a simple main effect for time is found. The results are set out in the following sections

Complexity

Clauses per AS-unit

Table 42: Means and standard deviations for clauses per AS-unit

	Performance 1	**Performance 2**	**Performance 3**
TR	1.59 (.36)	1.75 (.4)	1.77 (.46)
PR	1.46 (.31)	1.55 (.26)	1.54 (.28)

There were no outliers, as assessed by boxplot. The data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p < .05$). There was homogeneity or variances ($p > .05$) and covariances ($p < .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 3.128$, $p = .209$.

One-way repeated measures ANOVAs were performed for the TR and PR groups separately. For the TR group there was a significant within-subject effect for time $F(2,46) = 3.047$ $p = .05$, with a large effect size partial $\eta^2 = .117$ although pairwise comparisons did not reach significance. There was no significant effect for time on the PR group (F(2, 42) = .914 $p =.409$ $\eta^2 = .042$. In terms of a two-way interaction, there was no statistically significant interaction between type of repetition and time on clauses per AS-unit, $F(2, 88) = .471$, $p = .626$, partial $\eta^2 = .011$.

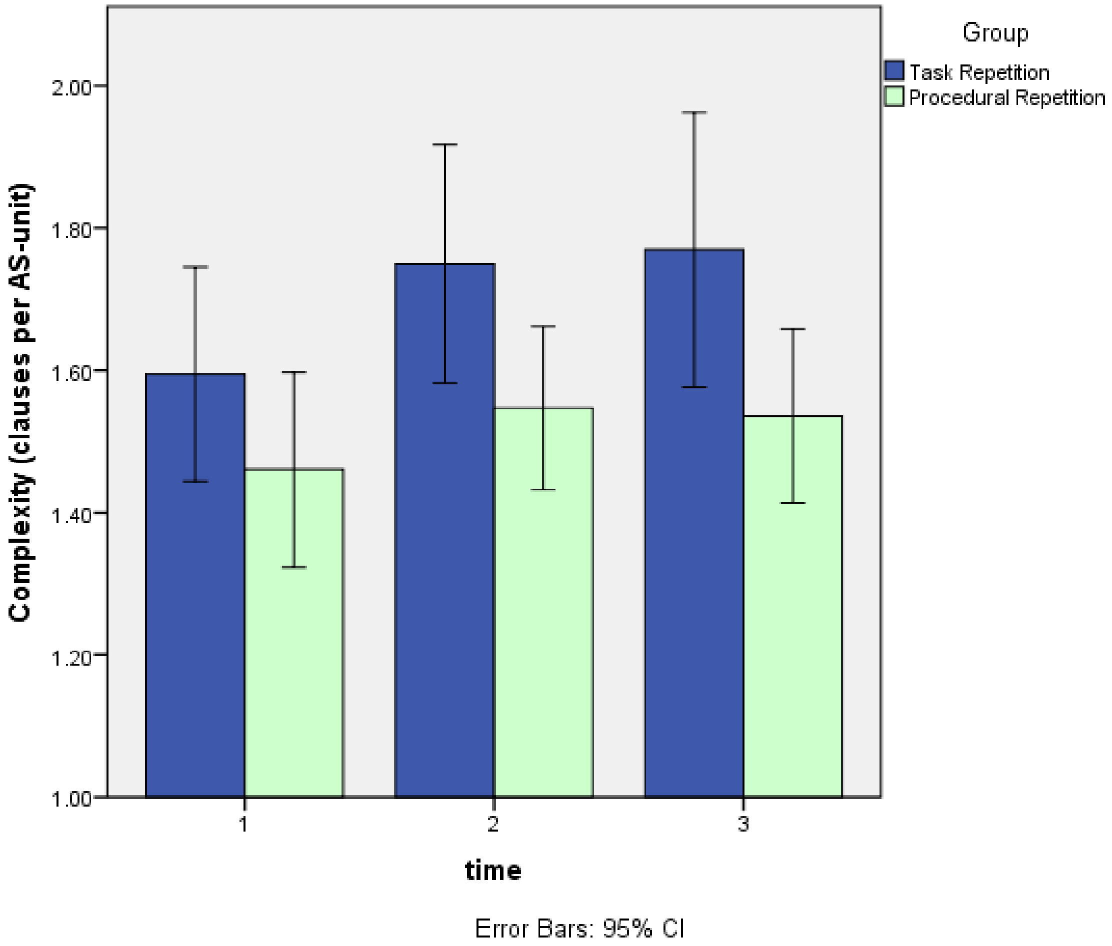


Figure 22: Training session: Clauses per AS-unit

Mean length of AS-unit

Table 43: Means and standard deviations for Mean length of AS-unit

	Performance 1	Performance 2	Performance 3
TR	11.90 (2.53)	13.25 (2.76)	12.96 (3.2)
PR	11.14 (3.77)	11.76 (4.04)	10.55 (3.03)

There were no outliers, as assessed by boxplot. The data was normally distributed, as assessed by Shapiro-Wilk's test of normality (p > .05). There was homogeneity or variances (p > .05) and covariances (p > .05), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 1.283$, p = .526. ANOVA was performed for the

TR and PR data. There was a trend observed for time for the TR group $F(2,46) = 2.734$ $p = .075$, partial $\eta^2 = .106$ and pairwise comparisons showed a significant increase between Time 1 and Time 2 M= 1.353 SE = .577 $p = .028$. Differences between Time 1 and Time 2 or Time 2 and Time 3 were not significant. There was no significant effect for time for the PR group (F (2, 42) = 1.113 $p = .338$ partial $\eta^2 = .050$. There was no statistically significant interaction between type of repetition and time on mean length of AS-unit, $F(2, 88) = 1.357$, $p = .263$, partial $\eta^2 = .030$.

Mean length of clause

Table 44: Means and standard deviations for mean length of clause

	Performance 1	**Performance 2**	**Performance 3**
TR	7.55 (1.12)	7.7 (1.34)	7.42 (.93)
PR	7.79 (2.6)	7.64 (2.34)	7 (1.65)

There were no outliers, as assessed by boxplot. The data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity or variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 1.283$, $p = .526$. ANOVA was performed for the TR and PR data separately using the 'split group' function. There was no significant effect for the TR group $F(2,46) = .666$ $p = .519$, partial $\eta^2 = .028$. Neither was there a significant effect for time for the PR group (F (2, 42) = 1.955 $p = .154$ partial $\eta^2 = .085$. There was no statistically significant interaction between type of repetition and time on mean length of AS-unit, $F(2, 88) = .957$, $p = .388$, partial $\eta^2 = .021$.

Accuracy

Table 45: Means and standard deviations for Accuracy

	Time 1	**Time 2**	**Time 3**
TR	.81 (.1)	.82 (.1)	.85 (.09)
PR	.71 (.17)	.72 (.2)	.73 (.15)

There were no outliers, as assessed by boxplot. The data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was homogeneity or variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi^2(2) = 1.731$, $p = .421$. ANOVA was performed for the TR and PR data separately using the 'split group' function. There was a significant within-subject effect for time for the TR group $F(2,46) = 3.461$ $p = .04$, partial $\eta^2 = .131$ and pairwise comparisons showed a significant increase between Time 1 and Time 3 M= .041 SE = .016 $p = .019$. Differences between Time 1 and Time 2 or Time 2 and Time 3 did not reach significance. There was no significant effect for time for the PR group $F(2.42)=.143$ $p =.867$ $\eta^2 = .007$. There was no statistically significant interaction between type of repetition and time on accuracy, $F(2, 88) = .253$, $p = .777$, partial $\eta^2 = .006$.

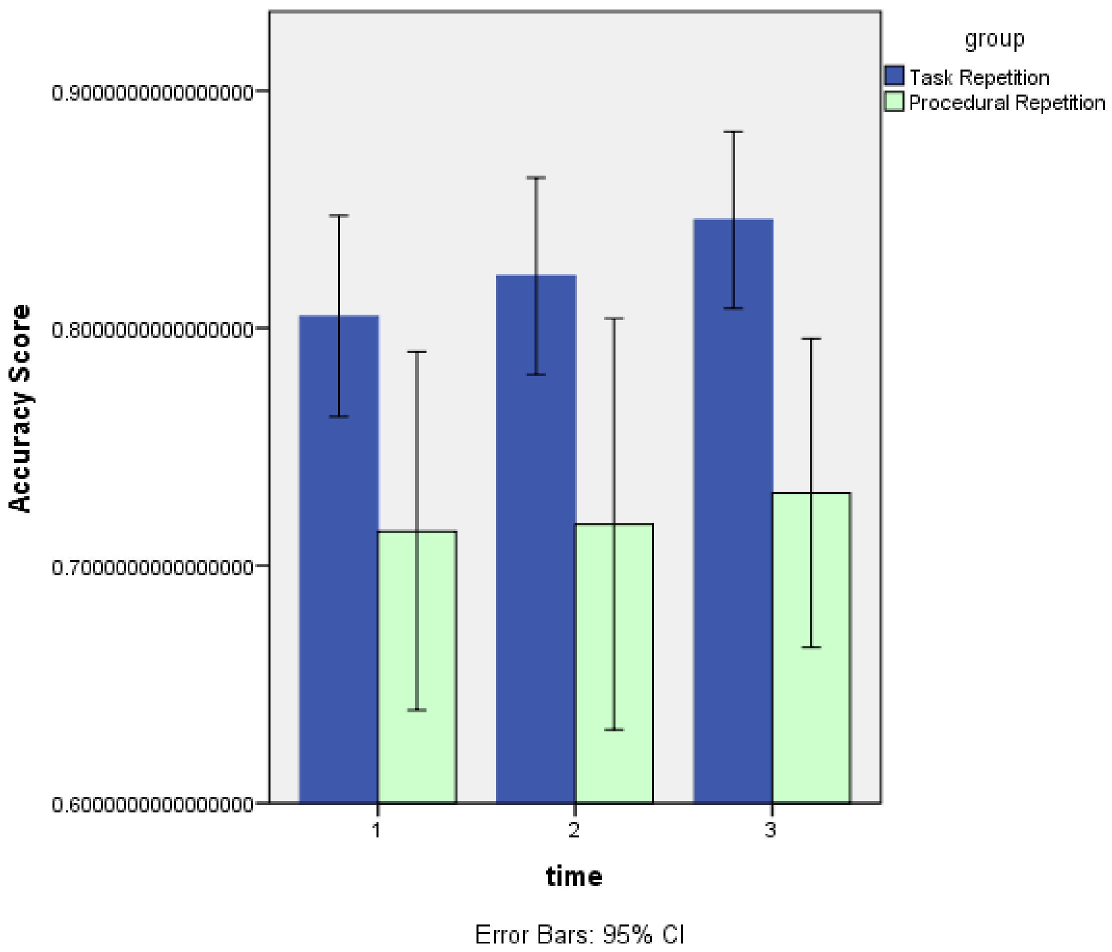


Figure 23: Training session: Accuracy

Table 46: Results of one-way and two-way ANOVAs for the three complexity measures and one accuracy measure

		Significant effect for time?	Significant 2-way interaction?
Mean length of AS-unit	*Task repetition*	NO	NO
	Procedural repetition	NO	
Amount of subordination	*Task repetition*	YES	NO
	Procedural repetition	NO	
Mean length of clause	*Task repetition*	NO	NO

	Procedural repetition	NO	
Weighted clause ratio	*Task repetition*	YES	NO
	Procedural repetition	NO	

Table 47: Impact of TR and PR carousel on global performance measures of CAF

	Performance 1	Performance 2	Performance 3
Complexity			
Amount of subordination			
Task repetition	1.59 (.36)	1.75 (.4)	1.77 (.46)
Procedural repetition	1.46 (.31)	1.55 (.26)	1.54 (.28)
Accuracy			
Weighted clause ratio			
Task repetition	.81 (.1)	.82 (.1)	.85 (.09)
Procedural repetition	.71 (.17)	.72 (.2)	.73 (.15)
Fluency			
Speech Rate			
Task repetition	129.5 (24.21)	146 (26.41)	147.86 (27.68)
Procedural repetition	120.49 (38.06)	121.15 (30.41)	121.31 (32.36)
Mean Length of Run			
Task repetition	5.03 (1.34)	5.99 (1.73)	6.04 (1.57)
Procedural repetition	4.68 (2.37)	4.84 (2.13)	4.50 (2.08)

Table 48: Differences for CAF measures between Time 1 and time 2, Time 2 and time 3, and Time 1 and time 3

	Time 2 – Time 1	**Time 3 – Time 2**	**Time 3 – Time 1**
Complexity			
Amount of subordination			
Task repetition	0.16	0.02	0.18*
Procedural repetition	0.09	-0.01	0.08
Accuracy			
Weighted clause ratio			
Task repetition	0.01	0.03	0.04*
Procedural repetition	0.01	0.01	0.02
Fluency			
Speech Rate			
Task repetition	16.5**	1.86	18.36**
Procedural repetition	0.66	0.16	0.82
Mean Length of Run			
Task repetition	0.96**	0.05	1.01**
Procedural repetition	0.16	-0.34	-0.18

1(denotes significance at .05 level, ** denotes significance at .005 level)*

10.2.6 Summary of short-term impact of carousels on performance

Global fluency is affected significantly on the first repetition but only for the TR group (Table 48). In terms of specific aspects of fluency, speed increases significantly between Performance 1 and 2, breakdown decreases significantly between Performance 1 and 3 and repair is not significantly affected. Holistic judgement follows the same pattern as global and speed fluency with significant increases between Performance 1 and Performance 2. Accuracy and the complexity measure of amount of subordination increase steadily across three performances giving rise to significant increase only between the first and third performances and only for the TR group.

10.3 Task Repetition, Procedural Repetition and fluency transfer

Hypothesis 2

There will be an increase in fluency between pre- and post-test only for those students that took part in TR carousel training (N. de Jong & Perfetti, 2011)

All analyses relating to fluency transfer, were conducted on the data elicited during the pre- and post-test data. It therefore involves all three groups: the task repetition (TR) group, the procedural repetition (PR) group and the control group (C).

10.3.1 Preliminary screening of pre- and post-test data (for fluency transfer analysis)

There were two outliers in the pre-test data for speech rate as assessed by examination of studentised residuals for values greater than ±3. These two participants from the PR group had unusually low speaking rates at time 1. The recordings were listened to and it was established that they had both abandoned the task early (they were actually partnered with each other). For this reason, their data was excluded from analysis for all measurements as it was likely that the inclusion of this data would give a disproportionally low fluency reading for this group for the pre-test.

All measures were approximately normally distributed as assessed by inspection of Q-Q plots and Levene's test and there was homogeneity of covariances, as assessed by Box's test of equality of covariance matrices on all dependent variables.

To establish the homogeneity of the three groups in terms of their starting levels of fluency, one-way ANOVAs were carried out on the pre-test performances. The results of the ANOVAs revealed no significant differences across the three groups in either the overall fluency measures (SR $F(2, 61) = 2.586$, $p = .084$; PMLR $F(2, 61) = .849$, $p = .433$), speed fluency (AR $F(2,$

61) = 2.404, p = .099), breakdown fluency (PTR $F(2, 61)$ = .676, p = .513; FreqMid $F(2, 61)$ = 2.063, p = .136, or repair fluency measure (freqRepair $F(2,61)$ = 1.790, p = .176);). However, there was a significant difference between groups on the holistic rater scores (F(2,61)= 6.809, p =.002) Post-hoc Bonferroni tests found that the TR group had a significantly higher score (M=1.08 SE.3 p =.002) than the PR group. For this reason, two-way interactions for pre- and post-test data in respect of the holistic scores must be interpreted with extreme caution.

10.3.2 Mixed 2-way ANOVAs

To determine if there was an effect for the intervention on the fluency of students over time, mixed two-way ANOVAs were carried out using the pre- and post-test data for each of the dependent fluency variables. 'Time' at two levels was the within-subjects factor which distinguished between pre- and post-test data and the between-subjects factor was 'group', which separated those participants that had undergone training and repeated the same story (TR), those that had presented three different stories (PR) and those that had no special training at all (Control). For each dependent variable, first a two-way interaction was sought, followed by simple main effects. Where no two-way interaction was found, the main effects of time are reported. In contrast to the analysis of the session data, here simple main effects (i.e. univariate RM ANOVAs) were only investigated if a significant two-way interaction was found. This is due to the presence of a true control group.

Table 49: Results of two-way ANOVAs and follow-up one-way ANOVAs (where appropriate)

		Significant 2-way interaction?	Significant effect for time?
Speech rate	*Task repetition*	TREND	NO
	Procedural repetition		YES
Mean length of run	*Task repetition*	NO	N/A
	Procedural repetition		
Articulation rate	*Task repetition*	YES	NO
	Procedural repetition		YES
Phonation Time Ratio	*Task repetition*	NO	N/A
	Procedural repetition		
Frequency of mid-clause pause	*Task repetition*	NO	N/A
	Procedural repetition		
Mean length of mid-clause pause	*Task repetition*	NO	N/A
	Procedural repetition		
Frequency of repair	*Task repetition*	NO	N/A
	Procedural repetition		
Holistic	*Task repetition*	NO	N/A
	Procedural repetition		

Pruned Speech Rate (SR)

Table 50: Means and standard deviations for SR (pre- post-tests)

	Pre-Test	Post-Test
TR	145.43 (28.85)	149.72 (28)
PR	123.6 (29.26)	130.84 (32.91)
Control	145.29 (29.25)	138.17 (28.82)

There was no statistically significant interaction between group and time on SR, $F(2, 58) = 2.946$, $p = .06$, partial $\eta^2 = .092$. However, as the result came very close to significance, follow-up simple main effects were investigated. Univariate RM ANOVAs revealed a significant difference between pre- and post-test for the PR group with a large effect size $F(1, 17) = 4.265$, $p = .05$, partial $\eta^2 = .201$ No significant change was found for either the TR group ($F(1, 23) = .731$, $p = .401$, partial $\eta^2 = .031$) or the control group ($F(1, 18) = 2.411$, $p = .138$, partial $\eta^2 = .118$).

Mean length of run

There was no statistically significant interaction between group and time on SR, $F(2, 59) = .397$, $p = .397$, partial $\eta^2 = .031$. There was no main effect for time $F(1, 59) = .128$, $p = .722$, partial $\eta^2 = .002$ or group , $F(2, 59) = 1.753$, $p = .182$, partial $\eta^2 = .056$.

Articulation Rate

Table 51: Pre- and post-test - Articulation rate

	Pre-Test	**Post-Test**
TR	223.96 (30.39)	224.68 (24.38)
PR	204.81 (23.46)	221.89 (26.46)
Control	225.39 (26.77)	219.91 (17.74)

There was a statistically significant interaction between group and time on AR, $F(2, 59) = 5.953$, $p = .004$, partial $\eta^2 = .168$. Simple main effects were therefore investigated. Univariate RM ANOVAs revealed a significant difference between pre- and post-test for the PR group with a very large effect size $F(1, 19) = 9.087$, $p = .007$, partial $\eta^2 = .324$ No significant change was found for either the TR group ($F(1, 23) = .031$, $p = .862$, partial $\eta^2 = .001$) or the control group ($F(1, 17) = 1.975$, $p = .178$, partial $\eta^2 = .104$).

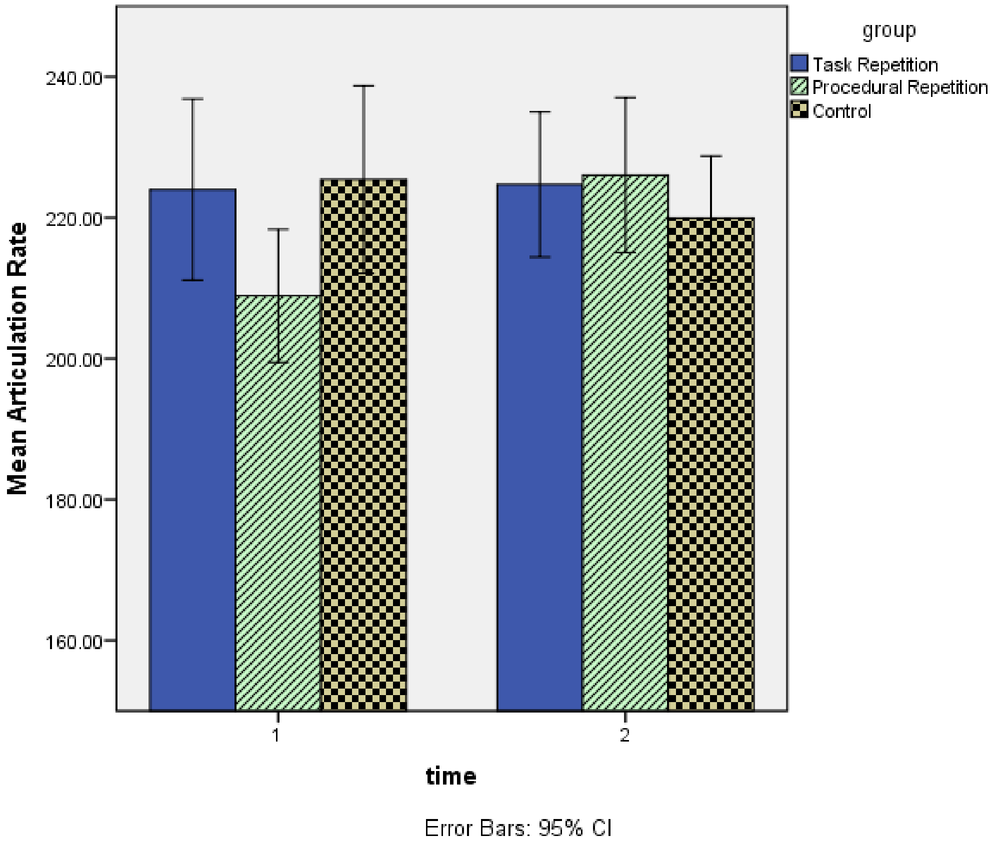


Figure 24: Pre- post-test: AR

Phonation time ratio

The data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p < .05$). There was no statistically significant interaction between the intervention and time on phonation time ratio, $F(2, 59) = .609$ $p = .547$, partial $\eta^2 = .020$. There was no main effect of time $F(1, 59) = .004$, p .947 , partial $\eta^2 = .000$. The main effect of group showed that there was no statistically significant difference in phonation time ratio between intervention groups $F(2, 59) = .910$, $p = .408$, partial $\eta^2 = .030$.

Frequency mid-clause pauses

The data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was no statistically significant interaction between the intervention and time on frequency of repetition, $F(2, 59) = 1.262$, $p = .291$, partial $\eta^2 = .041$. There was no main effect for time $F(1, 59) = .144$, p .705 , partial $\eta^2 = .002$ or group $F(2, 59) = 2.177$, $p = .122$, partial $\eta^2 = .069$.

Mean length mid-clause pauses

The data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was no statistically significant interaction between the intervention and time on mean length mid-clause pauses, $F(2, 59) = .139$, $p = .870$, partial $\eta^2 = .005$. There was no main effect for time $F(1, 59) = .133$, p .717 , partial $\eta^2 = .002$ or group $F(2, 59) = 1.402$, $p = .254$, partial $\eta^2 = .045$.

Frequency repair

There was no statistically significant interaction between the intervention and time on frequency of repetition, $F(2, 59) = .966$, $p = .387$, partial $\eta^2 = .032$. There was no main effect for time $F(1, 59) = 1.454$, p .233 , partial $\eta^2 = .024$ but there was a significant main effect for group $F(2, 59) = 8.242$, $p = .001$, partial $\eta^2 = .218$. The TR group was associated with a mean frequency of repair score 1.151 (95% CI, .192 to 2.110) points lower than the PR group, a statistically significant difference, $p = .013$. The TR group was also associated with a mean frequency of repair score which was 1.524 (95% CI, .536 to 2.511) lower than the control group (p= .001)

Holistic Judgement

The data was normally distributed, as assessed by Shapiro-Wilk's test of normality ($p > .05$). There was no statistically significant interaction between the intervention and time on holistic rater scoring, $F(2, 59) = .691$, $p = .505$, partial $\eta^2 = .023$. There was no main effect for time $F(1, 59) = .052$, p .820 , partial $\eta^2 = .001$

10.3.3 Summary of carousel and fluency transfer

The results show an effect for the intervention on the global fluency measure of speech rate for the procedural repetition group, and although this fell just short of 2-way significance (i.e. compared with control) the simple main effect of time was significant for the PR group. There was a significant increase in articulation rate for the PR group. No other change in fluency measures was observed for any group.

10.4 Chapter summary

In this chapter, I have presented the quantitative results from the study. The first section of this chapter dealt with the results of the study in relation to short-term fluency (i.e. fluency changes during the training session). I presented the preliminary analyses, one-way and two-way ANOVA results and followed these up with the relevant pairwise comparisons. Each set of results was presented in a section which began with the corresponding hypotheses. In the second section, I presented the results of the study in relation to fluency transfer. Once again, I presented the preliminary analyses, 2-way and follow-up one-way ANOVAs. In the next chapter, I explore these results in greater detail by looking at two case studies.

11 – Two case studies

11.1 Introduction

In this chapter, I explore possible explanations for the main findings through the analysis of two case studies. The chapter begins with the presentation of the case study of Kimmy[18], who was in the TR group. In particular, I examine her performances during the TR carousel in an attempt to shed some light on the cognitive processes which may be responsible for her increased fluency. Next, I present the case study of Eva[19], a member of the PR group. Here, I focus on finding potential explanations for the fact that there was a transfer effect for fluency observed. I conclude with a general discussion of these findings and how they fit in with the quantitative results.

These two case studies were chosen on the basis of 'typical sampling' (Dörnyei, 2007) drawing on 'nested samples' (Riazi & Candlin, 2014). Specifically, they were chosen because they appeared to be fairly typical profiles because their temporal fluency scores closely reflected the group scores (i.e. increased fluency during session for TR and increased fluency between pre- and post-test for PR). It was felt that they would be interesting cases to look at in order to elucidate the repetition process and its impact on fluency. In particular, the data is explored for potential explanations for a) short-term fluency increase for the TR group and b) fluency transfer for the PR group.

11.2 Case 1 – Kimmy - TR group

11.2.1 Quantitative results

In terms of the quantitative analysis of Kimmy's performances, the results were fairly typical of the group means (Table 52). Her speech rate increased considerably between Performance

[18] Pseudonym
[19] Pseudonym

one and Performance two and then more modestly between Performance two and Performance three. Frequency of mid-clause pause decreased considerably between Performance one and Performance two, and holistic score followed the same pattern as speech rate, with the final performance being judged as the most fluent (and very high).

There is a small increase in fluency for Kimmy between pre- and post-test and it is clear that her performance on the pre-task was substantially more fluent that her first performance on the session task. This is not a surprising finding as others have found an effect for task type on fluency (e.g. Bygate, 2001) and it makes sense intuitively that talking about personal information will elicit more fluent speech than an unfamiliar narrative. However, what is very interesting is that, by the third performance, Kimmy's speech rate, articulation rate and holistic score reach similar levels to her pre-test. In a sense, then, the effects of TR in the case of utterance and holistic variables of fluency appear to override the effects of task type.

Table 52: Results of quantitative analysis for Kimmy

	Pre-test (Day 1)	**Session P1 (Day 2)**	**Session P2**	**Session P3**	**Post-test (Day 5)**
Speech Rate	175.39	126.93	162.48	175.91	208.86
Mean length of Run	11.69	5.08	9.06	8.42	11.75
Articulation rate	233.31	215.66	228.11	235.66	271.79
Phonation time ratio	77.99	79.80	87.69	89.37	82.08
Frequency of mid-clause pause	.94	10	4	3	7.11
Mean length of mid-clause pause	.71	.79	1.11	.58	.42
Repair	1.88	2	1	1.1	3.56

Holistic judgement score	7.5	5.50	6.00	7.50	8
Accuracy	N/A	.91	.81	.91	N/A
Length of AS	N/A	18.14	14.82	13.33	N/A
Subordination	N/A	2.29	1.91	1.67	N/A
Length of clause	N/A	7.94	7.76	8.00	N/A

11.2.2 Session data transcriptions

Performance 1

(0.38ev) the story was about (0.48mv) a woman :: who worked for (0.31mu) charity (**1.28eu**) | and (0.35ev) yeah she was pregnant | and after work she went (1.01ev) she wanted :: to go with her friend (0.71mc) home (**1.98ec**)| but they were in the countryside :: so there were no other cars (0.68eu)| and (0.74ev) when they started (0.46eu) :: to go :: her friend (0.39mv) who was driving the car :: it started :: to rain very (0.38mu) heavily | and it was (0.4eu) the weather was so bad :: that her friend (0.28mv) couldn't see the road anymore (0.86eu) :: and couldn't control the steering (**1.29mu**) wheel | yeah (2.32ec) so they (**1.85mc**) went (**1.87mc**) away from the road :: or the because she couldn't see the road (0.43ev) :: and dropped {10} (0.27mu) in a ravine (*pron*)(**1.27eu**)| and (0.69ev) over (0.96ec) when they were (**1.77mc**) in the ravine trapped a stick of wood (0.31mu) was in her chest and in the yeah

Performance 2

the story was about a woman :: who worked for a charity | and (0.6ev) yeah (0.26eu) she was pregnant (0.97eu) | and (**1.07ev**) she was with her friend in the countryside | and they wanted :: to go back to their home (**1.27eu**) | and (0.55ev) her friend :: who was driving the car (**2.2ec**) :: had problems :: to see the road :: because there weather was very bad | so it started :: to rain very heavily | and (0.51ev) it wasn't (0.35mu) yeah possible :: to see the road | so they (**3.3mc**) get away from the road :: and dropped (0.26mu) in the ravine (*pron*)| and (0.62ev) yeah has big stick of wood (0.52mu) was in the chest of the pregnant woman (0.94eu)| and she was afraid :: that the baby has any injuries :: because she was in the eighth month :: and (0.55ev) yeah everything was with blood and then (**1.49eu**)| and (0.67ev) they had to wait for the help because (0.41eu)

Performance 3

So the story was about a woman (0.88ec) | she was pregnant :: and worked for a charity (**1.73mc**) on the countryside (0.8eu)| and (0.3ev) when she wanted to go back home with her friend (0.3ev) :: they started to drive :: but then the weather was very bad | it started :: to rain very heavily | so (0.38eu) her friend (0.33mv) couldn't see the road any more :: and (0.77eu) drove in a fence | and (**1.17ec**) yeah a big stick of wood was in the the pregnant (0.65mu) woman's chest (0.88eu)| and she was very afraid :: that (0.51ev) her child has any injuries :: or (0.75ev) if it's still alive | (0.47eu) they waited for help (0.68ev) | and then the helicopter brought her to the hospital (**1.14eu**)| and after some surgeries :: she (0.76mv) gave birth| and everything was ok with the child | so (0.57eu)

11.2.3 Analysis

A deductive approach was taken to the case study analysis in the sense that the transcriptions and recordings were investigated in order to shine light on the observed changes in utterance fluency at both the group and individual. Following a careful analysis of the three performances a number of interesting observations emerged:

Firstly, and in line with N. de Jong and Perfetti (2011) and Thai and Boers (2016), it was noted that certain phrases were repeated verbatim. Secondly, it was noted that these phrases tended to be delivered with a lesser degree of pausing and sound lengthening. Finally, and of particular interest, it was observed that Kimmy seemed to move away from verbatim repetition in the third performance and made changes to her speech that may have further enhanced her fluency scores. Each of these observations is dealt with in turn in the section below.

Verbatim repetition

In the discussion of his study into the effects of the 4-3-2 task sequence on learner performance, Boers (2014: 228) explains that "(t)he sheer amount of verbatim repetition across the [repeated] deliveries is striking. In some of the third deliveries, exact duplicates of word strings from previous deliveries spanned stretches of over 30 words (in one case even 50 words)."

Table 53: Showing examples of verbatim repetition across the three performances (Kimmy)

	Performance 1	Performance 2	Performance 3
Example 1	(0.38ev) the story was about (0.48mv) a woman who worked for (0.31mu) charity (1.28eu) and (0.35ev) yeah she was pregnant	the story was about a woman who worked for a charity and (0.6ev) yeah (0.26eu) she was pregnant (0.97eu)	So the story was about a woman (0.88ec) she was pregnant and worked for a charity (**1.73mc**) on the countryside
Example 2	her friend (0.39mv) who was driving the car	her friend who was driving the car (**2.2ec**)	N/A
Example 3	it started to rain very (0.38mu) heavily	it started to rain very heavily	it started to rain very heavily
Example 4	and dropped (0.27mu) in a ravine (*pron*)(1.27eu)	and dropped (0.26mu) in the ravine (*pron*)|	and (0.77eu) drove in a fence

note: 1 numbers in brackets denote pause length. 'e' denotes an end-clause pause 'm' denotes a mid-clause pause. 'v' means that the pause was 'voiced'(e.g. er, um) 'u' means that the pause was silent and 'c' means that this was a composite pause

Likewise, Kimmy also seems to rely heavily on verbatim repetition. As Table 53 demonstrates, a number of complete phrases are repeated during repetitions of the task. In general, these repeated phrases are delivered in one go, without interruption. It seems that this is true particularly between performance one and two: the third performance is a little more 'rogue' in the sense that Kimmy repeats some phrases verbatim but also re-phrases and repackages certain ideas (in Example 1 she repeats the words verbatim on the second performance but then plays about with the order on the third performance). Example 4 is also interesting because it follows a similar pattern, (near) verbatim repetition on the second performance followed by rephrasing on the third. We can also see here that, in general, phrases which were uttered during the first performance often included mid-clause pauses, these are less common on the second and third performances.

As we have seen (Chapter 2), increased fluency is generally linked to theories of automaticity in speech production and increased fluency during TR is often associated with theories of conceptual planning and/or priming. The verbatim repetition and concurrent increased speed and reduction in pausing provides support for this. Given the amount of verbatim repetition, it seems likely that the language that Kimmy is retrieving on subsequent performances is accessed in a speedier fashion because it is primed. This is evidenced by the reduction (and almost elimination) of mid-clause pauses which are understood to be primarily related to lexical searches. It also seems probable, then, that priming results in not only increases in speed at specific points in the speech production model (see discussion of Segalowitz' model in Chapter 2) but also that the speech is produced in an altogether different way (i.e. it is drawn in an entire chunk from memory as opposed to being re-conceptualised and re-formulated).

Lexical problem-solving mechanisms (increased monitoring)

As explained above, it seems that a possible explanation for the increased speed of speech and decreased pausing, may be, at least partially linked to the reduction in conceptual planning and opportunities to prime and subsequently retrieve 'primed' language. However, an additional explanation is that increased attentional capacity and increased monitoring led to further employment of strategies for overcoming communication issues.

Table 54: Example of substitution in Kimmy's performances

Performance 1	Performance 2	Performance 3
yeah (2.32ec) so they (1.85mc) went (1.87mc) away from the road or the because she couldn't see the road (0.43ev) and dropped (0.27mu) in a ravine (*pron*)(1.27eu)	so they (**3.3mc**) get away from the road and dropped (0.26mu) in the ravine (*pron*)	so (0.38eu) her friend (0.33mv) couldn't see the road any more and (0.77eu) drove in a fence

Table 54 shows the same section of the story during performance 1,2 and 3. In just this short section, there are three examples which may provide evidence of increased monitoring:

Firstly, Kimmy seems to struggle with her explanation of the car falling into the ravine. During performance one, this is evidenced by the lengthy mid-clause pauses before and during the incorrect formulation of verb+preposition "went away". This is also observed at performance 2 when there is a very lengthy pause before the incorrect utterance "get away" suggesting that she is still unsure of the collocation. However, in the third performance, this construction is not even attempted. Instead Kimmy relies on the modal construction "her friend couldn't see the road".

Secondly, there is a further example of this hesitation around verb+preposition formulation at the end of the extract. During the first performance, Kimmy pauses in the middle of "dropped in" as she does in the second performance. In the third performance, however, she resorts to "drove in a fence."

Finally, Kimmy incorrectly pronounces the word "ravine" during the on both occasions before apparently abandoning it altogether in the third performance when she chooses a different noun, 'fence' instead. These three examples neatly illustrate how Kimmy is able to develop strategies for dealing with difficult language that allow her to speak more fluently at the third performance.

An interesting question to consider at this stage is why the first two performances seem quite similar but the third is different. A tentative explanation might be that additional attentional resources freed up during performance two because of the familiarity with the task's requirements mean that there are greater available resources for monitoring. I would suggest that this monitoring alerts the learner to perceived deficiencies and episodes of dysfluency which can then be avoided on the third and final performance (see Figure 25 below for a tentative model)

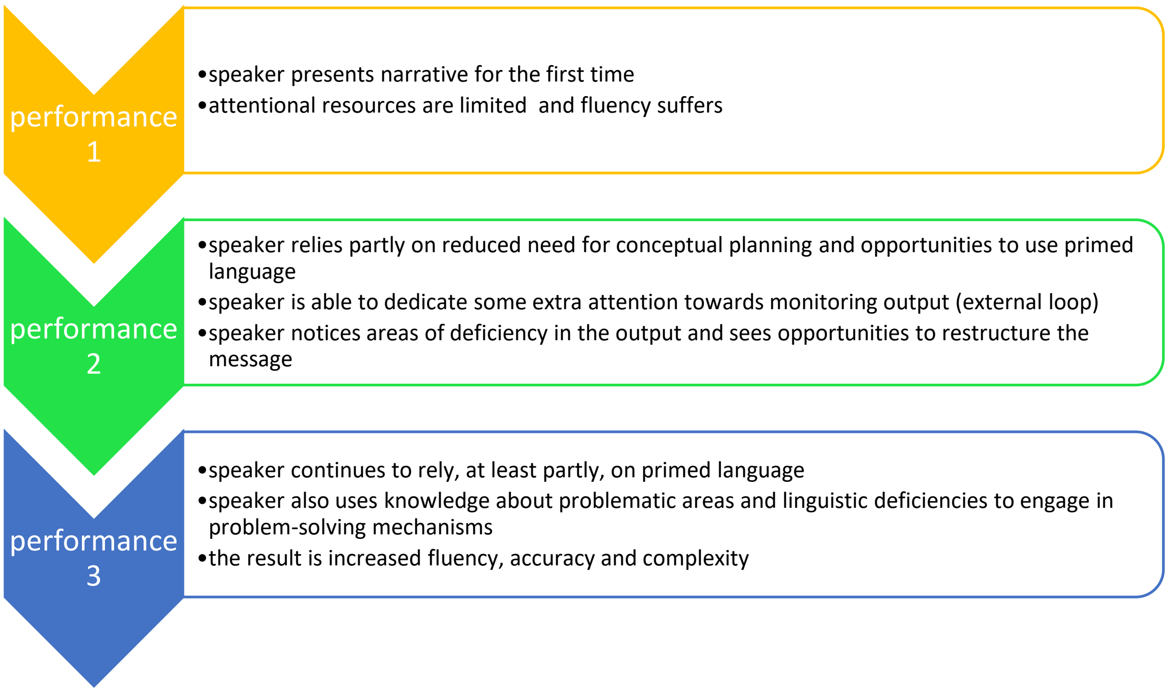


Figure 25: A tentative model for thinking about automaticity and monitoring during TR

11.3 Case 2 – Eva – Procedural Repetition group

11.3.1 Quantitative results

Similar to Kimmy, Eva's quantitative performance scores were fairly typical of the group scores (see Table 55), although her fluency did increase during the session as well as between pre- and post-test.

Table 55: Results of quantitative analysis for Eva

	Pre-test (Day 1)	Session P1 (Day 2)	Session P2	Session P3	Post-test (Day 5)
Speech Rate	163.21	142.10	140.39	187.57	185.5
Mean length of Run	9.71	8.56	6.70	10.72	11.71
Articulation rate	223.70	183.90	230.84	257.01	239.69
Phonation time ratio	77.38	81.75	65.81	74.11	80.9
Frequency of mid-clause pause	3.96	5.19	5.24	2.91	2.8
Mean length of mid-clause pause	.74	.88	.62	.64	1.02
Repair	3.96	1.04	5.24	.97	2.8
Holistic judgement score	7	5	6	8	7
Accuracy	N/A	.79	.90	.87	N/A
Length of AS	N/A	11.42	9.57	14.85	N/A
Subordination	N/A	1.25	1.57	2.08	N/A
Length of clause	N/A	9.13	6.09	7.15	N/A

11.3.2 Session data transcription

Performance 1 (elongated sounds are indicated with " ~ ")

Ok here we have a story about a ca~(0.42)t (1.25mu) a~(0.3)nd a woman (0.96eu)| at the first picture you~(0.6) can see a woman (1.21mu) with her cat (laughter)| a~(0.71)nd it's Christmas time | so~(0.99) (1.05eu) she goes out for shoppi~(0.36)ng |

and she~(0.42) (0.65mv) bought some Christmas prese~(0.57)nt | and yeah she is~(0.35) (0.53mc) diabetic (1.74eu)| and at the second picture you see :: that she's~(0.37) (0.75mu) came home with all her~(0.48) presents and Christmas stuff and everything| and she also~(0.57) ea~(0.62)t (1.03eu) she also ate some~(0.76) sweets (1.1eu) | and after~(0.3) (0.41eu) before she~(0.77) wants :: to go to the bathroom | she used the bathroom (0.38eu)| and suddenly she fell like really <makes noise> strange (laughs) :: because she~(0.38) suffering of diabetes (0.79eu) | and then she felled do~(0.34)wn |

Performance 2

At the first picture (laughing) (0.55eu) you see a guy~(0.42) (0.89eu) | and he ha~(0.28)s a lot of cows (1.66eu) | and in the second pi- (0.35mu)*picture you can see :: that one cow~(0.34) (0.27mu) trie~(0.66)s~(0.39) :: to attacked the guy (1.87eu)| and he tried :: to run away | but the cow was really fast | so the cow~(0.43) (0.39mu) catch the guy :: and attacked him | and he (0.65eu) the cow rammed him like (0.85eu) that he was lying on the floor (0.69eu)| and then the cow jumped on him (0.98eu) yeah (laughter) (0.48eu)| the cow jumped on him | and he was like sitting on the guy (0.83eu) :: and didn't move anymore (0.66eu)| but the guy~(0.36) (1.6mu) didn't want :: to die~(0.36) | so he tried :: to survive (0.27eu)| and he took his fingers :: and put it directly in the eyes of the cow (1.16eu)| so~(0.26) he could (2.72ec) he (0.5mu) could~(1.35) how you say he could (2.2ec) yeah ((reach the eyes))yeah and he could live [end of extract]

Performance 3

Ok the story it's abou~(0.68)t~(0.62) people (1.09eu) :: who~(0.85) went out to do kayaking (1.7eu)| and I think :: I'm not sure :: but it have to be like (0.37mu) in other country :: because (0.79ec) they were kayaking | and everything was ok | <makes noise> and suddenly all of the sudden (0.42eu) they came a hippo | you know the hippo out of the river (0.81eu) :: and attacked one of the guy :: who was inside the kayak (1.29eu) | and (0.53eu) I think (1.36eu) :: I don't know (1.66ec) the other one~(0.33) escaped :: or tried to help him | I don't know | but the other one left the river :: because they were so scared | and the one guy who~(0.52) (0.84mu) got attacked of the hippo (1.05eu) :: was alone (0.98eu)| and the hippo eat (0.72mu) the arm of the guy (1.93eu) | and~(0.68) yeah here I don't know :: what happened (0.7eu) to be honest | so (0.54eu) I think yeah whether the hippo eat the whole body :: or~(0.29) he lost a lot of blood | but at the end of the story he died |

11.3.3 Analysis

A deductive approach was taken to analysis in the sense that the data were analysed to identify reasons for the increase in fluency between pre- and post-test. From the analysis undertaken, there is some evidence to suggest that the PR process enables and allows Eva to develop confidence in using past forms, and to become more fluent as a consequence.

Frequency and accuracy of past forms

Given that one of the main commonalities between the training session tasks and the pre- and post-tests was grammatical tense, the training session transcriptions were consulted to identify behaviour related to simple past tense use. The detailed qualitative analysis of Eva's performances during the PR carousel, revealed differences in relation to frequency of past forms (Table 56). In the first performance, past forms are relatively few. Often the present tense is used.

Table 56: Verb tokens, accuracy and sound lengthening

	Total verb tokens	**Narrative tense tokens (number used accurately)**	**Number of drawls (over 250ms)**	**Drawl directly before verb**	**Drawl during verb**
Performance 1	16	6 (3)	17	5	2
Performance 2	25	16 (14)	10	4	4
Performance 3	27	15 (13)	7	3	0

Sound lengthening

It was noted that Eva had a tendency to elongate sounds immediately before and during a past form 'attempt'. In order to explore this further, I examined Eva's production for examples of elongated sounds. Following Hilton (2009) a drawl was considered to be a sound longer than 250ms. The PRAAT sound files and TEXTGRIDs for each of the three performances were examined once more. This time the runs of speech were investigated for sounds which exceeded 250ms.

The drawls were annotated on the textgrid and information regarding length of drawl was inserted into the orthographic transcriptions. A number of things became clear: Firstly, drawls often seemed to accompany other types of hesitation (filled/unfilled pauses), creating hesitation 'clusters' in the middle of clauses. In the same way that pauses were found to cluster together to create 'composite' pauses (or 'hesitation clusters' in Hilton, 2014), it seems that sound lengthening also occurs as part of these clusters, which, in turn, creates even longer hesitation clusters. A final observation is that, for Eva, the longest drawls and hesitation clusters seemed to occur before past tense structures as in the following examples.

Example 1

| so the cow~(0.43) (0.39mu) catch %the guy :: and attacked him |

Example 2

she's~(0.37) (0.75mu) %came home with all her~(0.48) presents%

This suggests that some of Eva's dysfluency was tied up in uncertainty about the simple past, particularly more complex sequences. What is interesting, however, is that during the second and third performances, Eva attempts to use a much higher number of verbs in the simple past and with greatly increased accuracy. Her tendency to drawl and pause before these forms is also reduced. Overall, then, this suggests that Eva gains some confidence either from attempting these past structures, from receiving input, or a combination of the two. This could be a possible sign that morphological rules relating to regular past tense usage (i.e. regular verb = add –ed), and in particular the morpho-phonological elements of the rules are becoming proceduralised. With regards to irregular past simple use, it is suggested that repetition of these same verbs across performances and across the entire week of the intervention, along with frequent interlocutor input, may have resulted in increased confidence and accuracy in using them, possibly through the memorisation of salient verbs. Here, then we have a possible explanation for increased fluency (especially articulation rate) at the post-tests for Eva. Simple past tense use was given sufficient and *broad* practice during the three sessions of training resulting in increased confidence in applying rules related to regular past tense morphology and/or memorisation of accurate irregular past forms.

11.4 Discussion and implications

Conclusions which are drawn from these case studies are, necessarily, speculative and exploratory in nature. Despite this, they have revealed some possible explanations for the main findings of the research, and also point to some directions for future research.

Firstly, it seems that Kimmy was using verbatim repetition to deliver chunks of speech in a faster, smoother manner. It was assumed that this was due to the combination of reduced conceptual planning and the opportunities to draw on primed language. This finding is tentative but interesting as it attempts to make a link between TR and the particular stages in speech production. Secondly, the observation that Kimmy apparently used task repetition to engage in communicative strategies that the first performance of the task did not afford her, is interesting. It seems that previous engagement with the task gave Kimmy the chance to monitor and notice deficiencies in her speech which she then dealt with on the next performance. The implications of this finding are limited, in that they relate to only one learner in a single context. However, the finding is an intriguing one and future research might examine the relationship among task repetition, monitoring and problem-solving mechanisms with greater numbers of participants.

There are also a number of findings related to the procedural repetition case study. Firstly it seems that PR provided Eva with the opportunity to produce a wider range of lexical items. In particular, it was noted that the learner encountered a wide range of verbs in the simple past for (i.e. he *bought, the cow attacked*). It was suggested that it was this practice of using and receiving input on a wide variety of past forms (both regular and irregular verbs) that contributed to increased fluency at the post-test.

Overall, these case studies, taken together with the group data, provide some tentative explanations for some of the main findings and point towards avenues for future research. In addition, it is hoped that these case studies provide an informative and accessible picture of the sort of language that is produced by learners during both types of carousel task sequence. To that end, it may be of particular interest to language teachers, testers and practitioners.

11.5 Chapter summary

This chapter has presented the findings of two case studies from the current data set. Through detailed examination of Kimmy's task performances, it emerged that there were at least two fluency-relevant processes taking effect. The first was the appearance of language chunks which were performed on subsequent performance much quicker and without interruption. A second observation was that the learner engaged in communicative strategies which lowered the length and frequency of mid-clause pauses and may have contributed to accuracy results but only on the final performance. This finding adds a new perspective to current discussions about the impact of task repetition and, in particular, the relatively recent discovery that *different aspects* of fluency (and therefore speech production processing) might be affected by different numbers of repetitions. Eva's performances, on the other hand, revealed the *breadth* of language that the participants in the PR group were exposed to. It was suggested that it is this breadth of practice of different regular verbs in the simple past may have led to the proceduralisation of related morphosyntactic rules and that frequent exposure to and practice of using common irregular verbs in the simple past may have allowed for the memorisation of those irregular forms. When combined, the overall effect might have been increased confidence in producing a narrative using the past simple, resulting in decreased tendency to elongate sounds prior to a past tense marker and therefore increased rate of speech.

12 – Discussion of findings

12.1 Introduction

In this chapter, I interpret and discuss the findings of this research. The chapter is organised into a number of sections. I begin with a brief summary of the original hypotheses and how they correspond with the research findings. The first substantive section will discuss the findings of the study as they relate to the impact of both carousel task sequences on *short-term* fluency. The second section looks at the findings in terms of *fluency transfer* effects and suggests explanations with reference to specific examples from the data. In the third section, I provide a summary and discussion of the learner perceptions research. I conclude with a short chapter summary.

12.2 Findings overview

Table 57 presents an overview of the original hypotheses and a commentary in light of the research results.

Table 57: Summary of hypotheses and research findings

Hypothesis	Supported?	Commentary
1a) Global fluency will increase during both TR and PR carousel task sequences but fluency will increase to a greater extent for the TR group	Partially	Fluency did increase significantly for the TR group but fluency did not increase for the PR group (*ns)*.
1b) For the TR group, different aspects of fluency (i.e. speed, breakdown,	Partially	Different aspects of fluency did increase on different performances. However speech rate did not increase steadily. Instead it increased

repair) will develop at different points in the carousel cycle. Speech rate will increase steadily over the course of the three performances. Subtler changes relating to breakdown fluency (e.g. mid-clause pausing) and repair will take place only on the third performance, if at all.		more between first and second performances and then levelled off. Breakdown fluency was affected only between first and third performance as hypothesised. Repair fluency was not significantly affected (ns).
1c) Fluency will increase for both the TR and PR groups in terms of holistic rater scoring of fluency.	Yes	Fluency did increase for both groups but *significant* (p = .01) gains were observed only in the TR group.
1d) Speech rate and mean length of run will correlate most strongly with the NS rater scores.	Yes	Speech rate and mean length of run correlated most strongly with the NS rater scores
1e) Complexity and accuracy will increase for both the TR group and PR group. Gains will be greater for the TR group.	Partially	Gains in complexity (subordination) and accuracy were only found in the TR group. Complexity and accuracy did not show a significant increase for the PR group.
1f) For the TR group, complexity and accuracy will increase on the first repetition For the PR group,	No	Accuracy and complexity increased only between the first and third performances and only for the TR group.

fluency will increase on the first repetition while complexity and accuracy will not increase until the third performance		
2)There will be an increase in fluency between pre- and post-test only for those students that took part in TR carousel training	No	There was not a significant increase in fluency between pre- and post-test for the TR group but there *was* a significant increase for the PR group.

12.3 The TR and PR carousels and short-term effects

In this section, I discuss the findings of the studies as they relate to Hypotheses 1a – 1e. I look at the impact of the TR and PR carousel on global measures of fluency (i.e. speech rate; mean length of run) and discuss how this relates to other findings in the TR literature. I also draw on the available literature, the case study analyses as well as other pertinent examples from the data in order to offer theoretical explanations for the main findings. Next, I look at the changes in fluency during the TR carousel as they relate to the fluency dimensions of *speed, breakdown* and *repair.* Again, I compare these findings with those in the relevant literature and I offer explanations for observations. This is followed by an exploration of the holistic score findings and the extent to which they correlate with utterance fluency measures. The final part of this section looks at how *complexity* and *accuracy* were affected by the carousel task sequences and provides explanations.

12.3.1 TR carousel and Global fluency

There was an increase found for Global fluency (i.e. speech rate; mean length of run) for the TR group. This means that Hypothesis 1a was partially supported by the results of the current study. The finding that fluency increased during TR echoes the findings of many studies that have also found increased global fluency (SR/MLR) during task repetition (Wang, 2014;

Sample & Michel, 2014; Lambert et al., 2017; Ahmadian, 2011; Ahmadian & Tavakoli, 2010; Bygate, 2001).

As explained in Chapter 5, only Fukuta (2016) did not find a short-term fluency-enhancing effect for TR. She explained that the lack of an observed increase in fluency might have been due to the delay between performances or the fact that the study only looked at a single repetition. I have argued that priming might have a particularly important role to play in TR effects on fluency and therefore it follows that recency of activation might be particularly salient. The findings of this study would therefore support Fukuta's supposition that fluency effects might be moderated by the distance in time between the original performance and the repetition.

It can therefore be said with some confidence that immediate TR is associated with an increase in short-term and/or task-specific fluency. The important contribution that this study makes, however, is that this it can now be said that this is also true of the immediate, ecological task repetition which characterises the Poster Carousel.

In the studies mentioned above, increased fluency through task repetition has invariably drawn on Levelt's (1989; 1999) models of speech production, insofar as the act of carrying out the task performance prior to the repeated performance is supposed to allow learners to draw on that first performance to increase fluency in the second (or subsequent) performance. However, while all these researchers agree that TR has a "general facilitory effect" on fluency (Segalowitz & Segalowitz, 1993), there seems to be less certainty about the specific cognitive process(es) that are affected by TR. Below, I will attempt to bring together the quantitative and qualitative findings of this study to offer some possible explanations. First, I will explore the impact that verbatim repetition might have had on fluency. Secondly, I consider how increased attention to monitoring processes may have led to message avoidance or replacement in the third performance.

Verbatim repetition, planning and priming

Verbatim repetition is sometimes considered to play an important role in explaining short-term fluency gains (for example see Boers (2014) and Thai and Boers (2016), in reference to the 4-3-2 task sequence). The analysis within this research of Kimmy's performances during the TR carousel revealed verbatim repetition of short phrases and whole chunks of speech.

Kimmy's performances were transcribed and annotated with key information related to pause frequency, location and length. These revealed that phrases uttered with interruptions in the first performance were delivered verbatim but without hesitation in the second and third performances. This was taken as evidence of increased automaticity (through lexical priming) of speech production on the second and third performances.

As explained in Chapter 2, L2 cognitive fluency is believed to be at least partly determined by the extent to which cognitive processes governing L2 speech production are automatic. I also explained that automaticity in L2 speech processing can be thought of as happening in two distinct ways. On one hand, speech may be automatic in that it involves the retrieval of memorised language in a single-step (Kormos, 2006). On the other hand, speech may be automatic because rules related to the TL have gone from being declarative knowledge about the language, through the proceduralisation process and finally become habituated (DeKeyser, 2017; Kormos, 2006).

Based on the existing research in TR effects, there are arguably two underlying reasons for increased automaticity in the speech production process during TR: *conceptual planning* and *lexical priming.* Drawing on the quantitative group results as well as the evidence from the case study, both of these explanations are likely. It was suggested in Chapter 11 that Kimmy's second performance was more fluent because the language needed to deliver her story was primed during the first performance and she was therefore able to draw on the primed language in subsequent performances. In other words, the speaker was able to assemble the message in a speedier fashion because lexical retrieval at fluency vulnerability points in the speech production process was quicker. This finding lends support for thinking about task repetition as much more than providing a form of strategic planning which is arguably concerned only with the conceptualisation or macroplanning stage of speech production. It seems that the act of assembling the speech (i.e. conceptualising, formulating and articulating) means that the language is produced more fluently. In other words, it is likely that *both* planning *and* priming are at play during the Carousel task sequence. However, relying solely on global measures of fluency, it is not possible to say unequivocally whether the group findings for increased fluency during TR were the result of reduction in the need to plan content and/or priming. In order to shed more light on specific speech production processes, it is necessary to look at more refined measures of fluency (see section below).

Problem-solving mechanisms

Another explanation for the TR group's improvements in fluency could be due to the avoidance of problematic structures and lexis. Learners may be taking advantage of increased processing capacity (due to the effects of planning/priming) to monitor their performance more closely, identifying problematic areas of performance (DeKeyser, 2017). Learners might use the third and final performance to deal with these problematic areas in different ways, possibly omitting the information altogether, or replacing problem lexis with better known lexis. This, in turn, seemed to have a further impact on fluency as problematic language tended to be coupled with pausing and hesitation.

Again, this was suggested by the case study analysis. Kimmy's behaviour during the training session suggested that the areas of language that were most problematic in terms of fluency (i.e. those that precipitated long mid-clause pauses) were avoided altogether in the final performance. Often, increased monitoring is discussed in association with increased accuracy and complexity but here I have shown how it can also impact on fluency because dysfluent stretches of speech can be noted and dealt with.

12.3.2 PR carousel and Global fluency

In the current study, and in contrast to Sample and Michel (2014), Pinter (2005) and N. de Jong and Perfetti (2011), there was *no* observed increase in fluency for the PR group. In the aforementioned studies, the increase in fluency observed was attributed to familiarity with the task demands and task structure ('discourse' in Bygate, 2001), in spite of the fact that the content in repeat performance may be novel.

It is important to note, however, that fluency gains for procedural repetition (and TR) in N. de Jong and Perfetti's (2011) study may have been due to the fact that the participants in their study were also performing their tasks (both TR and PR) under increasing time constraints. This also meant that a smaller speech sample was analysed during the second and third performances and this may also have impacted on their results (de Jong, personal communication; Boers, personal communication). In Sample and Michel's (2014) investigation of PR and CAF, gains in fluency could actually be attributed to content planning and priming (i.e. similar to TR) because the repeated versions of the task were actually *identical* in terms of the lexis they required of students. The details of fluency gains for participants in Pinter (2005) is limited to a single observation therefore it is difficult to draw further comparisons between her findings and those in the present study.

Of course, given the argument that TR resulted in increased fluency because of lexical priming, it follows that no, or little, increase for fluency would be observed for the PR group because the lexis required for each performance was different. However, it is important to also acknowledge some additional explanations for the lack of an increase in short-term fluency for the PR group. Firstly, I look at the impact that the competing demands of the PR carousel might have had on short-term fluency. Secondly, I consider the influence of interlocutor involvement and spontaneous speaking.

Competing demands

There are a number of possible theoretical explanations for a lack of fluency increase for the PR group in the current study. Firstly, the particular nature of the PR carousel which involved passing input from one participant to the next, meant that the task necessitated increasing creativity, inference and imagination on the part of the learners, as the input they received each time became less detailed and more vague. The examples below are taken from the beginnings of three separate speeches of students in the PR group.

> S26 - OK I don't have many details for about this story because someone don't explain well
>
> S5 - it's very difficult because I don't understand the story
>
> S77 - OK this is a I think a man who don't know who work maybe in the farm

It is clear that, for many of these students, the PR task was especially challenging because they did not necessarily understand the stories. All three of these students went on to present the stories and produced the same amount of speech overall as they did at the beginning of the carousel, however what often happened was that the story became less clear as it was passed along. See for example the extracts in Figure 26. Here we see how a cow became a *bull* and the farmer became a *bullfighter*. In another interaction the same story was given a particularly grisly ending by one student who insisted that the farmer got a friend to kill all his cows by driving into them with a truck! What is clear, then, is that, in the PR version of the carousel, students have to draw on other linguistic and cognitive skills to complete the task. Due to the reduction in clarity and amount of input, learners need to be imaginative with the language and they need to invest more time interpreting the pictures (i.e. guessing), meaning that they might not be able to dedicate attention to increasing their fluency.

Spontaneous speaking

version #2

ok this is a i think a man who don't know who work maybe in the (farm) not farm like zoo (zoo) the man who have arrived do you know (ah is cow cow) yeah yeah yeah i don't know the name but i'm the cow maybe the cow hurt him (ok possible possible) hurt him yeah it's possible and after a week the people called a ambulance and in the end i think maybe the ambulance arrive later late and the man lost a lot of blood maybe

version #5

this is a story about crazy cow the one day the man it's a like a competition in the with cow you know like with red handkerchief but suddenly a cow the saw a red handkerchief get angry and try to kill yeah try to hit him and after that he fall down on the ground and the cow tried to kill him and the man put him hands up please help me but (but is one cow or cows) one cow one cow tried to kill him but he say help me put him hands up

Figure 26: Showing how the story changes over the course of the PR carousel

As I have discussed above, *comprehension* was a key component of the PR task. Task completion in this task sequence relied upon the transferral of knowledge from one student to the next and so on. The precise demands of the PR task (in contrast with the TR task which was less reciprocal) meant that the performances were, on the whole, more interactive. This meant that students were required to respond to their interlocutors' questions and may have been less able to launch into a pre-prepared (planned) explanation based on the input. The reason for the increased interaction in the PR carousel is presumably due to the 'listeners' eagerness to understand the story so that they are not left floundering when it comes to their turn to present. They often interrupted interlocutors during their explanations in order to clarify meanings, check lexis, pronunciations and so on. Figure 27 below is an exchange that demonstrates interruptions and questions that were fairly characteristic of the PR carousel performances.

The result is that students in the PR group were more likely to be put ‘on the spot’ when they were speaking and not able to draw on any conceptual plan they may have had in their minds. Similarly, they were often unable to ‘get into the flow’ of speech because they were being interrupted more frequently.

> so this is the story of $ four friends | they they was in the park | and they wanted to do kayak (kayak?) yes and $ on a river | so they did they did kayak | but suddenly $ one $ one guy fall on the river :: but was eaten (by snake? who is that?) | is a hippo | (what is? ah hippo) yeah and so one guy falled | and was and the hippo ate the guy | and his friend wanted :: to save his friend | so he jump on the river (he dived into the.. why?)| because the hippo eat this one after the other guy :: so two two men died |

Figure 27: Sample performance from PR group showing interruptions and clarifications

It is certainly possible that any attentional resources which may have been freed up through task familiarity were immediately directed towards coping with the difficulty of dealing with spontaneous questions. The PR carousel was a less predictable, more spontaneous task sequence than the TR carousel and this is a possible reason for the lack of increased fluency observed.

12.3.3 Changes in specific aspects of fluency across repetitions (TR carousel only)

Hypothesis 1b was supported by the results of the study as it was found that different aspects of fluency were affected at different time points for the TR group and there was no change in the fluency of the PR group on any aspect of fluency. For the TR group, speed (articulation rate) and global measures increased on the first repetition of the task and then plateaued, mid-clause pausing decreased gradually over the course of the three performances for the TR group, other breakdown and repair fluency were not significantly affected by TR.

Table 58: Aspects of fluency over three performances

Aspect of fluency	Impact over the 3 performances (TR only)
Global	Significant increase between first and second performances. Sustained between second and third.
Speed	Significant increase between first and second performances. Sustained between second and third.
Breakdown	Significant change only between first and third performances. PTR increased considerably between first and second. FreqMid decreased gradually between first and third
Repair	No significant change

In the following section, I will provide a short breakdown of results for each of these aspects (global measures, speed, breakdown and repair) and discuss with relevance to the literature.

Global measures

For global measures a similar pattern was observed in both speech rate and mean length of run with considerable, significant gains on the second TR performance and smaller, non-significant gains between the second and third TR performances. Many TR studies (e.g. Wang, 2014; Bygate, 1996) only investigated a single repetition of the task and so it is not possible to say what the impact on these measures might have been over three or more performances (i.e. whether speech rate would level off or continue to increase).

Of special interest here, then, are those studies that looked at three or more task performances. Boers (2014) found that during straightforward TR (i.e. *not* 4-3-2) the highest proportion of gains for speech rate occurred between the first and second performance, with much smaller gains between the second and third. This finding was replicated in the current study. Similarly, Thai and Boers (2016) found that for repeated task performances, the speech rate of participants increased most noticeably between the first and second performances and to a much smaller extent between the second and third performance.

The fact that SR was found to plateau after a single repetition is in contrast with Lambert et al. (2017), however. In their study, Lambert et al. (2017) found that speech rate continued to increase between the second and third performances before levelling off. A possible explanation for the differences between their findings and those in the current study (and Boers, 2014; Thai & Boers, 2016) is their decision to reduce the allotted time for participants to complete the task between time two and time three. Although entirely understandable from a practical point of view (the room went very quiet when all the students had finished speaking but the time had not yet run out (Lambert, personal communication)), it is also feasible that the knowledge that they would have reduced time available in which to complete the task, pushed the students to speak at faster speeds. This is consistent with findings from 4-3-2 studies (including Boers, 2014 and Thai & Boers, 2016). It should be noted, however, that Lambert et al. (2017) did not believe that the reduction in allotted time would constitute time pressure because learners "generally did not use all of the time that they had available" (Lambert et al., 2017, p. 11). However, it remains a possibility that the mere suggestion that time was going to decrease (and beeping coming from the timer) might have impacted on the way that the learners oriented themselves to the third (and subsequent) performances.

In terms of explaining the 'plateau' effect for SR and MLR, I would offer that these measures correlate very highly with the pure speed measure (see Appendix) and that there is a limit in how much faster an individual can get regardless of the number of repetitions. I would suggest, therefore, that once speed has reached its peak for a given learner, *other aspects* of fluency then come into play. In other words, it is *because* of increased speed of processing that learners are able to improve their fluency in other ways.

Speed

This study found that articulation rate increased significantly between TR performance one and performance two only. This higher speed of articulation was then maintained at time three. The only other study which has looked at TR and its effect on articulation rate is N. de Jong and Perfetti (2011). In their study, they found that AR did increase significantly during TR but that this was not limited to gains between the first and second performances. Indeed, in most cases, the largest gains were between performance 2 and 3. De Jong and Perfetti's finding is in line with other 4-3-2 findings (Boers, 2014; Thai & Boers, 2016) who found that in the 4-3-2 condition, greatest gains were between Time 2 and 3. They explain that this could be because it was only the final ‘2 minute’ performance which constituted actual time pressure for the learners in their study.

As explained in Chapter 2, articulation rate reflects the extent to which a speaker lengthens sounds and also their tendency to intersperse their speech with micropauses (i.e. not connecting speech, Heike, 1980). Increases in articulation rate are therefore due to the speed at which a speaker moves from one sound to the next. Both micropausing and sound elongation are related to dysfluency because they reflect the extent to which a speaker ‘buys time’ to produce speech. The difference between buying time through sound elongation and through silence is that a speaker is more likely to hold the floor when the elongate sounds (Hilton, 2014). It is likely that the extent to which an L2 speaker elongates sounds as opposed to pausing silently might be a cross-over from L1, or might reflect the speaker’s awareness of interactional norms in relation to turn-taking.

The increase in AR attested to here, then, does represent a reduction in the need to hesitate. An important question is *why* this happens. It seems likely that at least some of the increase in articulation rate is related to the planning and/or priming effects discussed in the previous section. In the following example (Table 59) we can see the difference between Performance one and Performance two for one of the participants in the TR group. A closer look at the times taken to produce these runs of speech reveals that while the first was uttered in a drawn-out fashion, the second is uttered much more quickly. Although retrospective interview information would help support this explanation, it is arguable that the long sound “fo~r” is related to a lexical search for the subsequent item “charity”. During the second performance, there is no sound elongation, suggesting that the word is now ‘ready’ or ‘primed’ for use. In other words, increased AR is due to decreased lexical searching and encoding time.

Table 59: Example of 'priming'

Performance 1:	**Performance 2:**
(0.38ev) the story was abou~t (0.48mv) a woman :: who worked fo~r (0.31mu) charity (1.28eu) |	the story was about a woman :: who worked for a charity

An alternative explanation would be that all reductions in syllable length (increased AR) are due to reduced need to plan the content of the message (i.e. Levelt's *conceptualisation*). According to Skehan, Foster and Shum (2016), it is end-clause pauses that are related to conceptual planning and presumably the same is true for sound elongation. Future research might consider differentiating between sound elongations which occur at the start and end of a clause (i.e. an end-clause 'drawl') and those which occur mid-clause. This might go some way to establishing whether the changes in speed attested to in the present study were the result of conceptual planning and/or lexical priming. What we can do, however, is look in more detail at the breakdown aspect of fluency in order to identify fluency changes which may be related to specific stages in speech production.

Breakdown fluency

The current study operationalised breakdown fluency as both *phonation time ratio* and *frequency of mid-clause pauses.* I deal with each of these separately.

Phonation time ratio

The greatest increase in phonation time ratio came between the first two performances for the TR group, and yet, the increase did not reach the level of significance until the final performance. As explained in Chapter 2, phonation time ratio reflects the amount of time that a speaker spends actually speaking (including all repairs but excluding filled and silent pauses >250ms). It therefore provides an interesting gauge of the extent to which a speaker needs to pause during speaking.

In N. de Jong and Perfetti (2011), the researchers found that phonation time increased significantly for both TR and PR groups. However, once again, it is entirely possible that PTR is affected mainly because of the reduction in available time to complete the task. I have discussed with de Jong (personal communication) the possibility that changes in fluency *over the course of a performance* might be responsible for some of the fluency (and in particular PTR) gains observed in 4-3-2 studies. This is because, if a person is asked to talk about, for example, *what they think about pets* (N. de Jong & Perfetti, 2011) and are given a minute to think about what to say, they may begin with these ideas leading to initially high rates of fluency. As the performance time ticks on, however, the speaker runs out of planned ideas and has to search for additional content. The result is increased pausing towards the end of the performance. If we reduce the allotted time, and therefore also the *length of the sample of speech* that we analyse, we may only be analysing the initial 'burst' of information. Of course, there may also be an impact of just 'being told they are going to have less time to speak', but the fact that N. de Jong and Perfetti (2011), Boers (2014) and Thai and Boers (2016) chose to analyse the entire speech sample in their studies means that we are not able to rule out the impact that fluency changes *over the course of the performance* might have had, particularly on PTR.

All this means that there are no TR studies to which we may reliably compare results in relation to PTR. However, as PTR is related to both frequency and length of pausing, it might be useful to consider what other studies have found in relation to pause phenomena and TR. Lambert et al. (2017), for example, found that the frequency of filled end-clause pauses (as calculated by way of a ratio to total syllables) reduced between performance one and two. They suggested that this was likely to be due to a reduction in the need to plan content.

For reasons of space and succinctness, the current study did not investigate end-clause pauses, yet it is arguable that the significant increase in phonation time for the TR group between time one and two is largely due to a reduction in frequency and/or length of end-clause pause. This would be in line with Lambert et al.'s (2017) finding that end-clause pauses are largely affected between the first and second performances and that the learners in this study are using the first performance to partially 'plan content' resulting in decreased length and/or frequency of end-clause pauses.

Frequency of mid clause pauses

Frequency of mid-clause pauses reduced gradually over the three performances resulting in a significant reduction only on the final performance. This result mirrors that of Lambert et al. (2017), who similarly found a 'step-wise' reduction in frequency of mid-clause pause. Lambert et al. (2017) found that frequency of end-clause pauses (only filled) decreased considerably on the second performance and mid-clause pauses decreased gradually between the first and third performance. The researchers explain these findings in relation to the need to pause to formulate. They suggest that end-clause pauses are more indicative of the need to conceptualise for content and mid-clause pauses represent hesitation which is linked to linguistic formulation (i.e. lexical retrieval, phonological encoding etc). As Lambert et al. (2017, p. 25) explain:

> mid-clause pausing has been found to signal breakdowns in the linguistic encoding process and tends to occur because of difficulty in lexical access or syntactic encoding… in particular, midclause pausing reflects difficulties in retrieving relevant lemmas and accessing the morphosyntactic information associated with it… decreased midclause pausing is thus an indicator of more efficient linguistic encoding mechanisms [which] can be facilitated by priming effects

This line of argumentation has been more recently supported by Skehan, Foster and Shum (2016) and in Huensch and Tracey-Ventura (2017) who have both suggested that end-clause pauses in L2 speech are related to conceptual planning and mid-clause pauses to formulation. The implication of their findings and that of the current study, then is that *at least three performances of the task* might be necessary in order to bring about significant change in the cognitive processes of message formulation.

This is supported in the case study of Kimmy, which revealed that, even during the second performance, the learner continued to pause in the middle of clauses before 'problematic' items of language. These problematic words and phrases were carried over verbatim from the first performance. However, it is argued that increased monitoring at time two meant that these problem areas became more salient for the learner and she was then able to make necessary changes during the third performance. This strategy had the effect of almost eradicating mid-clause pauses entirely by the third performance.

On the whole then, The findings of the current study provide support for Lambert et al.'s (2017) findings that it is only after two performances that learners can start to ameliorate their performance at the level of *formulation.*

Repair

In the current study, there was no significant effect found for TR on the frequency of repair. This is in line with other studies which have similarly found no effect for TR on repair (e.g.

Wang, 2014). Lambert et al. (2017) also found no effect for repair (frequency of overt self-corrections) over the first three performances of the tasks, although they did find a reduction in the need to repair on subsequent task performances.

As mentioned in Chapter 2, there is much less consensus about the contribution that repair makes to perceived fluency and there is also less agreement about the impact that TR might have on this aspect of fluency. If repair represents the extent to which attentional resources are available for online monitoring (Kormos, 2006), then we might expect repair to increase, particularly on the second performance (i.e. when resources are more available). This is the pattern which is observed in the data (see Table 60) even though this result does not reach significance. This pattern, coupled with the increased accuracy at performance three, does suggest that there is a link between repair and TR. It seems possible that repair processes can only become activated when there is additional resources available to monitor and, once certain perceived 'problems' in the output have been dealt with, repair, subsequently decreases as the speech becomes more accurate on additional performances. This is, however, a question for further research.

Table 60: Means and standard deviations for frequency of repair (reformulations + overt self-corrections per minute)

	Performance 1	**Performance 2**	**Performance 3**
Task Repetition	3.03 (1.74)	3.36 (1.93)	2.85 (2.05)
Procedural Repetition	4.11 (2.35)	3.22 (2.23)	3.23 (1.82)

12.3.4 Holistic judgement

Hypothesis 1c refers to the holistic judgements of fluency which were provided by the two raters. The findings closely mirrored the findings from the analysis of global fluency measures (SR and MLR). One difference, however, was that the raters judged second and third performances in the PR group as higher for fluency (although this fell short of significance). This suggests that there may be some qualitative improvements in the participants' speech during the PR carousel which are not detected by the utterance fluency measures used in this study. Post-hoc discussion with the two raters revealed that they were sensitive to the 'ease' with which they spoke and that often this was noticeable in terms of how relaxed the speakers sounded and the extent to which they were able to use humour to convey their feelings.

That holistic judgement was found to be affected by performance is important from a pedagogic point of view. While it is useful for researchers to know that fluency (as defined in very technical terms) increases during task repetition, a teacher is more likely to be convinced by the usefulness of a task sequence when the changes can be *perceived.* Tavakoli and Hunter (2017) have argued that one reason for the divide between research and teaching practice in L2 fluency is that definitions of key constructs are not shared between researchers and teachers. Therefore, the finding that 'fluency' gains are perceivable to teachers as well as being indicated by temporal measures is a potentially useful insight.

As hypothesised (1d), measures for speech rate and mean length of run emerged as the strongest predictors of holistic fluency scores. Speech rate in particular correlated very highly with the individual rater scores and the combined score. Frequency of mid-clause pauses also correlated highly with rater judgements. This provides support for a number of studies which have suggested these measures are the best predictors of perceived scores of fluency (Kormos & Denes, 2004; Hilton, 2014; Derwing et al., 2004) and support for their inclusion as the measures of utterance fluency in the present study. As explained, the reason that speech rate and mean length of run correlate most highly is because they are 'inclusive' measures which integrate speed, breakdown and possibly even repair fluency.

12.3.5 Focus on form

In terms of the impact of the carousels on complexity and accuracy, hypothesis 1e was partially supported. There were significant gains for accuracy (weighted clause ratio) and complexity (amount of subordination) but only between the first and third performance. This stands in contrast to Wang (2014), who found that accuracy and complexity increased significantly on the second performance of the task. The current study instead replicates the findings of Thai and Boers (2016) and Boers (2014) who found that three performances of the same task can impact positively on complexity and accuracy. A possible explanation for this is that there was greater similarity between the tasks and task conditions of this study with the

Boers (2014) and Thai and Boers (2016) studies than with Wang (2014) whose study involved a simultaneous watch-and-narrate task.

Accuracy

The fact that accuracy increased significantly between the first and third performance for the TR group has important implications for language teaching, as teachers will want to know that learners are not using TR to proceduralise inaccurate or non-target language (Foster, 2001; Thai & Boers, 2014). In the TR literature more generally, there is disagreement about whether accuracy is affected by task repetition. While some studies (e.g. Wang, 2014; Boers, 2014; Thai & Boers, 2016) report accuracy gains for TR, others report no effect for accuracy (Bygate, 2001; Kim & Tracy-Ventura, 2013; Hsu, 2017). Others (e.g. Sheppard, 2006; Hsu, 2017) have found that task repetition must be combined with feedback in order to increase accuracy (e.g. teacher feedback, self-transcription).

One possible reason for the variation in findings is that many of the studies which have found an accuracy effect for TR tend to involve *immediate* rather than *delayed* task repetition (see Table 61). As discussed in Ahmadian (2016), Bui, Ahmadian & Hunter (forthcoming) and Ahmadian (2017, personal communication), it is likely that *immediate* repetition is more conducive to lexical priming than delayed repetition (when a task is repeated after a period of days, weeks or months, see also DeKeyser, 2017). Intuitively this would also seem to make sense: even in L1 we can imagine that if we repeat a particular task immediately we will be able to draw more easily on our first performance if it was only a couple of minutes ago than if it was three months ago.

Table 61: TR studies and accuracy findings

Study	Type of repetition	Increase in accuracy
Ahmadian (2011)	Delayed	No
Bygate (2001)	Delayed	No
Gass et al. (1999)	Delayed	No
Hsu (2017)	Delayed	No
Sample & Michel (2014)	Delayed	No
Fukuta (2016)	Delayed (one week)	Yes

Boers (2014)	Immediate	Yes
Lynch & Maclean (2000; 2001)	Immediate	Yes
Thai & Boers (2016)	Immediate	Yes
Wang (2014)	Immediate	Yes
Present study	*Immediate*	*Yes*

Those studies that report increased accuracy through task repetition are generally explained by attentional resources being 'freed up' by task repetition (i.e. Levelt, 1989) meaning that speakers can dedicate more attention to the formulation and articulation of the message (see Wang, 2014; Skehan, 2014; Fukuta, 2016). The current study adds to this body of literature by suggesting that an effect on accuracy appears to require *at least three repetitions* to be realised, This is evidenced in the exchanges in Table 62 which show repetition of inaccurate forms on the second performance followed by accurate production on the third performance.

Accuracy may have improved at the third iteration because participants were able to take advantage of the reduction in processing demands (through conceptual planning/priming) to focus more on monitoring output on the second performance (this is also evidenced by increased repair during the second performance, as discussed above), with the result that areas for improvement in the restructuring of the message are 'noticed' and the third performance can then be used to actively improve on the second performance. This may mean that errors are noticed during the second performance and avoided during the third.

Table 62: Extracts from the three performances for Toby (TR group)

Performance 1	Performance 2	Performance 3
So this is the story of a woman :: who have [ERR] been saved by er his * his [ERR] cat | so it *the (inaudible) was during the night | you know | she is	Ok so this is the story of a woman :: who has been saved by his cat | so during the night she woke * awoke [ERR] er :: to go to the bathroom |	So this is the story of er a woman :: who has been saved by a cat | so during the night she woke :: to go to the bathroom |

[ERR] wake up | and er she
go [ERR] to the bathroom |

Secondly, and as suggested by one of the learners in the focus groups, a further source of influence on accuracy may be the presence of interlocutor feedback. Although the apparent 'lack' of teacher feedback involved in the TR carousel task sequence was a major theme which emerged from the TR focus groups, it was also apparent that learners were actually getting corrective feedback from peers that, on occasion, was taken up and retained for subsequent deliveries. This peer feedback feature of the carousel task sequence was also mentioned in Richards (2011), and Lynch and Maclean (2000; 2001). Figure 28 exemplifies the sort of peer feedback which participants received and which may have contributed to the overall group gains in accuracy:

| and when she went back home :: she went to the bathroom :: and she fail fell out fall (fall down) out yes she fall down in the floor (on the floor) on the floor | and her cat saw hi- saw her |

Figure 28: Exchange highlighting CF from interlocutor

In Figure 28 we can see that the speaker is struggling with some lexical, grammatical and phonological issues within the same utterance "she fail fell out fall.." and the interlocutor assists by offering the correct phrasal verb (i.e. 'fall down'). Note that the tense is not used correctly in either the CF move or the speaker's response. Similarly, in the next AS-unit, the interlocutor provides the recast "*on* the floor" following the speaker's erroneous use of the preposition "*in* the floor". Taken together, the evidence from this study offers some possible explanations as to the reasons *why* improvements in accuracy take place. Further research is needed to verify these explanations.

Complexity

In the present study complexity was found to increase over the course of the three performances. This finding is consistent with much of the TR literature in which syntactic and/or lexical complexity is often found to increase through TR (Sample & Michel, 2014; Bygate, 2001; Wang, 2014). What is interesting in the current study, is that while amount of subordination increased (clauses per AS-unit), the overall length of AS-unit did not increase. This suggests that while additional clauses were being co-ordinated in later performances, the clauses themselves were becoming more compact. This, of course, may be taken as a positive sign in that the speakers are becoming more precise and concise.

A possible explanation for this might be that, in the same way it can result in increased accuracy, increased monitoring of output may play a key role to play in complexity development. This is because at the second or third repetition, having explained a particular narrative once and then a second time with increased attention to monitor, it will be more evident for speakers where there are logical connections to be made in the story. There is evidence to support this from the data. In the example below (Table 63), for example, a student explained their story during the first performance using mainly discrete mono-clausal AS-units. This was built upon slightly during the second performance and then even further on the final performance.

Table 63: Changes in complexity between performance 2 and 3 for student 50

Performance 2	**Performance 3**
| and then suddenly he heard something behind him | and it's a baby | the baby is fallen out of the window |	|and suddenly he heard that :: something is falling off out of the window | and when he looked :: he sees :: that it's a baby |

note: '|' denotes end of AS-unit '::' denotes clause boundary (within AS)

By way of an interim summary, I have linked global fluency increases to opportunities for conceptual planning and lexical priming and suggested that the increased automaticity in speech production allows learners the resources to monitor their message more closely which, in turn, gives rise to increased accuracy, complexity and more nuanced aspects of fluency (such as a reduction in mid-clause pausing). This argument was supported by the fact that there were no short-term fluency gains observed for the PR group (who had little or no opportunity to use planned/primed language). Alternative explanations given for the lack of changes in fluency for the PR group were the increased task demands and the emphasis that the PR task sequence placed on responding to interlocutors' questions.

12.4 The carousel task sequence and fluency transfer effects

In this section, I discuss the findings as related to Hypothesis 2. Hypothesis 2 was not supported by the current research because there was no fluency transfer effect for the TR group. However, there was a significant transfer effect for the PR group. I firstly look at the impact of the TR carousel on fluency transfer. I compare findings with the relevant literature and I offer a number of interpretations of the findings. Next, I turn to the impact of the PR carousel on fluency transfer. I try to situate findings within the landscape of existing research on long-term PR effects and I also provide explanations for them.

12.4.1 Task repetition and (lack of) fluency transfer effects

That there were no gains in fluency transfer found for the TR group replicates the findings of a number of studies (e.g. Bygate, 2001; Kim & Tracy-Ventura, 2013; Hsu, 2017) that also found no long-term/transfer fluency effect for TR. However, the finding is in contrast with N. de Jong and Perfetti (2011) who found some fluency transfer benefit. There are a number of possible explanations for this discrepancy. One possible explanation, and one that I have mentioned before, is that, along with task repetition, the participants in N. de Jong and Perfetti (2011) also performed their tasks under increasing time pressure. Other recent 4-3-2 studies such as Boers (2014) and Thai and Boers (2016) have looked only at online task performance data. It is therefore possible that the fluency transfer gains observed in N. de Jong and Perfetti (2011) are somehow linked to the fact that the intervention involved time pressure. Another difference between the present study and N. de Jong and Perfetti (2011) is that their participants were also given feedback on their performances as well as instruction specifically related to fluency. Elsewhere in the literature it has been suggested that this might be key for TR effects to transfer (Sheppard 2006; Hsu, 2017). Lynch and Maclean (2000) also suggest that there is:

> "a need for teachers to follow up task-based practice with 'noticing' activities, so that we can help learners consolidate for the longer term what may otherwise be fragile changes in their interlanguage" (Lynch & Maclean, 2000, p. 245)

One final difference between this study and N. de Jong and Perfetti (2011) is that the present study was larger in size. It is possible that any significant findings for fluency transfer in N. de Jong and Perfetti (2011) were linked to the relatively small number of participants.

What possible theoretical reasons might there be for the present study's finding that fluency increases did not transfer? One possible explanation might be related to lexis-specific automaticity. If I return for a moment to the findings for the *short-term* fluency effects of the poster carousel, it will be remembered that increased fluency seemed to be due to the priming of specific lexis. Following on from this, then, the lack of a transfer effect for fluency as a result of TR is not greatly surprising. Increased automaticity based on lexical priming is lexis-dependent. It is specific to the story being told. Therefore, the fact that the TR carousel increased fluency in a lexis-specific way means it is probable that fluency transfer is *only* likely to be on performances which require that specific language. The post-tests were different, not only in terms of lexical content but also in terms of the task demands. Lightbown (2008), in a thorough exploration of the literature on transfer effects of instruction, explains that instruction that elicits the same conditions for language processing as will be required for transfer is likely to have the most success. This includes such aspects as demands on processing. It is therefore offered as a potential reason for the lack of transfer effect observed in this study. This could therefore be a reason for the lack of a TR 'transfer effect' observed in the current research and also in similar studies which use pre- and post-tests that are very different in content and demands from the intervention tasks (e.g. Kim & Tracy-Ventura, 2013; Patanasorn, 2010). Indeed, Kim and Tracy-Ventura (2013) acknowledge this as a potential reason for the observed lack of transfer effect.

Another consideration when thinking about the fluency transfer efffect of the TR carousel, and an important one in terms of pedagogic implications, is that the extent to which the fluency effects of the TR carousel transfer might best be measured in alternative ways other than CAF in performance. As we have seen, fluency *does* increase in the short-term and it seems that learners actually *use* that increased cognitive fluency to monitor and ameliorate their performance in other ways. It is possible, then, that it is the increased *awareness* which is possible through TR that is key to its language learning potential. In other words, it's what learners do *because* of their increased fluency that is important, not the increased fluency itself. If, as Kormos (2006) has suggested, that monitoring might be key to 'noticing' (and, therefore, language *acquisition*) then it seems that TR might allow learners to interact in a mode of 'heightened sensitivity' in which they are more perceptive and aware of their own language and of input from their interlocutor which may be conducive for acquisition.

However, there is nothing to say that developments of this kind would immediately 'show up' on a post-test as increased fluency (Larsen-Freeman, 2009). As we have seen, fluency is affected in both the short- and long-term by so many factors that it is entirely likely that any restructuring of interlanguage brought about by TR might not be translated into increased speech rate on a particular task. What we need to do then as researchers is step back and think more carefully about how developments in interlanguage can be measured which go beyond performance on a single task.

12.4.2 Procedural repetition and fluency transfer

We move now to an unexpected and interesting finding of the current study: that *procedural* repetition did seem to have an effect on learners' fluency on a completely different task. This stands in contrast to other studies in this area: while some studies (e.g. Kim & Tracey-Ventura, 2013) have found a transfer effect for PR on complexity, none have found a transfer effect for fluency.

In particular, it was the *speed* of speaking that increased for the PR group between pre- and post-test. As I have explained, articulation rate (the measure of speed in this study) can arguably be affected by two factors: a decrease in average syllable length, or the linking together of words (i.e. omitting 'micropauses'). In other words, articulation rate gives an indication of the speed with which a speaker moves from one sound to the next in connected speech. Syllable length is likely to be related to hesitation (Dörnyei & Kormos, 1998) in that learners can 'buy time' for thinking by drawling (i.e. lengthening sounds). The linking of words, on the other hand (reduction in micro-pauses) can be related to hesitation but may also be linked to articulatory phenomena (e.g. consonant attraction, Heike, 1989).

In what follows, I explore two possible explanations for the finding that PR impacted on fluency transfer. These include (1) increased confidence in using past forms (proceduralisation); and (2) lower starting level of fluency.

Confidence in using past forms (proceduralisation/memorisation)

One possible reason why PR impacted upon fluency transfer is that the changing task content resulted in increased *familiarity* with past tense morphology. Intuitively it seems to make sense that providing learners with the need to use a wide range of exemplars within a single grammatical form may provide learners with a) the necessary practice to move from declarative knowledge to procedural knowledge in relation to grammatical rules and b) a sufficient amount of input and subsequent output repetition of irregular forms to promote memorisation. Taken together, these two occurrences might be expected to promote fluency *on a new task* because the speaker's overall confidence in attempting the simple past is increased.

The case study of Eva supports this interpretation of the results. During the first performance, past tense usage was lower than subsequent performances. When the simple past tense (both regular and irregular verbs) was attempted during the first performance, it was often characterised by errors and considerable pausing and drawling. This was taken as evidence for her hesitancy in attempting past forms. However, during the second and third performances, frequency of simple past use increases and error in using past tense reduces. The extent to which she hesitates (sound lengthening) before using the past tense also reduces. There is some support for this explanation elsewhere in the literature: Although Patanasorn (2010) did not find that fluency did not transfer to a new task as a result of PR, he did find that accuracy in using past forms was higher on a new task for the PR group than the content and task repetition groups. He explains that this is because they had the opportunity to practice with a much wider range of verbs.

The learners in the current study would have been presented with morpho-syntactic rules related to regular past tense use in a formal instructional setting they will also have been provided with lists of irregular verbs and their conjugation in the past simple. However, as many teachers will be painfully aware, knowing about the rules and identifying deviances from the rules does not mean that a learner can then use this information in spontaneous speech. Indeed, as I set out in the opening chapter, one of the main aims of language instruction should be to provide the necessary practice to allow learners to proceduralise rules and memorise and automatise (memorise) other aspects of the language that are not rule-based (e.g. irregular verbs in the past).

It is possible, then, that the repeated exposure to a wide range of verbs in the past tense may have helped the PR learners to practice and (partially) proceduralise rules relating to the morpho-phonological properties of past forms and memorise certain irregular verbs[20]. As suggested by the findings of the case study, this may have manifested itself in reduced drawling before past tense indicators and therefore increased articulation rate (speed) overall.

Lower initial levels of fluency

Although no significant differences in terms of fluency measures existed between groups at the outset of the study, it cannot be ignored that the PR group did seem to be generally lower in proficiency. This might be expected to impact upon fluency *gains* because this group had lower initial levers of fluency and therefore further to go while the TR group may have been already operating at reasonably high levels of fluency. This was an inherent risk of using intact groups of learners; that some groups, for any number of reasons, might display different CAF profiles. Although all the groups were selected at the same proficiency level, two of the groups used for the PR carousel were at the beginning of the level (i.e. they had just been 'moved up'). Again, this was an unavoidable consequence of randomly assigning groups to different conditions. This is discussed further in the limitations section in the next chapter along with the possibility of using ANCOVA analysis to help detangle these factors.

To give a short interim summary, the fact that there was no fluency transfer for the TR group was explained in terms of the lexis-specific nature of increased fluency during the TR training session. I also suggested that the fluency increase during TR might be better thought of as facilitative as opposed to an end (which can somehow be transferred) in and of itself. I gave two possible reasons for the observation of a fluency transfer effect which for the PR group. Firstly, I suggested that exposure and practice of a range of past forms may have resulted in proceduralisation and memorisation which transferred to the post-test because both training

[20] A question might be why, then, were there no gains in accuracy observed for the PR group. A possible answer would be that, for reasons set out in chapter 9, errors regarding tense were not included in the accuracy scoring unless they were particularly salient. It is possible that accuracy gains may have been observed if a stricter approach to scoring accuracy had been used.

session tasks and the pre- and post-tests required learners to work with past tense constructions. The alternative explanation was that the PR group had lower fluency at the outset of the study and therefore had 'further to go' in terms of fluency improvements.

12.5 Focus group findings and discussion

In this section, I explore the themes of the focus group sessions which were held with participants in the TR and PR task sequences. In total, there were seven focus groups. Each focus group took place at the end of the week. Present during the focus groups were all those learners that had been involved in the training sessions that week as well as the regular class teacher and myself. The focus groups can be described as 'semi-structured' because they involved minimal input from the researcher although facilitative questions were provided to generate discussion. The class teacher often assisted in this process. The focus groups were audio-recorded and transcribed in their entirety. Focus groups relating to TR and PR were grouped together accordingly in order to establish different attitudes towards the two different types of repetition. A thematic approach was taken to the analysis. In this section, I first present the main themes to emerge from the TR focus groups before moving on to discuss the PR focus group findings.

12.5.1 Task repetition

Overall, the impression gained from studying the TR focus group transcriptions is mixed. Three main themes emerged from the TR focus group data: 'improvements in performance', 'boredom' and 'feedback'. I will present each in turn and provide examples from the focus group transcriptions. In addition, I will discuss how these findings support or contradict other observations in the TR literature before presenting some explanations for these findings and the implications.

Improvements

A number of learners pointed out that their performances were 'improved' on subsequent deliveries as in the following example:

> **Student 6:** Yeah actually it er helped us to improve speaking

Student 8: I'd say maybe as we repeated the same er story er for three times I guess yeah sometimes the phrases that ..maybe I used were a bit.. complete

Researcher: When? The first time or the third time?

S8: The third time

Researcher: The third time

S8: It was more fluent yes I'd say

Student 2: yes we can improve er our story because we we repeat it three times I think that the the last one is always the better one because we I think we can see in which way we are er more understood or I don't know how to say that

As this extract demonstrates, some learners were under the impression that the task sequence gave them the opportunity to improve their performance.

Student 1: er when you talk about er tell the story about second time third time yeah it's not about using different phrases er how to explain the scene sometimes we add some more or our comments the next time we get used to the story so we maybe become more confident about it

In the second extract, the learner refers to adding "some more of our comments" and being more "confident". In general, this 'awareness' provides some support for Ahmadian et al. (2017) and Lynch and Maclean (2000; 2001) in that some learners do seem to be aware of the benefits of TR.

It would have been very interesting to see if there was any correlation between attitudes to the task sequence and performance on the tasks. Unfortunately, it was beyond the scope of the current study to see how learner perceptions correspond with fluency scores across performances, although this is an avenue that could prove fruitful.

Boredom

A second theme which emerged from the TR focus groups was 'boredom'. This theme cropped up in all three TR focus groups. It is clear that a number of learners felt that the structure of the TR task sequence was overly repetitive, to the extent that three performances were considered unnecessary and also that, as a whole, many were unconvinced that the task sequence as a whole would be fun to do on a regular basis (i.e. it was ok as a one-off). These two views are highlighted in the following examples from the data:

> **Focus group 3**
>
> **Student 30**: only thing maybe em was a little bit boring to explain it three times I know that's good to do because you change every time the story a little bit and you er can say more details about the story but it was sometimes boring to repeat and to repeat"
>
> **Researcher:** is there anything I mean is there anything that would have made it less boring?
>
> **S30:** yeah to explain it onetime and then to swap but maybe not that took too much time the stories were very good um no I think it was good yeah

This example from the focus group 3 presents a view that was fairly typical of the data; although learners did refer to the 'boredom factor' ("it was sometimes boring to repeat and repeat"), they often hedged this opinion with an acknowledgment of the benefits of the task sequence ("I know that's good to do because you change every time the story a little bit and you er can say more details about the story").

Often, there was a divide between those learners who found it boring and those who focused on the perceived benefits, as the following exchange demonstrates:

> **Focus group 2**
>
> **Student 1:** [the task sequence] was useful actually because er it help us to start thinking start organising our thought our thinking er and er get a vision of what we are we are planning to say er so it was very helpful
>
> **Student 2:** maybe er tell the story three times it's too much I think yeah maybe two twice will be better
>
> **S1**: Why? Because it gets boring?
>
> **S2**: yes yes one time is ok and then the second time because it's easier because you have already say it but after it yeah it's boring
>
> **Student 3:** you don't think when you talk something…

S2: yeah I think the second time was always the best the third you was bored you say yeah it's this story like this this this

S1: in my opinion if it was boring you er just er start organising the words and learn how to put the words in the correct orders and develop the way that you are going to say the same things to the other people so it (inaudible) the way you are going to tell the story actually and make better in learning the language and telling the story this is the most important thing in repeating and you learn more er repeating

S2: yeah of course of course

S1: this is the most important thing of the lesson itself I think my opinion

Others alluded to the fact that they had taken part in multiple carousels over the course of the week and felt that this was too much.

S2: I think for some days it's ok but after a week it would be a little bit boring

S1: Yeah I think so

S2: because yeah it's just always the same yeah I think for one week it's ok to do thinks like that yeah

S1 yeah yeah just for one week but yeah we did exactly the same thing during the three classes we had like read the story drawing something explain to the others make you know a turn in the class to to go in each poster and yeah tell the story again and again

Taken together, it is impossible to ignore the fact that the learners in this group *did* feel that TR could be 'boring'. It's true that this did not apply to all learners and some of those that did mention feeling bored also acknowledged the benefits. Nonetheless, there is support here for Kim (2013) and Van de Guchte et al. (2015) who similarly found (through different means) that learners might be less engaged while repeating the exact same task.

The findings differ from those of Ahmadian et al. (2017), Lynch and Maclean (2000; 2001) and Lambert et al. (2017), who found that learners did not find immediate task repetition boring, although this could be explained firstly by the fact that the current study required learners to take part in three carousel training sessions in as many days and, secondly, because, in comparison to Lambert et al. (2017), the current study took place during a scheduled lesson which the participants were actually *paying* for while Lambert et al.'s (2017) participants were volunteers who were reimbursed for taking part in the study.

It is surprising, however, that the findings differ from Lynch and Maclean (2000: 2001) who used a very similar task and cohort. An explanation for the differences here, then could be the different methods for data elicitation (questionnaire versus focus group). For example, it is possible that the focus group context created an atmosphere in which learners felt more willing to speak candidly.

Corrective Feedback

The theme of feedback emerged through the analysis of the TR focus group data. Feedback, or rather, the lack of it, was discussed in all three TR focus groups, making 'feedback' a very prominent theme in this data set. The following exchange from Focus group 1 exemplifies this theme:

> **Student 2**: Yeah I don't know if it's so useful to um to tell the same story again and again because if you make some failures
>
> **Student 1**: And you always say the same things for me it was a yeah I was always
>
> **S2**: and you do always the same faults
>
> **S1**: Yeah using the same words and yeah nobody was able to correct me so yeah that's why but I think that for the last class we had it was on Friday it was yeah it was quite interesting because I think that the story were more difficult to understand so it was uh because you can have you can learn some vocabulary and new context so that's why yeah that was good the last one was quite good
>
> **S5**: I we we are not er we don't receive correction so I think we we can't correct ourselves so I don't maybe I don't think we really improve

The finding that learners want more feedback and teacher involvement is not unique to this study or to task repetition. Lyster, Saito and Sato (2013), for example found that learners want more feedback than teachers would like to provide and, as already mentioned, often the class set-up (e.g. pair- and group-work) means that the teacher cannot be everywhere at once. It comes as little surprise, then, that teachers are keen to know in what way and when they should intervene during task performance (Samuda & Bygate, 2008). Some suggestions are presented in the section on pedagogic implications in the next chapter.

12.5.2 Procedural repetition

The learners' attitudes to PR were generally positive. No learner reported being bored by the task sequence. Instead, two main themes emerged from the PR focus group data: 'comprehension' and 'vocabulary'.

Comprehension and communication

A theme which emerged from the PR data was the emphasis that the task sequence placed on communication (effort) and comprehension skills as the following two extracts highlight.

> **Student 44**: I think it's really because you have to describe all for example when we drew some pictures we have to describe it with a lot of details and we have to find even when you don't know vocabulary you have to try to describe it that the people understand it and this is like a good way to learn a language to find always a way to explain things to other people

> **Student 60**: I think it's a very good way to study English because um the best thing er the most important thing is er maybe we heard the story and after one minute we have to we have to speak about it what did the previous person say so we have to we have to move our brain so quickly so it's no time to rest so yeah I think it's good way to study English I like this kind of studying English

It is possible that the second student's comment about moving their brain so quickly reflects the flexibility which the task demands in terms of constantly alternating storyteller and interviewer roles.

Vocabulary

A number of participants commented that it was interesting to work with lots of different stories and one even mentioned that if the story had been the same each time (as in the TR group) it may have made the task easier but it would have been boring.

In the following example, a student is explaining which part of the procedure he enjoyed the most.

> **Student**: …when er we saw the picture and describe what the picture happen for me it was the best because I learn a lot more different words like my vocabulary

This would seem to support the quantitative and case study findings in the sense that the PR group were exposed to more linguistic input and practice. It might also point to another pedagogic use for the PR carousel: that of vocabulary learning.

12.6 Chapter summary

In this chapter, I have discussed the findings of the study as they relate to short-term fluency (Hypotheses 1a-e), fluency transfer (Hypothesis 2) and learner perceptions of the task sequences (Hypothesis 3). Firstly, I explored the effect of the two versions of the carousel on short-term fluency. I first dealt with the TR carousel effects on fluency, linking findings from the case study analysis to the group results. I suggested that conceptual planning and lexical priming play key roles in the observed increase in fluency for the TR group. I also suggested that increased opportunity to monitor performance resulted in the use of certain avoidance strategies which further increased utterance fluency at Performance three. In terms of the PR carousel, I explained that possible reasons for a lack of fluency gains could be the complex demands of the task and the need to respond to interlocutors' questions spontaneously. I went on to explain how different aspects of fluency were affected by the TR carousel. I noted that speed and global measures were mainly affected between the first and second performances and that this was likely due to lexical priming as well as reduced content planning. Increases in accuracy and complexity were explained in relation to increased monitoring opportunities and peer-to-peer negotiation of meaning.

In the second section, I discussed the impact that the carousel task sequences had on fluency transfer. Here I explained the lack of long-term gains for the TR group in terms of (a lack of) transfer appropriate processing and lexis-specific priming. I provided explanations for the long-term gains observed for the PR group, among these, the increased confidence and accuracy in using the simple past and reduced L2.

In the third section, I presented the findings of the focus group analysis. I explained that those in the TR group *did* mention boredom while the PR group did not. The TR focus groups revealed a trade-off whereby learners seemed to be aware of the purpose and some potential benefits but they also found it boring and frustrating to not have any feedback on their performance. On the whole they seemed to find repetition without feedback frustrating because they felt they were repeating incorrect language. The lack of feedback did not come up in any of the PR focus groups, instead the key themes were around the need for effort to comprehend the input received, and the opportunities for learning new vocabulary.

In the next chapter, I explore the implications of these findings, discuss the limitations of the research and recommend some avenues for future research.

13 – Theoretical, methodological and pedagogic implications

13.1 Introduction

In this chapter, I outline the implications of the present research in three main domains: theoretical, methodological and pedagogic. In terms of the theoretical implications, I focus on how the current study contributes to perspectives on automaticity in L2 speech production. I also suggest that the findings can fit into Segalowitz' (2010) framework for thinking about ISLA and the impact on fluency transfer. The methodological implications section outlines the strengths and weaknesses of a mixed methods approach, the challenges and tribulations of conducting classroom-based research and the contributions that this study has made in terms of speech analysis. Pedagogic implications include a discussion of the pros and cons of each type of task repetition, ideas for how to 'sell' repetition to learners and teachers, ways to integrate feedback and options and ideas for personalising the carousel task sequence. To conclude, I consider the extent to which the findings of this research are generalisable to other contexts, the limitations of the study and some possible directions for future research.

13.2 Theoretical implications

The literature on TR/PR effects on CAF has drawn primarily on theories related to speech production processes with the majority of TR studies attributing rises in fluency during TR to a reduction in the need for the speaker to plan the message content (Bygate, 2001). A secondary factor is often suggested to be the so-called 'priming' effects of having already produced a particular message and being able to draw on the pre-verbal plan for that message on a subsequent occasion (Wang, 2014; Skehan, 2014; Lambert et al., 2017). The fluency transfer effects of TR are attributed to 'proceduralisation' (de Jong & Perfetti, 2011).

It has to be said that, with a few exceptions (e.g. Lambert et al., 2017), researchers often allude to these theories but provide little indication of the precise ways in which TR is believed to impact on fluency. Another crucial question which often fails to be asked and answered is: *so what?* What does it matter if fluency increases in the short-term? What is the impact on the learner? Or on learning?

The current study has attempted to shed some light on these questions and look for specific ways in which TR might impact on short-term fluency and fluency transfer, offering support for theories related to speech production.

In this section, I describe these theoretical implications of the research, focusing on evidence which links the processes of TR and PR with specific aspects of L2 speech production processing. First, I look at the impact of strategic *planning* processes. Secondly, I tackle the slippery issue of '*priming*' and its relationship with TR. Next I look at how both these processes can be linked to *monitoring*. Finally, I explain how procedural repetition might ultimately be related to a *proceduralisation* route to automaticity.

13.2.1 Planning and Priming

As we have seen, task repetition is often suggested to provide speakers with conceptual planning in the sense that the first engagement of the task allows learners to plan content which leads to a more fluent delivery on the second attempt because attention that would be diverted towards conceptualisation can be available for the microplanning, formulation and articulation stages of production. It is also suggested that the fact that TR gives learners a chance to actually produce speech means that language is primed on subsequent performances: something which might also be believed to impact on utterance fluency.

I have argued that the increased fluency which was observed for the learners during the TR carousel is related to both conceptual planning *and* priming. Ahmadian (2016) and Ahmadian and Bui (in prep) have suggested that immediate repetition sees greater increases in fluency than spaced or delayed repetition. I would suggest that the reason for this is that priming effects are greater if the time between the initial performance and repeated performance is lessened. I would also argue that this goes some way to explaining the findings in the literature that *immediate* repetition seems to result in increased accuracy as well as fluency. If cognitive fluency (i.e. lexical retrieval) is faster because of priming then the amount of attentional capacity which is freed up by repetition is much greater if that repetition is immediate. In turn, this means that more resources can be directed towards the *monitoring* of the message, which is key to increasing complexity and accuracy. In other words, this study has supported the line of argumentation that immediate TR might be more conducive for performance than delayed TR.

This study has also highlighted the relevance of sound elongation (drawls) for establishing how changes in speed of speech and articulation may be linked to either conceptual planning or lexical priming. I have suggested that those elongated sounds could be treated as hesitations and therefore analysed similarly to filled and silent pauses, including information about their position in a clause. I have argued that those elongated sounds which come in the middle of clauses are likely indicative of difficulties related to lexical and syntactic encoding. To date, there are few studies that have considered sound elongation as a measure of utterance fluency. Future work in this area would be valuable.

13.2.2 Monitoring

The current study has foregrounded the role of monitoring in fluency, accuracy and complexity gains during task repetition. The analysis of the quantitative data (increased tendency to repair, albeit *ns*), taken together with the case studies provided evidence of increased monitoring at Time two and consequently significant increases in accuracy and complexity at Time three. It is argued that increased cognitive fluency during the production of speech at time two frees up attention which can be directed towards monitoring output.
Further research into the role that monitoring plays during TR would be welcome which seeks to integrate the role of monitoring into Skehan's trade-off hypothesis. In particular, it would be useful to use post-hoc stimulated recall interviews with learners to gain more of an insight into their thought processes during TR.

13.2.3 Proceduralisation

The finding that there were long-term gains in fluency for the PR group when compared to TR and control suggests that there has been some significant change in the cognitive processes of speech production for this group of participants. Kormos (2006, p. 156) explains that "two interrelated processes are responsible for the (long-term) development of fluency in L2: automatization of encoding processes and the use of prefabricated language units". This study has suggested that, while short-term fluency gains through TR are likely linked to be related to conceptual planning and priming, long-term gains for the PR group may be linked to the automatization of encoding processes of (in particular) morpho-phonological rules related to regular past tense grammatical forms and the memorisation of irregular simple past forms..

The support for this line of argumentation comes from the fact that the case study analysis suggests that, at least for that learner, hesitations which were related to past tense use reduced during the intervention, while the frequency of past tense use and accuracy of use actually increased. I took this as suggestive of a general increase in confidence in using past forms, theoretically linked to proceduralisation. Findings of other studies (Kim & Tracy-Ventura, 2013; Patanasorn, 2010) in which PR resulted in long-term improvements in accuracy and complexity could also be taken as support for this line of argumentation.

13.2.4 A framework for thinking about fluency transfer and formal instruction

Although I have offered one potential explanation for fluency transfer for PR (see above), along with Larsen-Freeman (2009), Norris and Ortega (2009), Lambert and Kormos (2014) and Segalowitz (2010) I also acknowledge that It is likely that many factors are at play which may exert their influence on CAF over time. Therefore, in order to cite the theoretical findings of this research in some sort of broader context and allow for systematic comparison with other (and future) studies, it is useful to refer to a framework for thinking about the connection between instructed SLA (ISLA) and fluency transfer. Segalowitz (2010, p. 7-8), for example, explains that frameworks are very useful for discussing research findings:

> Frameworks for thinking about issues can be most useful. A framework helps to integrate results, at least to clarify connections in what otherwise might appear to be

> disparate research findings. A good framework for fluency would prompt new questions in a systematic way and stimulate thinking that goes beyond simply looking for practical answers to questions about language training; it would guide the search for answers grounded in theory. Finally, a good framework holds the seeds of its own destruction insofar as it is able to raise questions that will eventually make clear its own limitations and point in the direction of a new framework that will replace it.

Segalowitz' (2010) provides a cognitive science perspective on language instruction and fluency transfer. He suggests that there are five main ideas which emerge from a cognitive science perspective of ISLA and fluency (see Table 64). These are: transfer appropriate processing; genuine communication; repetition; motivation and existing fluency level. Firstly, he suggests that fluency is only likely to be influenced by instruction which engages learners in *transfer appropriate processing.* This means that the sort of cognitive processing which learners engage in within a classroom needs to be the same as those which are required for communication in the 'real world' and, although he uses this term more broadly than can be found elsewhere in the psycholinguistic literature (e.g. Blaxton, 1989; Morris et al., 1977) it is similar to a notion of TAP as described in Lightbown (2008) in which TAP incorporates a number of ideas related to similarities between *learning* context and *test* context. Secondly, Segalowitz suggests that learners need to be exposed to the necessary linguistic resources in situations when they are needed for genuinely communicative purposes. The third criteria is that tasks are repetitive in order to "activate and reactivate the same set of cognitive processes" (p. 176). In addition he notes that individual differences such as L2 sense of self and current level of fluency will also have a knock-on effect on fluency development.

Table 64 A framework for thinking about ISLA and fluency transfer (adapted from Segalowitz, 2010)

Concept	Explanation
Transfer appropriate processing	"transfer of learning to a new setting will be most effective if the kinds of mental processing required at the time of learning match those that will be elicited in the transfer context" (p.173-174)

Genuine communication	"exposure to [linguistic] resources should occur in a context where learners have to use them to fulfil genuinely communicative goals" (p. 176)
Repetition	"learning activities in the classroom should..provide learners with opportunities for systematic repetition in order to activate and reactivate the same set of cognitive processes" (p.176)
L2 sense of self	"L2-specific sense of self plays an important role in the acquisition process leading to fluency, and …this L2-specific sense of self can change (grow) as learning progresses" (p.177)
Existing fluency level	"perhaps by achieving high levels of cognitive fluency through instruction, the neurocognitive network underlying the planning, assembly, and execution of L2 utterances can become more self-sustaining" (p177)

We can therefore view the findings of the present study in terms of their relationship to this framework. Firstly, in terms of transfer-appropriate processing (TAP) between the session task (the TR and PR carousels) and the post-test, neither of the tasks fully meet this criteria. As Segalowitz explains:

> "L2 communication in the real world places various cognitive and perceptual processing demands on the L2 user. Fluent retrieval of earlier learned L2 knowledge and skills will depend in a significant way on the extent to which the cognitive and perceptual processes elicited at the time of communication match those that had been previously elicited at the time of learning" (Segalowitz, 2010, p. 65)

There was a clear rationale for using the particular pre- and post-test that were selected for the study. However, the cognitive demands of the task in the carousel task sequence (narrate a story using picture prompts) and the post-test (talk about a past event) were not perfectly matched from a TAP perspective.

In terms of how both tasks compare in how 'genuinely communicative' they were, I have shown that the PR carousel task elicited a genuine *need* to communicate. While the TR carousel was still 'a task' in a broader sense, it failed to meet Gatbonton and Segalowitz' (2005) criteria for a 'genuinely communicative' task, because there was no explicit need for the learners to enter into an interaction other than that they had been told to do so by the class teacher. This is an area for further investigation.

Both carousels were inherently *repetitive,* but in different ways. In the two case studies it seemed that they were repeating different things. The repetition in the TR group was much more focused and specific, meaning that learners could draw on already-primed language to reduce the processing load. It was argued that this meant learners could divert some of their attention that would otherwise be spent on producing the message elsewhere (i.e. on monitoring processes). The sort of repetition which characterised the PR carousel was much broader, more akin to the repetition described in Larsen-Freeman (2012) and DeKeyser (2017) and perhaps better dubbed 'practice'. It was argued that during this form of repetition, learners are not able to rely on primed language and instead must produce speech 'for the first time' each time. However, the tasks were selected such that learners in the PR group had the opportunity to repeat a particular structure (past simple tense) even if the lexical content was different each time. It was suggested that this may have had an effect on cognitive processing in terms of the proceduralisation of morpho-syntactic and phonological rules related to regular past simple verb usage and the memorisation of irregular past simple verb usage.

Although this study did not attempt to gauge the impact of the intervention on participants' *motivation*, willingness to learn and sense of L2 self (although I would argue that future research in this area would be welcome, see below), the post-hoc focus groups as well as the impression gained by the researcher and ESL teachers provide an exploratory base for discussing the extent to which the task sequences had any impact on motivation. It is true that a number of participants did find the TR carousel 'boring'. This follows other studies (e.g. Kim, 2013; Van de Guchte et al., 2015) which also found that TR was boring for some learners. Although, to the best of my knowledge, there has been no research which has looked at the impact that attitudes to the task might have on fluency, arguably a negative attitude towards the task may have impacted on performance at the post-task to some extent. Similarly, it was found that learners in the TR group did not necessarily *perceive* that they had made improvements in their speeches and, in particular, were often frustrated about the lack of teacher feedback which was given and which they believed would have allowed them to improve. Once again, it is possible that this lack of confidence in the task sequence may have impacted on motivation and levels of engagement which in turn led to a lack of fluency transfer.

In terms of initial levels of fluency, the TR carousel does seem to impact on cognitive fluency in the short-term. It is therefore possible that this might help to make the underlying processes of speech production more “self-sustaining” (Segalowitz, 2010, p. 177). In a dynamic systems approach to L2 fluency, it is also suggested that if fluency is increased, for example, through explicit instruction, it may precipitate a ‘virtuous circle’ or ‘upward spiral’ in terms of fluency development. This is because, as outlined by Segalowitz (2010; 2016) increased cognitive fluency may impact positively on sense of L2 self which in turn drives learners to seek out opportunities to interact in the TL leading to further fluency gains. Segalowitz (2010, p. 113) explains that

> It seems likely that there is a mutually supportive process involved here, where a certain threshold level of general cognitive and oral fluency is required before an L2 user will be able to attend to the sociolinguistic dimension of speaking and to learn from social encounters. Once achieved, however, this in turn opens up new possibilities for future and possibly even richer environments leading to even greater levels of fluency development

This means that language instruction provides an opportunity to ‘intervene’ in the dynamic system of fluency development, to offer learners the opportunity to increase cognitive fluency (even if only in the short-term) which may then contribute to longer-term development.

Segalowitz’ framework, then, provides a way of framing the findings of the current study with regards to fluency transfer and of offering some pathways for future investigations.

13.3 Methodological implications

The current study has a number of findings which may be of methodological value to fluency and task-based research. These will be addressed in three sections: (1) The challenges of carrying out L2 classroom-based research; (2) PRAAT annotation and coding; and (3) The use of mixed methods in TR research.

13.3.1 The challenges of carrying out L2 classroom-based research

The present study was carried out in an authentic L2 classroom, during regular scheduled classes and with intact classes and class teachers. Carrying out quantitative analysis in the classroom with intact classes presents the L2 researcher with a number of methodological challenges which need to be overcome (Foster, 1998; Dörnyei, 2007; Rossiter, 2001; Duff & Early, 1996). Explaining the particular challenges of conducting classroom research, Rossiter (2001, p. 36) writes:

> Faced with developing constraints related to non-equivalent groups, student and teacher participants, data collection, data analysis, task differences, and ethical considerations, the temptation for many classroom-oriented researchers in my position might be to curtail or even abandon their study. I maintain, however, that what are often perceived as problems by researchers are in fact the daily realities of the contexts in which most teachers practise. The limitations in these research settings may frustrate investigators… they are, however, part and parcel of the classroom context.

Rossiter (2001) discusses the particular challenges she faced while carrying out her doctoral research in a classroom environment. In particular she identifies challenges related to the use of intact classes, specific teachers, data collection, analysis and ethical considerations. In a similar vein, I will present here the challenges I faced while conducting the current study and present the ways in which I attempted to overcome them. Perhaps unsurprisingly, there is considerable overlap between Rossiter's experience and my own. Indeed, it would seem that some of these challenges are, indeed, part of the very fabric of the classroom environment.

Intact classes

Firstly, it is important to note the very particular challenge of working with intact classes of learners. Working with classes of this sort meant that there was often considerable variation between participants both within a class at an individual level and *between* classes at a group level. This is due to a number of realities involved with working with intact classes. The fluency levels of learners varied because learners were not put into classes based on their fluency level. Instead they were grouped together based on an in-house grammar test, written sample, and interview. This meant that even within a class, there were some who were deemed to be at a particular level because of their knowledge of grammar and others who were confident communicators. This had a big effect on variation in fluency levels which was even more pronounced between classes, as some classes seemed to be much more confident and fluent speakers than others, even classes at the same level of proficiency. Possible reasons for this are: (1) The impact that a particular teacher may have on learners' fluency; (2) The impact that point in the term may have; and (3) The absence or presence of 'loud' learners in the class.

In terms of the impact having a particular teacher may have, there was support for this theory in the student focus groups as one learner said that, in the normal Speaking and Listening class, they didn't usually have the opportunity to speak. Furthermore, teachers clearly differ in the emphasis they place on developing speaking skills, their attitudes to task-based approaches, the extent to which they record learners speaking. In terms of the latter, for example, I remember having a conversation with a class teacher prior to data collection and she said (I paraphrase): "Oh, you're going to record them speaking? That's fine! They are totally used to being recorded! We do it every week!"

The second point concerned the impact that being at a certain point in the 'term' may have had. In order to collect a sufficient amount of data to conduct effective statistical analysis, it was necessary to spread data collection over ten weeks. This was to carry out the research (which lasted a week for each class) with enough classes of learners. However, the data collection 'straddled' two 'terms'. This meant that data which was collected later in the study came from classes who were just beginning a term (i.e. they had just been moved up to intermediate level from pre-intermediate level) whereas the data that was collected earlier in the project came from classes who were in the middle or nearing the end of the term (and therefore level). I was aware of this potential source of influence at the time of data collection. However, the fact that classes were randomly assigned to conditions meant that, quite by chance, there were a higher number of ‘start-of-term’ classes in the PR group.

A final point in relation to fluency differences between classes is that some classes are just more talkative than others. This may be due to certain key members of the class that are naturally ‘loud’ or ‘extrovert’ and create a class atmosphere which is more chatty than others. It can also be related to the mix of linguistic and cultural backgrounds within a particular class. Although the school administrators took care to ensure a mix of L1 backgrounds within each class, it was still apparent that some classes were more keen to use the TL because of necessity whereas others were split into pockets of L1 friendship groups.

Continual enrolment

Another challenge was the school’s policy of ‘continual enrolment’. This policy meant that new learners could join a course at any point during a term (although, thankfully, newcomers were only added to classes on Mondays.) This meant that the design of the research had to be confined to a single week because class make-up changed from week to week with some learners leaving and new learners joining. Other learners would be ‘moved up’ to the next level if the teacher thought they were too strong for the class and others would be ‘moved down’ if they were struggling to keep up. This also meant that for some learners, they were thrust into this research project in the very first lesson on their very first day in a language school. For obvious reasons this was not ideal and generated a certain amount of guilt and anxiety for the researcher. In an attempt to ‘soften the blow’ for these new learners, I asked their morning class teacher to explain that I would be coming into the class later that day.

Scheduled lessons

As well as using intact classes, the current study took place during scheduled lessons (i.e. lessons that the learners were actually paying for) This presented a unique set of challenges and issues that needed to be overcome. Firstly, and most obviously, the research was restricted in terms of the available time for data collection. This meant that certain phases of the research had to be hurried. This undoubtedly had an impact on learners’ enjoyment and execution of the tasks involved in the research.

The fact that the research was taking place in a normal scheduled lesson meant that there was considerable pressure to 'sell' the research to the learners and teachers involved. I lived in fear of a student saying to me something like: "Wait a minute, why am I paying money to help you with your research?" It also meant that I needed to make arrangements with other teachers who were teaching classes simultaneously to 'babysit' any learners who did not want to take part in the research in their own classes. Luckily, this was not necessary as all the learners agreed to take part. I can't help wondering, however, whether any of the student absences were because they were not comfortable with the idea of the research project.

Speech recording in authentic classrooms

A particular challenge of this research was that multiple conversations needed to be recorded simultaneously, as this is one of the features of the carousel task sequence, and also one of the features of communicative classrooms more generally. This meant that there was at least two potential areas of difficulty. The first was that there would be considerable background noise during recordings which is dealt with in a separate section, below. The second was that I would need to find a way to record multiple conversations at once.

Issues with data recording meant that very explicit instructions had to be given to participants and both the researcher and class teacher were required to monitor how the recording devices were being used in order to ensure that they were held at an optimum height. It also meant that all recording devices needed to be 'locked' while they were recording to prevent learners accidentally or deliberately turning the recorders off during the tasks. It was also necessary to label each recorder and to make notes about which participant had which recorder. Detailed plans also needed to be made, particularly for the TR classes, in terms of how the learners were paired and what position they were in the classroom. This was to facilitate with ascribing the recordings on each Dictaphone with the person that had produced them. Unfortunately, I only had enough funds to acquire six voice recorders which meant there were only enough recorders for one per pair of learners, resulting in recordings from multiple participants on a single voice recorder.

The teachers

Contrary to that which is reported in Dörnyei (2007) and Rounds (1996), in general, the teachers involved in the study were entirely co-operative and enthusiastic about the research. I think that this was probably facilitated by two things. Firstly, I was a fellow teacher at the school and so the teachers may have been happier to help me than if I had been an 'outsider'. A second point is that I 'sold' the project to the teachers by explaining to them that it would essentially mean that they could observe me teaching their lesson for them. These teachers were all familiar with my research and the potential benefits of TR as I had already presented the results of some of my preliminary research at a teacher INSET session. I believe that these three factors may have meant that the teachers responded favourably to the research. There are clear implications for other novice researchers who would like to conduct research in L2 classrooms.

Data analysis

It is a natural consequence of working with learners in a normal classroom environment that they will interpret and perform the task in their own way, despite the best intentions of the researcher (Dörnyei & Kormos, 2000; Duff & Early, 1996). In the present study, thankfully, most learners stayed on task most of the time and it was not necessary to exclude data on the grounds that learners had not completed or subverted the task. However, learners did occasionally struggle to fill the allotted time with talk. This resulted in awkward silences at the ends of recordings as learners wondered what they should do next.

Some data was not able to be used because the 'visitor' had completely dominated the interaction, instead of asking questions to which the 'presenter' could respond, they attempted to divine the nature of the story based on the pictures that they could see:

> S34 (visitor): ah ok ok I understand I understand er this is a man and a baby and the baby fall fall down yes yes

Very occasionally, it was clear that anxiety, either in relation to the task itself or to the research can be observed to be having a direct effect on learners' performance:

> S77: a person I'm very nervous (ok don't worry) I understand but a little bit |

In the current study, given the relatively large number of participants, it is hoped that the effect of such factors are minimised overall.

General discussion

It is my hope that by highlighting the particular challenges I faced as a research whilst working on this project, that other researchers might find encouragement. I was certainly reassured to find that my experiences were not unique and that others (e.g. Rossiter, 2001; Duff & Early, 1996; Pica, 2005) had had similar experiences. I think it is very important indeed that we continue to discuss our difficulties as classroom researchers and that we seek ways to overcome or embrace the particular challenges that the L2 classroom provides. I also feel strongly that it was my own experience as a language teacher, who was familiar with the particular context of the language school and the school policies and procedures that allowed me to avoid a number of the pitfalls identified in, for example, Dörnyei (2007). There is therefore support for (a) providing training for teachers to conduct research in their own classes, (b) collaboration between researchers and language teachers, (c) more research for researchers who are also involved as teachers of the classes they are researching.

A final point is on the amount of detail given to methodological challenges of classroom-based research in published articles. It is relatively uncommon to find explanations of the difficulties faced by classroom researchers in research publications. This can be partly due to space but also, perhaps, due to a desire to present a polished impression of the research (Schachter & Gass, 1996). However, anyone that has worked as a researcher in a classroom knows that the process is not as clear-cut as it often appears in research articles. This could arguably be very off-putting for novice researchers and may dissuade them from going ahead with the research for fear that they must be getting something wrong (Rossiter, 2001; Duff & Early, 1996). I would like to say to those researchers that, yes, the classroom is a tricky environment for L2 data collection and , yes, it does pose unique problems and challenges, but, ultimately, the rewards of working in such a fluid and complex environment are considerable and provide a perspective of SLA which is messy, confusing, but also vital and dynamic.

13.3.2 Speech analysis, PRAAT scripts and noisy classrooms

As noted in Chapter 2, the precise temporal nature of measures of utterance fluency has precipitated the use of computer technology and specialist software such as PRAAT (Boersma & Weenink, 2008) which has been used to automatically detect silence in speech samples (N. de Jong & Perfetti, 2011; Cucchiarini et al., 2000; 2002) or used in conjunction with a specific computer script to identify syllable nuclei (de Jong & Wempe, 2009) thereby allowing for the automatic computation of speech rate (as in Boers, 2014 and Thai & Boers, 2016). Automatic fluency analysis makes dealing with larger amounts of data feasible, more objective and very precise. However, a drawback is that it requires relatively clear, straightforward speech data of the type that is elicited in a language laboratory and not in a working language classroom, where background noise and recording irregularities are likely to feature (Hilton, 2014). And yet, if fluency research is to be meaningful to both SLA researchers and language practitioners alike, it needs to be both scientifically and *ecologically* valid (DeKeyser, 2010). Some studies which have attempted to measure fluency in interactive scenarios (e.g. Lambert et al., 2017) have decided to use alternative measures of fluency because they were not able to measure silence automatically. This means, however, that their findings do not compare as readily with other studies that have used standardised measures of fluency (which include silent pauses, for example). In other words, there seems to be a trade-off necessary: if you want to conduct temporal fluency analysis in a classroom, you need to use different measures of fluency.

However, as set out in the opening chapter, it was important that this research had a dual perspective. I wanted it to be both scientifically and ecologically valid. This meant that I had to secure training for myself in how to use PRAAT effectively with noisy data and how to annotate speech samples quickly and precisely. I also needed to learn how to develop my own PRAAT programming scripts that would analyse my files in the way that I wanted. I developed two new PRAAT scripts designed to be used with my noisy interactive data (see appendix). These scripts have since been used on a range of projects (e.g. Tavakoli et al., 2016; Tavakoli et al., 2017) and have been shared with other fluency researchers in a series of workshops.

Not only did the manual annotation allow me to investigate *precisely* the features of speech that I found to be appropriate based on my reading of the literature, but also it allowed me to get really close to the data so that I was aware of certain patterns and behaviours and was able to link group data observations with individual-level explanations more easily than if I had simply run an automatic programme.

I would therefore argue for the benefits of manual annotation for utterance fluency analysis because it allows researchers to work with data that is not necessarily 'clean' but which comes from authentic classroom environments. It also means that reserachers can work with the particular fluency measures that they believe (based on their reading of the literature) really reflect cognitive and/or perceived fluency. Another point is that it allows the researcher to get a very good overview of the data and identify interesting patterns and irregularities. I feel that these benefits make the additional work involved worthwhile.

13.3.3 Mixing methods

The current study has hopefully shown that a mixed-methods approach provides an ideal platform from which to explore L2 task performance within a classroom context. I have shown how group and case study data can be used together to offer *complementarity* and a more complete picture of TR effects on fluency and I have shown how post-intervention focus groups have hopefully *expanded* the scope of the research. The combination of methods used in the current study meant that it was both deeper and wider than would otherwise have been the case.

13.4 Pedagogic implications

Based on the findings of this study, it can be said that there are benefits of using both types of carousel for language teaching purposes. On the one hand, the TR carousel allows learners to perform at higher levels of complexity, accuracy and fluency in the short-term. On the other hand, the PR carousel seems to provide opportunities for broader practice which may lead to fluency transfer. In this section, I attempt to make specific pedagogic recommendations based on the findings of this study. Firstly, I try to weigh up the benefits and drawbacks of each version of the carousel task sequence. Secondly, I make suggestions for ways in which TR and PR can be 'sold' to learners. Next I deal with a common concern of the learners in this study and therefore a challenge for language teachers: how to integrate feedback into the task sequences. I end this section for a look at different ways that the carousel task sequences can be enhanced and adapted to maximise learner and teacher engagement.

13.4.1 Task repetition or procedural repetition: which is most beneficial?

DeKeyser (2017, p. 27) advises teachers that "when planning activities, always think of whether they are meant to advance declarative knowledge, proceduralization, or automatization—is the aim to provide more understanding, getting to apply that understanding, or getting to use it faster, with less effort, more spontaneously?"

I have shown that the TR carousel increases cognitive processing speed (through conceptual planning and priming) and that this manifests itself in higher levels of utterance and perceived fluency. It can therefore be seen as a task sequence which targets automaticity or cognitive fluency but a very specific form of automaticity which is short-term and specific to the language required by the task. I have also shown that learners might use this temporary cognitive fluency to monitor their speech, noticing errors and opportunities for improvement in their output (and potentially in the input). Learners are then able to use this knowledge to make changes to their final performance whilst maintaining higher levels of fluency. The result is a third performance which is significantly more fluent, complex and accurate than the first. It therefore provides learners with an opportunity to integrate their language knowledge in active speech (Bygate & Samuda, 2005). However, there is no apparent fluency transfer effect following TR and, what is more, the learners seemed to find it, at times, boring and frustrating.

On the other hand, the PR carousel, which was a task sequence created specifically for this project, did not have any significant effect on fluency (automaticity) in the short term. It did, however, seem to be related to fluency transfer which I explained could be the result of the practice and subsequent proceduralisation and memorisation of past simple rules and forms. It can therefore been seen, in part, as a task sequence meant to advance and proceduralise declarative knowledge.

Table 65 outlines the key benefits and drawbacks of each version of the carousel.

Table 65: Benefits and drawbacks of both versions of the carousel task sequence

	Benefits	Drawbacks
Task repetition carousel	• Allows learners to perform at higher-than-usual levels of fluency • Allows learners to ameliorate performance in terms of accuracy and complexity • Provides learners with the opportunity to *monitor* their output (and interlocutor's input) more closely	• No *obvious* long-term effect on fluency • less stimulating for learners • Limited in terms of vocabulary
Procedural repetition carousel	• fluency increased on new task (transfer) • Involves productive and receptive skills • Enjoyable for learners	• Not conducive for increasing short-term accuracy/complexity • More complicated set-up for teacher

In terms of choosing between a TR or PR version of the task, then, I would suggest that teachers make a selection based on their knowledge of their class (proficiency level, attitude to tasks, age, context) as all of these things are potentially related to their acceptance of TR. In addition, teachers might like to think about the particular aims of their lesson. If the aim of a lesson is to *consolidate* or *integrate* language into active speech, then TR might be the best choice. On the other hand, if the aim of the teacher is to help learners to proceduralise grammatical or morphological rules, then the PR carousel in potentially more appropriate.

13.4.2 The hard sell

If task repetition is going to be used in the classroom, it is advisable to provide a clear rationale to learners and/or give feedback in order that learners have a clear understanding of the reasons for repeating a task and do not see it as an exercise in futility. In terms of explaining to learners the value in TR, teachers could take a number of different approaches. Firstly, as in N. de Jong and Perfetti (2011), TR could be presented as 'fluency training'. There, explicit explanations were provided on fluency and after each performance, learners had the chance to reflect on their performance in terms of its fluency. This reflection took the form of questions which popped up on the computer screen after learners had performed their speeches. This would obviously be more complicated to arrange using the carousel task sequence but, for example, learners could be asked to discuss how they felt about their fluency after each performance with their interlocutor.

Another option, and one which I feel particularly strongly about, is that teachers could make use of speech analysis software such as PRAAT to show learners how their language benefits from TR (or any other classroom task sequence). This could be achieved in a number of ways: (1) larger schools and universities could employ (or train) an in-house speech analyst who could provide speech analysis for the teachers of the school, (2) teachers themselves could be trained in how to analyse speech using software such as PRAAT or (3) a new 'user-friendly' computer program could be developed so that teachers could analyse learners' speech themselves. Hilton (2014) suggests that teachers could "have at their disposal a simple application for easily marking up sound files: hesitations could be selected, and tagged when occurring within a clause. An automatic count of numbers and lengths would give a clear, quantitative picture of the processing effort required by the task…[with] a much simpler user interface, for purposes of classroom assessment and action research" (Hilton, 2014, p. 45). Future research might take up the challenge presented by Hilton (2014) in order to empower teachers to analyse utterance fluency in their own classrooms and provide this information to their learners.

There are implications for this kind of approach to language assessment in the classroom beyond the benefits for the learners themselves. Firstly, Tavakoli and Hunter (2017) have suggested that the 'gap' between research and practice in L2 fluency can be partly explained by differing approaches to fluency analysis. For language teachers, the business of fluency assessment is a purely holistic and subjective one while for fluency researchers, it is related to the computation of temporal variables using specialised software. In order to have 'a common language' it might be very useful indeed to provide teachers with the skills to analyse language in the way that researchers do. This would have a knock-on effect on the validity of action research projects, teacher-researcher collaboration and so on. Although this would be a very different way of working with language in the classroom, I think there are potentially huge benefits. Perhaps even allowing learners to analyse the temporal features of their speech in quantitative and scientific ways could be a great leap towards the unification of L2 research and L2 teaching practice. It is an area that warrants further exploration in the coming years.

13.4.3 Integrating feedback

Teachers have a number of options when it comes to providing feedback during the TR and PR carousels. Firstly, teachers can circulate during the task performance, take notes and provide general feedback between first and second performances. Secondly, teachers can move around the carousel alongside a 'visitor' and provide feedback to individuals during their performances (the drawback of this is that some learners would get feedback sooner in the process than others). Thirdly, learners could listen to their recordings and transcribe them before repeating the task with additional interlocutors (Hsu, 2017; Lynch, 2007). The problem with this option, however, is that it would interrupt the natural rhythm of the task sequence and the learners who were 'visitors' would need to have another activity to do while the hosts transcribed their speeches. Finally, and perhaps most practically, learners themselves should be encouraged to give feedback. Partly, this can be facilitated by task design. In order to maximise the tasks' potential to induce learners to negotiate for meaning and engage in LREs, the task should have a clear communicative goal and be truly interactive. In other words, it may be necessary to tweak the task so that both learners who engage in the task have a really strong motivation to understand each other. In the current study, explicit instructions were given to learners to refrain from asking questions. This, it has been noted, may have limited the potential of the task sequence for fluency transfer. For example, as one of the teachers who took part in the study suggested, it could be that the learners who are 'visiting' the posters are 'journalists' who are writing up the 'witness accounts' for a newspaper. They would therefore be able to take notes and then they would be given the task (possibly for homework) of writing up the stories in a newspaper article about "unbelievable events".

Alternatively, or in addition, learners could actually be *trained* to provide better feedback. Lyster, Sato and Saito (2012) have shown how learners that were given training in how to provide corrective feedback to peers during interaction scored more highly on post-tests than those who had not been trained. There is clearly scope, then, for exploring the benefits of TR with learners who have been trained to give each other corrective feedback. An issue with peer feedback is that even though learners were receiving feedback through interaction with their peers, the attitudes of learners was that this was not useful for language learning. There were a number of references to the need for more "teacher feeling" and "professional" feedback. This is related to a wider issue in TBLT and that is how to convince learners of the usefulness of engaging in peer interaction when they would naturally prefer to have one-on-one feedback from their teacher.

13.4.4 Options and improvements

In the TR focus groups, it was suggested that there was a need for more "teacher feeling". Other learners from both groups said that they would have liked to interact with native speakers. Unfortunately, the reality of language classrooms are often such that teachers are stretched and learners are unable to be regularly partnered with native (or more proficient) speakers. However, a possible way of integrating an element of this into the carousel task sequences would be to provide learners with native speaker input in the first instance. Foster (2001) suggested that:

> to help prevent learners from committing inappropriate word sequences to memory, and/or to encourage them to build a more native-like memory store, it may be useful for teachers to enable learners to reflect upon their own language use and to compare it to native speaker norms… to give learners the chance to see what language choices native speakers make in specific contexts. A task-based classroom could be an ideal place for this. Learners who have done or who are about to do a language tasks could

listen to a tape or read a transcript of native speakers doing exactly the same task" (Foster, 2001, p. 90).

Along similar lines, in a relatively recent study, there were substantial benefits found when learners watched NSs doing a similar task to the one that they went on to perform (Arslanyilmaz and Pedersen, 2010). The way to integrate this sort of pre-task input would be to film native (or high proficiency) TL speakers telling the target story. If you had a class of 10 learners, you would need to have five different videos. Instead of providing the 'input material' to learners in the form of written accounts, then, it could be provided in oral (video) form. The carousel procedure would be the same. Learners work in pairs and share a computer. They watch the NS videos which were set up on computers (or tablets/mobile phones) around the room. When the video has been watched, the pairs are split into 'As' and 'Bs'. The 'As' stay with the video while the Bs move around the room listening to the As stories. When they have listened to all the stories, they return to their partner and the As and Bs swap roles.

The findings of the focus groups in the present study revealed that learners did not necessarily want to take part in carousels on a regular basis. Obviously, the learners in the current study took part in relatively similar carousels over the course of the week. In a real classroom, the carousels could take very different forms. Foster and Hunter (2016), for example, suggested a number of ways that the carousel can be adapted.

Lambert, Philp et al. (2016) found that learners were more engaged when they were talking about storyboards that they had created themselves based on their own experiences. This is certainly something which could be integrated into the carousel task sequence. For example, following the exploration of a first-person account of an incredible life experience like the ones featured in this study, learners could work on creating storyboards which represent their own incredible life experiences.

13.5 Generalisability of findings

As with all classroom-based research, there are certain limitations in terms of the generalisability of findings. On one hand, it can be said that research which is carried out in a classroom is more relevant to teaching than research which is carried out in a decontextualised laboratory because the environments, tasks, interaction patterns are very different. However, the flip-side of this is that the research is highly contextualised, meaning that findings may or may not be generalisable to other classes in other contexts and with other groups of learners etc. I take the view, however, that with well-conducted SLA research, whether conducted in a classroom or laboratory, certain observations can be said to have wider application.

An advantage of this study is that it has used a multilingual group of learners. At the very least, this suggests that findings of this study are not due to learners being from a particular L1 background (much of the TR research has dealt with speakers of a particular language). Of course, this may mean that the findings are *only* applicable to multilingual groups, in the sense that the *interactional patterns* and TL use are such that fluency is fostered to a greater extent in this sort of environment. An additional point here is that the context is an ESL one in which learners were studying in a TL-speaking host country. Following on from the discussion of a dynamic systems approach to fluency development outlined in Chapter 2, it is possible to assert that any long-term effects of the intervention may have been impacted by additional opportunities that learners may have had to interact with native speakers of the TL. In other words, it may be an *interplay* between TR/PR and opportunities to interact in the TL that ultimately led to the results of the present study.

The learners in this study were B1-B2 level and, even though Lambert et al. (2017) did not find any significant interaction between TR and proficiency level, I have argued that initial levels of fluency may, indeed, impact on the ways learners use the repetition which is inherent in the carousel. I think it is likely that looking only at *overall, mean* or *group* figures can be misleading when it comes to establishing the role that proficiency plays in TR benefits. These figures can mask enormous variation between individual participants.

It is also important to bear in mind that this research has dealt with a very specific type of repetition (immediate, ecological) and therefore it cannot be said that repetition defined differently would have the same effects. Indeed, I have argued that it is the specific nature of the carousel task sequence that may bring about certain changes to performance.

That being said, the findings of this study, coming as they do from an authentic classroom, with intact classes should provide ample evidence for the benefits of these sort of task sequences in another ESL context. This is supported by the fact that fluency gains are based not only on utterance fluency but also on *teacher judgements,* and the fact that data has not only been subjected to quantitative statistical analysis but also qualitative analysis.

Perhaps most importantly, this study has involved teachers and has put their opinions and feelings about the tasks at the very heart of the study. For this reason, it is my belief that the task sequences described in this study are not only beneficial for fluency (albeit in different ways) but they also "work" with real classes and with real teachers. For example, I know that following on from this research a number of teachers that I had worked with continued to use the carousel with their classes, adapting it and adding things as they went. This is the hallmark of a successful pedagogic task sequence, in my opinion, if teachers can imagine ways that they can put their own spin on the task sequence in order to serve particular pedagogic purposes and to fit with their particular teaching style. Ultimately, however, the findings of the focus groups show that carousels might be best administered *in moderation* alongside other tasks with different conditions.

13.6 Limitations

While care has been taken to maximise the reliability and of this study, it is necessary to acknowledge a number of limitations. Firstly, although the study is one of the largest of its kind in the field of TR, it is still quite small. However, as I and others (Witton-Davies, 2014; Hilton, 2014) have pointed out, manual fluency analysis of classroom-elicited speech data is intensive and time-consuming. Therefore, these larger studies may need to be carried out in more controlled environments which may make them more amenable to automated fluency analysis.

The pre- and post-test were selected carefully and were considered appropriate for assessing 'transfer' in the current study. However, analysis has revealed that they may have been too different in terms of cognitive processing demands from the demands of the session tasks. For this reason, the choice of pre- and post-test may be seen as a limitation of the current study. Future research might consider pre- and post-tests which are much closer to the session tasks in terms of demands, possibly alongside another task which is more different. This would allow for a comparison between transfer to a new task of a similar type and transfer to a new task of a different type. It would also have been interesting to have seen how participants fared on a *delayed* post-test (as in Kim & Tracy-Ventura, 2013; N. de Jong & Perfetti, 2011). In particular, because it has been suggested that the effects of *massed* practice of the sort which was provided in the present study may have an effect on *immediate* post-tests but not on delayed post-tests (DeKeyser, 2017). However, this was not an option in the present study because the make-up of classes changed each week. Participant attrition would therefore have been an issue (as in Rossiter, 2001).

Although the research questions in the current study were not concerned with long-term accuracy and complexity development, clearly the analysis of both these aspects of performance in the pre- and post-test data is an interesting next step. In a similar vein, the data from the remaining two sessions (Session 2 and Session 3) is currently unanalysed. As explained, for reasons of space, these were not analysed as part of the current study and was not necessary in order to test the hypotheses on the present study. Nonetheless, it would be interesting to see what changes there are to learners' fluency over the course of the week and to add further flesh to theories provided here, in particular to explore the relationship between the PR carousel and past tense usage.

An additional limitation from an empirical point of view, is the fact that the three groups of participants in the study were not perfectly homogenous. Certain between-groups findings (e.g. that the PR group made greater improvements between pre- and post-test) may have been influenced by the different CAF profiles of the groups. An advantage of the study, however, is that there is also a within-subjects component and did not rely simply on cross-sectional analysis between groups at a particular time-point (e.g. post-test). The latter is an approach taken in a number of TR studies (e.g Bygate, 2001; Fukuta, 2016) and is not ideal for observing change brought about *over time* through TR. Unfortunately, differences between groups in this study was part and parcel of the reality of working with intact groups of learners over a particular period of time.

Although there was a significant amount of informal and 'behind-the-scenes' teacher input into the current study (see Chapter 5), it would have been ideal to obtain the teachers' views of the task sequence (as in Richards, 2011) on a more formal basis. The next phase of this research would ideally be to present the research findings to groups of ESL teachers, elicit their feedback and encourage them to experiment with the task and procedural repetition carousels. I would then envisage a feedback stage whereby the teachers would be able to explain what worked and what didn't work, how they adapted the task sequence and so on.

A final limitation regards the case study analysis (Chapter 11). Only two cases were selected for this thesis for reasons of space. It would have been optimal to analyse a larger number of cases to provide further support or contradiction for the claims I have made in relation to the underlying *reasons* for fluency gains during TR and PR.

13.7 Future research

As is usual in ISLA research, and particularly in studies which include a qualitative element (Riazi & Candlin, 2014), as well as answering research questions, the study has generated new ones. In this section, I will highlight a number of potential directions for future research in the area of TR and L2 fluency.

Firstly, the suggestion that complexity and accuracy gains during TR may be linked to monitoring processes warrants further investigation. For example, it would be very interesting indeed to carry out stimulated recall interviews with participants who have been recorded during the carousel in order to gain a deeper insight into how TR impacts on monitoring and, ultimately, performance

The study's finding that procedural repetition may have a fluency transfer effect also demands further investigation. Given the fact that it is *teachers* who may need convincing about the usefulness of TR as a teaching tool (Ahmadian et al., 2017), and this study's finding that immediate PR might have particular benefits for L2 learning, there is clearly scope for further investigation of this form of repetition, the extent to which it impacts on learning, learners' attitudes to it and teachers' willingness to use it in a classroom. I would suggest that this sort of research provides an ideal opportunity for L2 researchers and language teachers to collaborate and design studies, the findings of which will have relevance for L2 research and language pedagogy.

Drawing on Segalowitz' (2010) account of ISLA and fluency transfer, it is possible to imagine a number of research projects which investigate the relationship among factors that are purported to impact on language learning. For example, and following Larsen-Freeman (2009) it would be interesting to look at long-term effects that TR might have using alternative indices than CAF. For example, there is support here for the theory that learner *engagement* and *motivation* may have a part to play in the ultimate success of TR and PR as the focus groups in the current study hinted at there being a relationship between eagerness to learn and willingness to repeat. It would therefore be very interesting (and relatively straightforward) to gauge the extent to which TR and PR impacts on learners' motivation to interact with native speakers, for example, and also the extent to which motivation impacts on the way in which learners engage with task repetition. This sort of research could build on the findings of MacIntyre (2007) and MacIntyre et al. (1998) by using the Willingness to Communicate (WTC) framework to interpret results.

Another key question which this research has raised is the connection between transfer-appropriate processing (TAP), TR/PR and fluency transfer. It was argued that in the current study and elsewhere (e.g. Kim & Tracy-Ventura, 2013; Patanasorn, 2010) that post-test tasks have been too dissimilar from treatment tasks to promote transfer. It would be interesting to see a study which operationalised TAP and compared conditions in which post-tests were more or less likely to capture TAP, for example a similar and completely different task type (as in Bygate, 2001).

Although, it was not discussed at length here, further research is needed which explores the impact of initial fluency levels on fluency gains through TR. This is something which also features in Segalowitz' (2010) framework. He suggests, however, that higher cognitive fluency might 'beget' even more fluency development. This dynamic relationship between current levels of performance and the impact that may have on fluency development is a crucial one in terms of understanding the interplay of factors involved in second language acquisition.

Along methodological lines, future research might also consider classroom *observation* as a means for triangulating findings and understanding exactly what it is that learners do when they repeat the task. Research questions could focus on, for example, non-verbal interaction, body language and the role of the teacher. Similarly, future research might focus on the role of teacher cognition in TR use, building on Tavakoli and Hunter's (2017) findings in relation to defining fluency and promoting fluency in the classroom.

As explained in the pedagogic implications section above, I suggested that learners could be trained to provide feedback during peer interaction (Lyster, Sato & Saito, 2014). As well as being a pedagogic recommendation, this could be interpreted as a research question: *what is the impact of peer feedback on CAF gains during task repetition?* The same is true for a number of the pedagogic recommendations made. Again, these present ideal opportunities to collaborate with language teachers on research.

13.8 Chapter summary

I have discussed the theoretical, methodological and pedagogic implications of this research. In terms of support for theoretical models, I explained that the current study emphasises the impact of *priming* in immediate task repetition and explained that one of the effects of priming is increased cognitive fluency on the same task. I explained that there is also support for theories of monitoring in SLA in the sense that learners in the TR group may have been diverting attention towards monitoring during TR. I explained that *proceduralisation* of morpho-syntactic rules is a possible explanation for the findings of the current study. In terms of a framework for thinking about fluency transfer from ISLA, I suggested that Segalowitz' (2010) framework was a useful point of departure. I also argued that the findings of the present study could add to this framework in interesting ways (i.e. by specifying the impact that procedural repetition might have).

Next, I presented the methodological implications. I believe that this study, as it was undertaken by a single researcher (me) has shown that certain 'difficulties' that arise when considering mixed-methods approaches can be overcome. Similarly, I have shown that the classroom environment, while it does present considerable challenges to the mixed-methods researcher can become a place that is conducive to this sort of analysis. I explained that classroom researchers need to be flexible, responsive and patient in order to reap the rewards that come with working in a simultaneously ecologically and scientifically valid way.

In terms of pedagogic implications, I drew on the study's findings and my own experience of working with the carousel task sequences as a teacher to outline some recommendations for language teaching. I suggested that both task and procedural repetition both have much to offer in language teaching but that procedural repetition may hold the key to longer-term fluency development. Given that this is a new finding, I stressed that further research is needed and that this research might be best carried out by teachers in their own classes. I explained different ways of 'selling' task repetition to learners including the possibility of using speech analysis software to show learners how their output changes over time. I explained a range of ways of integrating different forms of feedback into the carousel task sequence, which emerged as a major concern of learners in the present study. Finally, I looked at a number of options for adapting the TR and PR carousels for different teaching purposes.

I explained that, while the study was carried out in a specific context with specific groups of learners, a strength of the research is that it was carried out in a 'normal' ESL class. It was therefore suggested that the findings of this study may be generalisable to other ESL classrooms. As with all classroom-based research, this study has revealed a number of areas for future research. Chief among these, is the need to further explore immediate PR and its impact on SLA as this may lead to great developments in the way in which 'practice' is provided in ISLA.

13.9 Conclusion

At the very start of this thesis, I explained that I was motivated by two concerns: (1) understanding the relationship between TR and L2 fluency and (2) finding useful ways to work with TR in the language classroom and uncovering factors that might be salient to the implementation of TR with language learners.

My reading of the literature in these two intersecting areas led me to investigate the impact of different *types* of repetition on L2 fluency (i.e. TR versus PR) and involved looking at fluency on a number of different *levels* (cognitive, utterance, perceived) and also at different *aspects* of fluency (i.e. speed, breakdown, repair). The case studies which were investigated in this study provided a glimpse at fluency in even further detail, which allowed the slippery concept of automaticity to be explored in greater depth.

The nature of the relationship between TR and L2 fluency, as proposed in this study, is that immediate exact task repetition (TR) allows learners to plan their utterance and also draw on primed language. It would seem that the resultant increased speed of processing allows learners to dedicate more attention to monitoring during the subsequent performances which further increases other aspects of fluency (i.e. mid-clause pausing) as well as other aspects of performance (i.e. accuracy and complexity).

Procedural repetition, on the other hand, seems to have more in common with a much broader definition of repetition, one which is closer to 'practice' (DeKeyser, 2010; 2017; Larsen-Freeman, 2012). This is because PR does *not* allow learners to draw on primed language in the short-term although it does provide ample practice for the exploration of language *rules* and possibly the memorisation of salient irregular verb forms. The current study has shown that this breadth of input and experience within a particular mode of discourse may bring about changes to underlying representation (i.e. proceduralisation) to such an extent that hesitation related to a particular phenomena are greatly reduced.

In terms of how TR can be operationalised in the classroom, the current study has suggested that both TR and PR may have something to offer the language teacher. While TR may be more suited for integrating or consolidating existing knowledge, PR may be better for providing learners with broader practice within a language classroom. I have also explored how both TR and PR can be supplied to learners through a pedagogically convincing task sequence, using Lynch and Maclean's 'Poster Carousel' as the basis for the TR version and my own adapted version to provide immediate PR. The focus groups revealed that learners *did* seem more reluctant to repeat the exact same task than they did to engage in procedural repetition. The implications of these findings were discussed in the previous chapters and include the need to integrate feedback in some form and to clearly explain the intended aims of TR. It was suggested that certain alterations could be made to the TR carousel which take on board some of the learners' concerns.

Attending to both of these aims in a single research project meant employing a combination of research methodology and carrying out precise fluency measurement in an authentic language classroom. Both these decisions meant that the design, data collection and analysis were more challenging than might otherwise have been the case. This study has shown that mixed-methods approaches can go some way to facilitating research into complex systems such as fluency and in complex environments such as the language classroom and I have also shown that highly scientific tools of speech data analysis can be compatible with data which has been collected in a language classroom.

More research needs to be conducted to continue to investigate the intricate interplay among factors involved in TR, PR and L2 fluency in the context of language teaching. Ultimately, I hope to have conducted a study into task repetition which joins this important discussion, which has both scientific and ecological validity and that may be of interest to those who are researching L2 fluency and those who are busy trying to find a way to *teach* L2 fluency. For both these groups of people, fluency surely remains a slippery and, at times, frustrating construct to work with, but it is also at the very heart of what it means to know another language and I, for one, have found it an exciting and important concept to explore.

References

Ahmadian, M. J., Masouri, A., Ghominejad, S. (2017) Language learners' and teachers' perceptions of task repetition *Language Teaching Research*

Ahmadian, M. J. (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. J. (2012). Task repetition in ELT. *ELT journal*, *66*(3), 380-382.

Ahmadian, M. J., & 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.

Amiryousefi, M. (2016). The differential effects of two types of task repetition on the complexity, accuracy, and fluency in computer-mediated L2 written production: a focus on computer anxiety. *Computer Assisted Language Learning*, *29*(5), 1050-1066.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Andon, N., & Eckerth, J. (2009). Chacun à son gout? Task-based L2 pedagogy from the teacher's point of view. *International Journal of Applied Linguistics*, *19*(3), 286-310.

Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC journal*, *22*(1), 84-94.

Azkarai, A., & García Mayo, M. D. P. (2016). Task repetition effects on L1 use in EFL child task-based interaction. *Language Teaching Research*, 1362168816654169.

Azkarai, A., & Oliver, R. (2016). Negative feedback on task repetition: ESL vs. EFL child settings. *The Language Learning Journal*, 1-12.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Bailey, K. M., & Savage, L. (1994). *New Ways in Teaching Speaking. New Ways in TESOL Series: Innovative Classroom Techniques*. TESOL, 1600 Cameron Street, Suite 300, Alexandria, VA 22314.

Baker-Smemoe, W., Dewey, D. P., Bown, J., & Martinsen, R. A. (2014). Does measuring L2 utterance fluency equal measuring overall L2 proficiency? Evidence from five languages. *Foreign Language Annals*, *47*(4), 707-728.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 657.

Boers, F. (2014). A Reappraisal of the 4/3/2 Activity. *RELC Journal*, *45*(3), 221-235.

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, *32*, 83-110.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language teaching research*, *10*(3), 245-261.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 5.3.04). Retrieved from http://www.praat.org/

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*(2), 159-175.

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of memory and language*, *75*, 104-116.

Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks. *ETS Research Report Series*, *2005*(1).

Brown, R. (1973). *A first language: The early stages*. Harvard U. Press.

Brumfit, C. J., & Johnson, K. (1979). The linguistic background. *The communicative approach to language teaching*, 1-45.

Brumfit, C., & Bennett, T. J. A. (1979). Accuracy and fluency as polarities in foreign language teaching materials and methodology. *Bulletin CILA (Commission interuniversitaire suisse de linguistique appliquée)(«Bulletin VALS-ASLA» depuis 1994)*, *29*, 89-99.

Bui, G., Ahmadian, M. J., & Hunter, A. (in prep) The effects of 'spacing' of repeated L2 task performance

Bui, G., & Huang, Z. (2016). L2 fluency as influenced by content familiarity and planning: Performance, measurement, and pedagogy. *Language Teaching Research*, 1362168816656650.

Bybee, J. (2008). *Usage-based grammar and second language acquisition*.

Bygate, M. (1988). Units of oral expression and language learning in small group interaction. *Applied linguistics*, *9*(1), 59-82.

Bygate, M. (1996). Effects of task repetition: appraising the developing language of learners. In D. Willis & J. Willis (Eds.), *Challenge and change in language teaching* (pp. 134-146). London: Heinemann.

Bygate, M. (1999). Quality of language and purpose of task: Patterns of learners' language on two oral communication tasks. *Language teaching research*, *3*(3), 185-214.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate et al.., (Eds.), *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 23-48). Harlow: Longman.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (Vol. 11, pp. 37–74). Amsterdam: Benjamins.

Bygate, M., Swain, M., & Skehan, P. (2001). *Researching pedagogic tasks: Second language learning, teaching, and testing*. Routledge.

Cambridge ESOL. (n.d.). IELTS speaking test: public band descriptors. Retrieved May 19th, 2017, from https://www.teachers.cambridgeesol.org/ts/exams/academicandprofessional/ielts/speaking

Campione, E., & Véronis, J. (2005). Pauses and hesitations in French spontaneous speech. In *Disfluency in Spontaneous Speech*.

Chafe, W.L. (1980). Some reasons for hesitating In H.-W. Dechert & M. Raupach (Eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 271-286). The Hague,The Netherlands: Mouton.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 535-544.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in second language acquisition*, *11*(04), 367-383.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *The Journal of the Acoustical Society of America*, *107*(2), 989-999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *The Journal of the Acoustical Society of America*, *111*(6), 2862-2873.

De Bot, K. (1992). Applied linguistics. *Applied linguistics*, *13*(1), 1-24.

De Jong N (2012) Does time pressure help or hinder oral fluency? In: De Jong N, Juffermans K, Keijzer M, and Rasier L (eds) *Papers of the Anéla 2012 applied linguistics conference*. Delft: Eburon, 43–52.

De Jong, N.H. (2016) Predicting pauses in L1 and L2 speech *IRAL* (aop)

De Jong, J. H., & Van Ginkel, L. W. (1992). 15 Dimensions in Oral Foreign Language Proficiency. *The construct of language proficiency: Applications of psychological models to language assessment*, 187.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods*, *41*(2), 385-390.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2013). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 1–21.

De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(02), 223-243.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012a).Facets of Speaking Proficiency. *Studies in Second Language Acquisition*, *34*(01), 5–34.

De Jong , N. H. , Steinel , M. P. , Florijn , A. F. , Schoonen , R. , & Hulstijn , J. H . (2012b). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and nonnative speakers . In A. Housen , F. Kuiken , & I. Vedder (Eds.), *Dimensions of L2 performance and profi ciency investigating complexity, accuracy and fluency in SLA*. Amsterdam : Benjamins .

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 1–24.

De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*(2), 533-568.

DeKeyser, R., M. (2017) Knowledge and Skill in ISLA. In Loewen, S., & Sato, M. (Eds.). *The Routledge Handbook of Instructed Second Language Acquisition*. Routledge.

DeKeyser, R. (2007a). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.

DeKeyser, R. (2010). Practice for second language learning: Don't throw out the baby with the bathwater. *International Journal of English Studies*, *10*(1), 155-165.

DeKeyser, R. M. (2007b). Study abroad as foreign language practice. *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*, 208-226.

Derwing, T., M. (2017) L2 fluency development in Loewen, S., & Sato, M. (Eds.). (2017). *The Routledge Handbook of Instructed Second Language Acquisition*. Routledge.

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163-185.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359-380.

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(04), 533-557.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language learning*, *54*(4), 655-679.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, *34*(2), 183-193.

Deschamps, A. (1980). The syntactical distribution of pauses in English spoken as a foreign language by French students. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 255-266). The Hague, Netherlands: Mouton.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford University Press.

Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication. *Studies in second language acquisition*, *20*(3), 349-385.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language teaching research*, *4*(3), 275-300.

Dörnyei, Z., & Scott, M. L. (1997). Communication strategies in a second language: Definitions and taxonomies. *Language learning*, *47*(1), 173-210.

Doughty, C., & Pica, T. (1986). "Information gap" tasks: do they facilitate second language acquisition?. *TESOL quarterly*, *20*(2), 305-325.

Ejzenberg, R. (2000). The juggling act of fluency. In H. Riggenbach (Ed.), *Perspectives on Fluency* (pp. 243–265). Michigan: University of Michigan Press.

Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in second language acquisition*, *9*(01), 1-19.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, amp042.

Ellis, R. (Ed.). (2005). *Planning and task performance in a second language* (Vol. 11). John Benjamins Publishing.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford University Press, USA.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in second Language acquisition*, *26*(01), 59-84.

Essig, W. (2005). Storytelling: Effects of planning, repetition and context. In *Teachers exploring tasks in English language teaching* (pp. 201-213). Palgrave Macmillan UK.

Evans, M. A. (1985). Self-initiated speech repairs: A reflection of communicative monitoring in young children. *Developmental Psychology*, *21*(2), 365.

Fillmore, C. (1979). On Fluency. In C. Fillmore, charles, D. Kempler, & W. Wang (Eds.), *Individual differences in language ability and language behavior* (pp.85–101). New York: Academic Press.

Foster, P. (1998). A classroom perspective on the negotiation of meaning. *Applied linguistics*, *19*(1), 1-23.

Foster, P. (2009). Task-based language learning research: expecting too much or too little?. *International Journal of Applied Linguistics*, *19*(3), 247-263.

Foster, P., & Hunter, A. (2016). When it's not what you do but the way that you do it. *SLA Research and Materials Development for Language Learning*, 280.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second language acquisition*, *18*(03), 299-323.

Foster, P., & Skehan, P. (2013). Anticipating a post-task activity: The effects on accuracy, complexity, and fluency of second language performance. *Canadian modern language review*, *69*(3), 249-273.

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language learning*, *59*(4), 866-896.

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.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp.123-148). Amsterdam: Benjamins.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in second language acquisition*, *26*(02), 275-301.

Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research*, *20*(3), 321-340.

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing*, *13*(2), 208-238.

García Mayo, M. D. P., & Imaz Agirre, A. (2016). Task repetition and its impact on EFL children's negotiation of meaning strategies and pair dynamics: An exploratory study. *The Language Learning Journal*, *44*(4), 451-466.

García-Fuentes, C., & McDonough, K. (2016). The effect of explicit instruction and task repetition on Colombian EFL students' use of politeness strategies during disagreements. *The Language Learning Journal*, 1-13.

Gass, S. M., & Mackey, A. (2013). *The Routledge handbook of second language acquisition*. Routledge.

Gass, S., Mackey, A., & Ross-Feldman, L. (2011). Task-Based Interactions in Classroom and Laboratory Settings. *Language learning*, *61*(s1), 189-220.

Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, *49*(4), 549-581.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL quarterly*, *22*(3), 473-492.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*(3), 325-353.

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*(3), 379-399.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. Academic Press Inc.

Guillot, M.-N. (1999). *Fluency and Its Teaching*. Clevedon, UK: Multilingual Matters.

Guz, E. (2016). Refining the Methodology for Investigating the Relationship Between Fluency and the Use of Formulaic Language in Learner Speech. *Research in Language*, *14*(2), 95-122.

Hakuta, K. (1975). Learning to speak a second language: What exactly does the child learn. *Developmental psycholinguistics: Theory and applications*, 193-207.

Hansen, L., Gardner, J., & Pollard, J. (1998). The measurement of fluency in a second language: Evidence from the acquisition and attrition of Japanese. *On JALT*, *97*, 37-45.

Harmer, J. (1991). The practice of English language teaching. *London/New York*.

Hawkes, M. L. (2012). Using task repetition to direct learner attention and focus on form. *ELT journal*, *66*(3), 327-336.

Hedge, T. (2001). *Teaching and learning in the language classroom* (Vol. 106). Oxford,, UK: Oxford University Press.

Higgs, T. V., & Clifford, R. (1982). The push towards communication. In T. Higgs (Ed.), Curriculum, competence, and the foreign language teacher (pp. 57–79). Lincolnwood, IL: National Textbook Company.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, *36*(2), 153-166.

Hilton, H. (2014). Oral fluency and spoken proficiency: Considerations for research and testing. *Measuring L2 proficiency: Perspectives from SLA*, 27-53.

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.

Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* (pp. 121–142). Amsterdam ; Philadelphia: John Benjamins.

Hsu, H. C. (2017). The combined effect of task repetition and post-task transcribing on L2 speaking complexity, accuracy, and fluency. *The Language Learning Journal*, 1-16.

Huensch, A., & Tracy–Ventura, N. (2017). L2 Utterance Fluency Development Before, During, and After Residence Abroad: A Multidimensional Investigation. *The Modern Language Journal*, *101*(2), 275-293.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in second language acquisition*, *27*(02), 129-140.

Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.

Hunt, K. W. (1970). Do sentences in the second language grow like those in the first?. *TESOL Quarterly*, 195-202.

Indrarathne, B. (2016). *Attentional processing of different types of written L2 input and its relationship with learners' working memory capacity* (Doctoral dissertation, Lancaster University).

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct?. *Applied linguistics*, *29*(1), 24-49.

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.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809-854.

Kim, Y. (2009). The effects of task complexity on learner–learner interaction. *System*, *37*(2), 254-268.

Kim, Y. (2013). Promoting attention to form through task repetition in. *Second language interaction in diverse educational contexts*, *34*, 3.

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction?. *System*, *41*(3), 829-840.

Kinginger, C. (2011). Enhancing language learning in study abroad. *Annual review of applied linguistics*, *31*, 58-73.

Kinginger, C. (2013). Identity and language learning in study abroad. *Foreign Language Annals*, *46*(3), 339-358.

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. (2003). Attention and monitoring in a second language. A qualitative analysis. *Fremdsprachen Lehren und Lernen*, *32*, 116-132.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, N.J.: Lawrence Erlbaum.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145-164.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and cognition*, *11*(02), 261-271.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of memory and language*, *33*(2), 149.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. *Handbook of bilingualism: Psycholinguistic approaches*, 531-553.

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*, amu047.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167-196.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied linguistics*, *18*(2), 141-165.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. Oxford University.

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*, *30*(4), 579-589.

Larsen-Freeman, D. (2012). On the roles of repetition in language teaching and learning. *Applied Linguistics Review*, *3*(2), 195-210.

Larsen-Freeman, D. E. (1976). An explanation for the morpheme acquisition order of second language learners. *Language learning*, *26*(1), 125-134.

Lennon, P. (1990a). Investigating fluency in EFL: A quantitative approach. *Language learning*, *40*(3), 387-417.

Lennon, P. (1990b). The advanced learner at large in the L2 community: Developments in spoken performance. *IRAL-International Review of Applied Linguistics in Language Teaching*, *28*(4), 309-324.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H.Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Ann Arbor: University of Michigan Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W. J. M. (1999). Producing spoken language: a blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *Neurocognition of language*. (pp. 83-122). Oxford, England: Oxford University Press.

Lewis, M., & Gough, C. (1997). *Implementing the lexical approach: Putting theory into practice* (Vol. 3, No. 1, pp. 223-232). Hove: Language teaching publications.

Lightbown, P. M. (2000). Anniversary article. Classroom SLA research and second language teaching. *Applied linguistics*, *21*(4), 431-462.

Lightbown, P. M. (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.

Llanes, A., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, *37*(3), 353-365.

Llanes, À., & Muñoz, C. (2013). Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning*, *63*(1), 63-90.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological review*, *95*(4), 492.

Long, M. (2014). *Second language acquisition and task-based language teaching*. John Wiley & Sons.

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. *Modelling and assessing second language acquisition*, *18*, 77-99.

Lynch, T. (2001). Seeing what they meant: Transcribing as a route to noticing. *ELT journal*, *55*(2), 124-132.

Lynch, T. (2007). Learning from the transcripts of an oral communication task. *ELT journal*, *61*(4), 311-320.

Lynch, T., & Maclean, J. (1994). The Poster Carousel. In Bailey, K. M., & Savage, L. (Eds), *New Ways in Teaching Speaking. New Ways in TESOL Series: Innovative Classroom Techniques*. Alexandria, VA.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*(3), 221-250.

Lynch, T., & Maclean, J. (2001). Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 99–118). Harlow: Longman.

Lynch, T., & Maclean, J. (2003). Effects of Feedback on Performance: A Study of Advanced Learners on an ESP Speaking Course. *Edinburgh Working Papers in Applied Linguistics*, *12*, 19-44.

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, *46*(1), 1.

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., & Gass, S. M. (2015). *Second language research: Methodology and design*. Routledge.

Mackey, A., & Silver, R. E. (2005). Interactional tasks and English L2 learning by immigrant children in Singapore. *System*, *33*(2), 239-260.

Mackey, A., Abbuhl, R., & Gass, S. M. (2012). Interactionist approach. *The Routledge handbook of second language acquisition*, 7-23.

Maurice, K. (1983) The fluency workshop. *TESOL Newsletter* 17(4): 29.

Maurice, K. (1994) The Fluency workshop. in Bailey, K. M., & Savage, L. (Eds.). *New Ways in Teaching Speaking. New Ways in TESOL Series: Innovative Classroom Techniques*. Alexandria, VA.

McCarthy, M. (2010). Spoken fluency revisited. *English Profile Journal*, *1*(01), e4.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in second language acquisition*, *20*(01), 83-108.

Meunier, F. (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, *32*, 111-129.

Michel, M. (2017) Complexity, Accuracy and Fluency in Loewen, S., & Sato, M. (Eds.). (2017). *The Routledge Handbook of Instructed Second Language Acquisition*. Routledge.

Mora, J. C. (2006). Age effects on oral fluency development. *Age and the rate of foreign language learning*, *19*, 65-88.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *Tesol Quarterly*, *46*(4), 610-641.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, *16*(5), 519-533.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in second language acquisition*, *23*(04), 451-468.

Nassaji, H. (2012). The relationship between SLA research and language pedagogy: Teachers' perspectives. *Language Teaching Research*, *16*(3), 337-365.

Nation, P. (1989). Improving speaking fluency. *System*, *17*(3), 377-384.

Németh, N., & Kormos, J. (2001). Pragmatic aspects of task-performance: The case of argumentation. *Language Teaching Research*, *5*(3), 213-240.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555-578.

Norris, J. M., & Ortega, L. (2006). *Synthesizing research on language learning and teaching* (Vol. 13). John Benjamins Publishing.

Norris, J. M., & Ortega, L. (2003). Defining and measuring SLA. *The handbook of second language acquisition*, 716-761.

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.

Nunan, D. (1991). Methods in second language classroom-oriented research. *Studies in second language acquisition*, *13*(02), 249-274.

Nunan, D. (2005). Classroom research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 225-240). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Connor, N. (1988). Repairs as indicative of interlanguage variation and change. *Georgetown University round table in languages and linguistics*, 251-259.

Ohta, A. S. (2000). Rethinking interaction in SLA: Developmentally appropriate assistance in the zone of proximal development and the acquisition of L2 grammar. *Sociocultural theory and second language learning*, *4*, 51-78.

Olynyk, M., d'Anglejan, A., & Sankoff, D. (1987). A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. *Applied psycholinguistics*, *8*(2), 121-136.

Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International journal of social research methodology*, *8*(5), 375-387.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in second language acquisition*, *21*(01), 109-148.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590-601.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117-134.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism* (Vol. 18). John Benjamins Publishing.

Patanasorn, C. (2010). *Effects of procedural content and task repetition on accuracy and fluency in an EFL context*. Northern Arizona University.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, *191*, 225.

Payant, C., & Reagan, D. (2016). Manipulating task implementation variables with incipient Spanish language learners: A classroom-based study. *Language Teaching Research*, 1362168816669742.

Pica, T. (1987). Second-language acquisition, social interaction, and the classroom. *Applied linguistics*, *8*(1), 3-21.

Pica, T., Kanagy, R., & Falodun, J. (1993). Choosing and using communication tasks for second language instruction. *Multilingual Matters*, 9-9.

Pinget, A. F., Bosker, H. R., Quené, H., & de Jong, N. H. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349-365.

Pinter, A. (2005). Task repetition with 10-year-old children. In *Teachers exploring tasks in English language teaching* (pp. 113-126). Palgrave Macmillan UK.

Pinter, A. (2007). Some benefits of peer–peer interaction: 10-year-old children practising with a communication task. *Language Teaching Research*, *11*(2), 189-207.

Plonsky, L. (2013). Study quality in SLA. *Studies in Second Language Acquisition*, *35*(04), 655-687.

Plonsky, L., & Gass, S. (2011). Quantitative research methods, study quality, and outcomes: The case of interaction research. *Language Learning*, *61*(2), 325-366.

Plough, I., & Gass, S. M. (1993). Interlocutor and task familiarity: Effects on interactional structure. *Multilingual Matters*, 35-35.

Prabhu, N. S. (1987). *Second language pedagogy* (Vol. 24). Oxford: Oxford University Press.

Préfontaine, Y. (2013). Perceptions of French fluency in second language speech production. *Canadian modern language review*, *69*(3), 324-348.

Préfontaine, Y., & Kormos, J. (2016). A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 151-169.

Riazantseva, A. (2001). Second language proficiency and pausing a study of Russian speakers of English. *Studies in Second Language Acquisition*, *23*(04), 497-526.

Riazi, A. M., & Candlin, C. N. (2014). Mixed-methods research in language teaching and learning: Opportunities, issues and challenges. *Language Teaching*, *47*(02), 135-173.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse processes*, *14*(4), 423-441.

Roberts, B., & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes*, *15*(2), 129-157.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, *22*(1), 27-57.

Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (Vol. 2). John Benjamins Publishing.

Rossiter, M. (2003). " It's Like Chicken but Bigger": Effects of Communication Strategy in the ESL Classroom. *Canadian Modern Language Review*, *60*(2), 105-121.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, *65*(3), 395-412.

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *Canadian Modern Language Review*, *66*(4), 583-606.

Sadri Mirdamadi, F., & De Jong, N. H. (2015). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research*, *31*(1), 105-116.

Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. *Psycholinguistic models of production*, 45-65.

Sample, E., & Michel, M. (2014). An Exploratory Study into Trade-Off Effects of Complexity, Accuracy, and Fluency on Young Learners' Oral Task Repetition. *TESL Canada Journal*, *31*, 23-46.

Samuda, V. (2001). Guiding relationships between form and meaning during task performance: The role of the teacher. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 119–140). Harlow: Longman.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. New York: Palgrave Macmillan.

Sandelowski, M. (2003). Tables or tableaux? The challenges of writing and reading mixed methods studies. *Handbook of mixed methods in social and behavioral research*, 321-350.

Schachter, J., & Gass, S. M. (2013). *Second language classroom research: Issues and opportunities*. Routledge.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in second language acquisition*, *14*(4), 357-385.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. *Attention and awareness in foreign language learning*, *9*, 1-63.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. 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. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*(03), 369-385.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, *26*(02), 173-199.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in second language acquisition*, *26*(02), 173-199.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. *Handbook of bilingualism: Psycholinguistic approaches*, 371-388.

Sheppard, C. (2006). The Effects of Instruction Directed at the Gaps Second Language Learners Noticed in their Oral Production. *Unpublished PhD Thesis*, University of Auckland.

Shintani, N. (2012). Repeating input-based tasks with young beginner learners. *RELC Journal*, *43*(1), 39-51.

Shintani, N. (2014). Using tasks with young beginner learners: The role of the teacher. *Innovation in Language Learning and Teaching*, *8*(3), 279-294.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38-62.

Skehan, P. (1998). Task-based instruction. *Annual review of applied linguistics*, *18*, 268-286.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Skehan, P. (2001). Tasks and language performance. *Researching pedagogic tasks: Second language learning, teaching, and testing*, 167-186.

Skehan, P. (2003). Task-based instruction. *Language teaching*, *36*(01), 1-14.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied linguistics*, *30*(4), 510-532.

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. (2007). Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing research. *Complexity, accuracy, and fluency in second language use, learning, and teaching*, 207-226.

Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*, *32*, 199.

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 97-111.

Spada, N., & Lightbown, P. M. (1999). Instruction, first language influence, and developmental readiness in second language acquisition. *The Modern Language Journal*, *83*(1), 1-22.

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), Handbook on research in second language teaching and learning. (pp. 471-483) Mahwah, NJ: Erlbaum.

Swain, M., & Lapkin, S. (2000). Task-based second language learning: The uses of the first language. *Language teaching research*, *4*(3), 251-274.

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: Second language learning, teaching and testing (pp. 99–118). Harlow, UK: Pearson Education.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of pragmatics*, *30*(4), 485-496.

Takimoto, M. (2012). Assessing the effects of identical task repetition and task-type repetition on learners' recognition and production of second language request downgraders. *Intercultural Pragmatics*, *9*(1), 71-96.

Tavakoli, P. (2015). Connecting research and practice in TESOL: A community of practice perspective. *RELC Journal*, *46*(1), 37-52.

Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System*, *37*(3), 482-495.

Tavakoli, P. (2011). Pausing patterns: differences between L2 learners and native speakers. *ELT journal*, *65*(1), 71-79.

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 133-150.

Tavakoli, P., & Foster, P. (2011). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *61*(s1), 37-72.

Tavakoli, P., & Hunter, A. M. (2017). Is fluency being 'neglected' in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*.

Tavakoli, P., & Skehan, S. (2005). Strategic planning, task structure, and performance testing in R. Ellis (ed.) *Planning and Task Performance in a Second Language.* John Benjamins

Tavakoli, P., Campbell, C., & McCormack, J. (2015). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*.

Tavakoli, P., Nakatsuhara, F., & Hunter, A. (in print). Scoring validity of the Aptis Speaking Test: Investigating fluency across tasks and levels of proficiency. *Aptis Final Report.* London: British Council.

Thai, C., & Boers, F. (2016). Repeating a Monologue under Increasing Time Pressure: Effects on Fluency, Complexity, and Accuracy. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, *50*(2), 369-393.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84-119.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition*, *28*(01), 1-30.

Valls–Ferrer, M., & Mora, J. C. (2014). L2 fluency development in formal instruction and study abroad: The role of initial fluency level and language contacts. In C. Pe´rez–Vidal (Ed.), Language acquisition in study abroad and formal instruction contexts (pp. 111–136). Philadelphia/ Amsterdam: John Benjamins.

van de Guchte, M., Braaksma, M., Rijlaarsdam, G., & Bimmel, P. (2016). Focus on form through task repetition in TBLT. *Language Teaching Research*, *20*(3), 300-320.

Van den Branden, K. (2016). Task-based language teaching. In Hall, G. (Ed.). (pp. 238) *The Routledge handbook of English language teaching*. Routledge.

Verhoeven, L. T. (1989). Monitoring in children's second language speech. *Second Language Research*, *5*(2), 141-155.

Wang, Z. (2014). Developing accuracy and fluency in spoken English of Chinese EFL learners. *English language teaching*, *7*(2), 110.

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.

Wennerstrom, A. (2000). The role of intonation in second language fluency. In M.Koponen & H. Riggenbach (Eds.), *Perspectives on Fluency* (pp. 102–127). University of Michigan: Ann Arbor.

Widdowson, H. G. (1978). *Teaching language as communication*. Oxford University Press.

Willis, D., & Willis, J. (2008). *Doing task-based teaching*. Oxford University Press.

Willis, J. (1996). A flexible framework for task-based learning. In Willis, J., & Willis, D. (Eds.), *Challenge and change in language teaching* (pp. 52-62). Macmillan Education Australia.

Witton-Davies, G. (2014). The study of fluency and its development in monologue and dialogue. *Unpublished doctoral dissertation, Lancaster University, Lancaster, UK*.

Wood, D. (2001). In search of fluency: What is it and how can we teach it?. *Canadian Modern Language Review*, *57*(4), 573-589.

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review*, *63*(1), 13-33.

Wood, D. (2007). Mastering the English formula: Fluency development of Japanese learners in a study abroad context. *JALT JOURNAL*, *29*(2), 209.

Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *The Canadian Journal of Applied Linguistics*, *12*(1), 39.

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Bloomsbury Publishing.

Wray, A. (2002). Formulaic language in computer-supported communication: theory meets reality. *Language Awareness*, *11*(2), 114-131.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231-254.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-Line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied linguistics*, *24*(1), 1-27.

Yule, G., & Macdonald, D. (1990). Resolving referential conflicts in L2 interaction: The effect of proficiency and interactive role. *Language learning*, *40*(4), 539-556.

Appendices

Appendix 1 - Different stories involved in the carousel

Story number	Story title
1	I saved a baby who fell from a window
2	I was crushed by a cow
3	My cat saved my life
4	I was trapped in a ravine for eight days
5	I was impaled while pregnant
6	I was swallowed by a hippo
7	I saved a toddler trapped on a roof

Appendix 2 - Table showing numbers of students in each class

Class	**Group**	**Total number of students involved in the session**	**Number of students eligible for study***	**Number of different stories****
1	TR	12	10	6
2	TR	12	8	6
3	TR	10	6	5
4	PR	10	8	5
5	PR	10	3	5
6	PR	8	4	4
7	PR	12	7	6
8	C	10	10	5
9	C	10	8	5

*eligibility was related to being present for *both* the training session *and* the pre- and post-tests

**stories were included in the task sequence in the order that they appear in Appendix 1

Appendix 3 – IELTS performance descriptors

IELTS™

SPEAKING: Band Descriptors (public version)

Band	Fluency and coherence	Lexical resource	Grammatical range and accuracy	Pronunciation
9	• speaks fluently with only rare repetition or self-correction; • any hesitation is content-related rather than to find words or grammar • speaks coherently with fully appropriate cohesive features • develops topics fully and appropriately	• uses vocabulary with full flexibility and precision in all topics • uses idiomatic language naturally and accurately	• uses a full range of structures naturally and appropriately • produces consistently accurate structures apart from 'slips' characteristic of native speaker speech	• uses a full range of pronunciation features with precision and subtlety • sustains flexible use of features throughout • is effortless to understand
8	• speaks fluently with only occasional repetition or self-correction; hesitation is usually content-related and only rarely to search for language • develops topics coherently and appropriately	• uses a wide vocabulary resource readily and flexibly to convey precise meaning • uses less common and idiomatic vocabulary skilfully, with occasional inaccuracies • uses paraphrase effectively as required	• uses a wide range of structures flexibly • produces a majority of error-free sentences with only very occasional inappropriacies or basic/non-systematic errors	• uses a wide range of pronunciation features • sustains flexible use of features, with only occasional lapses • is easy to understand throughout; L1 accent has minimal effect on intelligibility
7	• speaks at length without noticeable effort or loss of coherence • may demonstrate language-related hesitation at times, or some repetition and/or self-correction • uses a range of connectives and discourse markers with some flexibility	• uses vocabulary resource flexibly to discuss a variety of topics • uses some less common and idiomatic vocabulary and shows some awareness of style and collocation, with some inappropriate choices • uses paraphrase effectively	• uses a range of complex structures with some flexibility • frequently produces error-free sentences, though some grammatical mistakes persist	• shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8
6	• is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation • uses a range of connectives and discourse markers but not always appropriately	• has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inappropriacies • generally paraphrases successfully	• uses a mix of simple and complex structures, but with limited flexibility • may make frequent mistakes with complex structures though these rarely cause comprehension problems	• uses a range of pronunciation features with mixed control • shows some effective use of features but this is not sustained • can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times
5	• usually maintains flow of speech but uses repetition, self correction and/or slow speech to keep going • may over-use certain connectives and discourse markers • produces simple speech fluently, but more complex communication causes fluency problems	• manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility • attempts to use paraphrase but with mixed success	• produces basic sentence forms with reasonable accuracy • uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems	• shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6
4	• cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction • links basic sentences but with repetitious use of simple connectives and some breakdowns in coherence	• is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice • rarely attempts paraphrase	• produces basic sentence forms and some correct simple sentences but subordinate structures are rare • errors are frequent and may lead to misunderstanding	• uses a limited range of pronunciation features • attempts to control features but lapses are frequent • mispronunciations are frequent and cause some difficulty for the listener
3	• speaks with long pauses • has limited ability to link simple sentences • gives only simple responses and is frequently unable to convey basic message	• uses simple vocabulary to convey personal information • has insufficient vocabulary for less familiar topics	• attempts basic sentence forms but with limited success, or relies on apparently memorised utterances • makes numerous errors except in memorised expressions	• shows some of the features of Band 2 and some, but not all, of the positive features of Band 4
2	• pauses lengthily before most words • little communication possible	• only produces isolated words or memorised utterances	• cannot produce basic sentence forms	• Speech is often unintelligible
1	• no communication possible • no rateable language			
0	• does not attend			

Appendix 4 – Spearman's correlations for all fluency measures

Correlations

			rater1	rater2	Meanrater	speechrate	MLR	AR	PTR	Freqmid	freqrepair	Accuracy	MeanLengthAS	Subordination	MeanLengthClause
Spearman's rho	rater1	Correlation Coefficient	1.000	.797**	.941**	.749**	.692**	.513**	.445**	-.442**	-.281**	.422**	.148	.227**	-.109
		Sig. (2-tailed)	.	.000	.000	.000	.000	.000	.000	.000	.001	.000	.083	.007	.205
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	rater2	Correlation Coefficient	.797**	1.000	.952**	.820**	.759**	.571**	.497**	-.432**	-.355**	.506**	.265**	.310**	-.035
		Sig. (2-tailed)	.000	.	.000	.000	.000	.000	.000	.000	.000	.000	.002	.000	.682
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	Meanrater	Correlation Coefficient	.941**	.952**	1.000	.830**	.765**	.571**	.497**	-.457**	-.333**	.483**	.215*	.278**	-.075
		Sig. (2-tailed)	.000	.000	.	.000	.000	.000	.000	.000	.000	.000	.011	.001	.383
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	speechrate	Correlation Coefficient	.749**	.820**	.830**	1.000	.888**	.772**	.503**	-.492**	-.316**	.407**	.238**	.214*	.028
		Sig. (2-tailed)	.000	.000	.000	.	.000	.000	.000	.000	.000	.000	.005	.012	.746
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	MLR	Correlation Coefficient	.692**	.759**	.765**	.888**	1.000	.569**	.555**	-.599**	-.322**	.489**	.254**	.182*	.088
		Sig. (2-tailed)	.000	.000	.000	.000	.	.000	.000	.000	.000	.000	.003	.033	.303
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	AR	Correlation Coefficient	.513**	.571**	.571**	.772**	.569**	1.000	.147	-.282**	-.063	.314**	.281**	.248**	.034
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.	.085	.001	.463	.000	.001	.003	.690
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	PTR	Correlation Coefficient	.445**	.497**	.497**	.503**	.555**	.147	1.000	-.270**	-.316**	.221**	.198*	.117	.108
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.085	.	.001	.000	.009	.020	.170	.206
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	Freqmid	Correlation Coefficient	-.442**	-.432**	-.457**	-.492**	-.599**	-.282**	-.270**	1.000	.067	-.347**	.014	.098	-.079
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.001	.001	.	.437	.000	.869	.255	.357
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	freqrepair	Correlation Coefficient	-.281**	-.355**	-.333**	-.316**	-.322**	-.063	-.316**	.067	1.000	-.236**	-.242**	-.316**	.042
		Sig. (2-tailed)	.001	.000	.000	.000	.000	.463	.000	.437	.	.005	.004	.000	.624
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	Accuracy	Correlation Coefficient	.422**	.506**	.483**	.407**	.489**	.314**	.221**	-.347**	-.236**	1.000	.191*	.221**	-.024
		Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.009	.000	.005	.	.025	.009	.779
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	MeanLengthAS	Correlation Coefficient	.148	.265**	.215*	.238**	.254**	.281**	.198*	.014	-.242**	.191*	1.000	.600**	.526**
		Sig. (2-tailed)	.083	.002	.011	.005	.003	.001	.020	.869	.004	.025	.	.000	.000
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	Subordination	Correlation Coefficient	.227**	.310**	.278**	.214*	.182*	.248**	.117	.098	-.316**	.221**	.600**	1.000	-.295**
		Sig. (2-tailed)	.007	.000	.001	.012	.033	.003	.170	.255	.000	.009	.000	.	.000
		N	138	138	138	138	138	138	138	138	138	138	138	138	138
	MeanLengthClause	Correlation Coefficient	-.109	-.035	-.075	.028	.088	.034	.108	-.079	.042	-.024	.526**	-.295**	1.000
		Sig. (2-tailed)	.205	.682	.383	.746	.303	.690	.206	.357	.624	.779	.000	.000	.
		N	138	138	138	138	138	138	138	138	138	138	138	138	138

**. Correlation is significant at the 0.01 level (2-tailed).

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

Appendix 5 – Session task materials

Experience: I saved a baby who fell from a window

On a mild October evening, I was walking down a busy London street in search of a cashpoint, pausing briefly to look at a menu in a restaurant window.

As I turned from the window, I felt a rush of air, and there was a sudden movement in my peripheral vision. Then was a noise as something hit the pavement to my right. I looked down. At first I couldn't comprehend what was there. A pile of rags? No, wait, a doll. To my horror, I realised it was a baby. I looked up and saw an open sash window.

My initial reaction was that she must be dead – she was so still and her eyes were closed. I looked around to see if anyone was going to do something. No one came. Then I realised that I had to do something.

I looked at the tiny child dressed in a pale pink babygrow. I looked in my bag for my phone and rang 999 while shouting to the now-gathering crowd to search the many nearby restaurants for a doctor or a nurse. I lay down next to the baby.

I had done a first-aid course a few months earlier and was grateful for it as I gently checked for a pulse. It was there, and a finger above her mouth confirmed she was breathing.

I clapped my hands in front of her face, calling, "Sweetheart, look at me. Open your eyes, darling, come on." She cried and then her eyes opened to look at me.

With my hand resting on her upper back, I just talked and talked to her, keeping my voice smooth and mellow. I knew she had to keep her eyes open so she wouldn't lapse back into unconsciousness. I wanted to keep her as still as I could to avoid aggravating any spinal injuries. One passerby shouted to pick her up, but I knew this was wrong.

After several minutes, the police arrived, followed by the ambulance. A police officer ran up to the apartment and emerged, minutes later, with the baby's parents, who were stunned, then distraught.I stepped back into a doorway and watched as the baby was strapped on to a rigid board and taken into the ambulance. In the next few days, the news reported the accident and mentioned a "passerby", which made me feel strange. Not particularly heroic, just glad that I had known what to do.

Experience: I saved a toddler trapped on a roof

I'd arranged to meet my wife and a couple of friends in the beer garden of a local hotel. As I approached, I could hear raised voices. As I walked into the garden, I saw everyone looking up, including my wife and friends. Following their gaze, I saw a child in a red hat perched on top of a roof jutting out from the hotel.

I realised the people in the garden had arranged themselves underneath, in the hope that one of them would be able to catch him if he fell. From the other side of the roof, I heard a woman's voice imploring him to sit down and keep still. It was the boy's mother, who had tried to climb up to him and become stranded halfway along a flat roof.

The child ignored his mother's voice and his every movement brought a fresh gasp of anxiety from onlookers. The fire brigade had been called, but no one knew how long they'd take. I scanned the side of the building, I quickly worked out the quickest way up and moved towards the drainpipe.

The drainpipe was old and looked as if it might break away from the wall, but I pulled myself up, grabbed on to a gutter that cracked as it took my weight, then quickly heaved myself on to the tiles. Above me, the child paid me little attention and didn't appear to be frightened at all. He was younger than I'd thought – no older than two – and the roof was steep. The situation suddenly seemed even more grave.

My only option now was to keep climbing. I launched myself upwards and tried to ignore the tiles fracturing beneath me, a couple slipping free underfoot and clattering down.

I planted myself as firmly as I could on the roof, grabbed the boy and swung him on to my lap. In my precarious position, just holding on to him was a challenge – he wouldn't stop wriggling.

It took only about 15 minutes for the fire brigade to arrive. A long ladder slid towards us, and a fireman climbed up, warning me not to move until he was directly below me. I passed the boy over, then waited to take my own turn climbing down, glad the situation was now in someone else's hands.

The boy's mum thanked me over and over and called me an "angel", but I don't think that term really fits. I'd acted instinctively rather than heroically.

Experience: my cat saved my life

When I first got my cat Charley, she was a tiny kitten, no more than five weeks old.

I have been diabetic since I was 12 and need two injections of insulin a day. I can usually control my diabetes well, but once or twice, usually due to illness or exhaustion, I have ended up having hypoglycaemia, when your blood sugar drops to dangerously low levels, causing dizziness, palpitations and even loss of consciousness. Normally during a hypo, you can recover by eating something sweet, but sometimes the attack is so sudden, you don't get the chance.

I had been Christmas shopping all day and came home exhausted. I had some food and went to bed, but I obviously didn't eat enough, because that night I can remember getting up to go to the toilet… and then nothing. I collapsed in the bathroom.

Within minutes, the cat realised something was wrong. Instead of just sitting by my side, she went to our bedroom and jumped on the bed where Kevin, always a heavy sleeper, lay undisturbed. Then she began to pat his hand with her paw and lick his face, something she'd never done before.

We've no idea how long she continued this, but it must have taken a while for her to rouse him. He woke up because she was licking him, but he shooed her off the bed and fell back to sleep. Charley didn't give up, though, and kept on patting and licking, all the while making a strange squeaking noise that my husband had never heard her make before.

After batting her away another couple of times, he finally sat up. Charley instantly shot off the bed and out of the door. Kevin noticed that my side of the bed was empty, and followed the cat to the bathroom, where he found me unconscious on the floor. Thankfully, he had been trained in how to give me an injection of glucagon, which makes the body release glucose, and within minutes I came round. "What am I doing here?" I asked groggily.

As I recovered with a drink and some toast, Charley sat on my lap, purring. She slept by my side that night, and the next day hovered around me. My other two cats slept through the whole thing – much more typical cat behaviour. Now I am more careful with my blood sugar levels – it has taught me an important lesson. It shook Kevin that I could become so ill so quickly – he hadn't had to inject me before. We were both so grateful to the cat.

Experience: I was trapped in a ravine for eight days

It was 20 September 2007 and I had just finished the night shift at a local supermarket near my home in Maple Valley, Washington.

After leaving the supermarket, I climbed into my blue Honda four-wheel drive and pulled out on to the motorway. But then I veered off the road and plummeted into a 20ft-deep ravine.

I was hanging from my seat at a strange angle, jammed hard against the steering wheel, with my seatbelt cutting into my chest. I tried to move, and cried out as pain shot through me. Looking down, I could see there was something wrong with my shoulder. My left arm was hanging at a weird angle and I couldn't move my fingers. I could feel the sickening crunch of broken ribs. My left leg was wedged tightly between the seat and dashboard, and had gone completely numb.

I felt confused, frightened and overwhelmingly tired.

Day faded into night. At some point I became aware of a blue light glowing in front of me. I realised it was my mobile phone. I stretched as far as I could, but my body was pinned tight by the seatbelt and the steering wheel, and it remained just out of reach.

I was slipping in and out of consciousness, but could tell from the cycles of light that several days had passed. Then I heard a noise. There were faces outside the window. I assumed it was a hallucination until I heard someone shout, "She's alive!"

I was cut out of the car and put into a medically induced coma while doctors catalogued my injuries. My kidneys were failing, I had a dislocated left shoulder, fractured ribs and vertebra, and my left clavicle had been snapped in two.

I haven't been back to the spot since the accident. Even now, four years later, I don't remember what happened or why I crashed, and I hope I never do. By some miracle I survived eight days at the bottom of a ravine, with terrible injuries and no food or water. I don't want to ask any questions.

Experience: I was impaled while pregnant

Long past midnight one Sunday in 2002, after a tiring weekend working as a charity collector, I climbed into my friend's car for the long drive back home to Bournemouth. I was eight months pregnant and exhausted. Even before the car had reached the dark country roads outside the city, I had slipped into a deep, untroubled sleep.

The accident happened when a fox ran into the road. As my friend swerved to avoid it, he lost control of the car, which skidded off the road and down a slope into woodland. I remember the windscreen exploding as the car crashed through a fence, showering me with broken glass and splintered wood.

As it broke through the fence, one of the wooden posts plunged deep into my chest. Another splinter pierced my left arm, and a piece lodged in the side of my neck, narrowly missing a main artery. I ran my fingers through my hair and clumps of it came away, along with shards of glass. Broken glass was sticking out of my face, too. My friend tried to help me out of the car, but every time he attempted to undo my seatbelt, I'd cry out in pain.

I passed out before the emergency services arrived, and they spent an hour trying to free me. A helicopter airlifted me to hospital. I remember looking down and seeing the fence post jutting out of me, like a spear, and screaming, "What the hell's that?" before blacking out again. The helicopter crew feared they were going to lose me at that point.

It was several hours before I surfaced again, this time in the hospital. I had tubes and clips attached everywhere, but the wood was no longer in my chest. My first thought was for my baby. He was fine, I was told, but the wood had travelled a full six inches into my body, passing right though my breast and into my stomach, missing the uterus by centimetres.

I was advised to stay in hospital until the baby was born, so the hospital could monitor both of us, but I was keen to get home and after a few days I was released.

In the end, I went into labour a few days late. It wasn't an easy birth – the impact of the accident had caused the baby to move, so we were spine to spine, which made my contractions very painful.

I called my little boy Kai – it means "safe harbour" or "survivor" in German.

Experience: I was swallowed by a hippo

I was 27 and I'd been working as a guide on kayaks on this stretch of the Zambezi river for years. That day I'd taken clients out with three apprentice guides – Mike, Ben and Evans – all in kayaks. We were near the end of the tour. The solid whack I felt behind me took me by surprise.

I turned just in time to see Evans, who had been flung out of his boat, flying through the air. His boat, with his two clients still in it, had been lifted half out of the water on the back of a huge bull hippo.

There was a cluster of rocks nearby and I yelled at the nearest apprentice to guide everyone there, to safety. Then I turned my boat and paddled furiously towards Evans.

I reached over to grab his outstretched hand but as our fingers were about to touch, I was engulfed in darkness. There was no transition at all, no sense of approaching danger. It was as if I had suddenly gone blind and deaf.

My legs were surrounded by water, but my top half was almost dry. I seemed to be trapped in something slimy. There was a terrible, sulphurous smell, like rotten eggs, and a tremendous pressure against my chest. It was only then that I realised I was underwater, trapped in the hippo's mouth. I wriggled as hard as I could, and in the few seconds for which he opened his jaws, I managed to escape. I swam towards Evans, but the hippo struck again, dragging me back under the surface.

Then, the hippo lurched suddenly for the surface, spitting me out as it rose. Mike was still waiting for me in his kayak and managed to paddle me to safety. I was a mess. My left arm was crushed to a pulp, blood poured from the wounds in my chest.

Luckily, he knew first aid and was able to seal the wounds in my chest which almost certainly stopped my lungs from collapsing and saved my life. By chance, a medical team was nearby, on an emergency drill, and with their help I stayed alive long enough to reach a hospital with a surgeon. He warned me he would probably have to take off both my arms and the bottom of my injured leg. In the end, I lost only my left arm – they managed to patch up the rest. Evans' body was found down river two days later.

Experience: I was crushed by a cow

Three years ago, I was moving the herd of 300 cows out of a field and one cow, who hadn't stayed with the others, crept up silently behind me. I waved my hands at her and shouted, "Get back!" I wasn't concerned, just slightly irritated that she'd followed me. But as I turned to walk back to the yard, the cow suddenly ran at me and hit me in my back, which threw me face-down on the concrete track.

Shocked and winded, I rolled over and tried to get up, but the cow lowered her head and pushed it into my chest and stomach, crushing me into the ground. I felt the back legs of the 1,000lb beast folding, and she sat on me.

She was making the most terrible bellowing, bawling sound – I think she genuinely wanted to kill me. Desperate and gasping for air, I took the only action I could – I pressed a thumb into each of her eyes and twisted as hard as I could.

Looking back, it seems cruel, but it was a purely instinctive action. At first I thought it wasn't going to work. The massive black-and-white head continued to bear down on me even as I screwed my thumbs in deeper. "This is it," I thought. "I just can't get out of this one."

Then, quite suddenly, the pressure was released. The cow wandered away, shaking her head – I wondered at first if I'd blinded her. I tried to stand, but my back was in searing agony and my legs wouldn't work. Gasping with pain, I managed to retrieve my mobile phone.

My wife ran to me and we waited for the ambulance, but when it arrived the crew were kept at bay by the furious cow. A farmhand drove over in a Land Rover and the cow went for him as well. None of us had ever seen anything like it.

Eventually, he managed to drive her away and the paramedics assessed me. Fearing my back might be seriously injured, they called for an air ambulance. I doubt it took more than half an hour, but the wait seemed to last for ever – the pain was unbearable.

Eventually I was given morphine, which left me too spaced out to take in much of what followed. At the hospital, I was assured I'd suffered no permanent damage, though I was told if I'd been a smaller-framed guy, I wouldn't have stood a chance – the weight of that great head bearing down on my chest would simply have crushed my ribcage. After a fortnight, I returned to work. The cow had been destroyed.

Appendix 6 – Letter for participants

School of Communication, Culture and Creative Arts

ST MARY'S UNIVERSITY COLLEGE

Waldegrave Road

Strawberry Hill

Twickenham, TW1 4SX

020 8240 4000

Title of Study: Fluency Training in the Second Language Classroom: The Carousel Technique

Name of researcher: Ann-Marie Richards

Email: ann-marie.richards@smuc.ac.uk

The study will form part of my PhD research in the School of Communication, Culture, and Creative Arts, St Mary's University College, Twickenham.

The aim of the research is to investigate the effectiveness of a particular type of fluency training for people who are learning English as a second language. The type of training I am investigating is a speaking activity called 'The Carousel Technique'.

If you agree to participate, you will be involved in up to 4 fluency training sessions with the researcher, a British Study Centres teacher and other students of English. These sessions will take place in your normal 'Speaking and Listening' classes at 12.20pm, Monday - Friday. Parts of the activity will be recorded using a voice recorder. You may also be asked to complete a short questionnaire about your experience.

A code will be attached to your data so it remains totally anonymous and your data will be kept safe and confidential. You will not be identifiable in the write-up of this research or any publication which might follow.

You are free to stop your involvement in the project and withdraw at any time.

If you would like to receive a summary of the results of this research, please write your email address on the attached consent form.

Thank you,

Ann-Marie Richards

Appendix 7 – Consent form

Consent form

Title of Study: Fluency Training in the second language classroom: The Carousel Technique.

Name of researcher: Ann-Marie Richards

I have been informed about the nature of this study and willingly consent to take part in it.

I understand that I will be recorded and that the content of the recordings and questionnaires will be kept confidential.

I understand that I may withdraw from the study at any time.

I am over 16 years of age.

Name ____

Signed ____

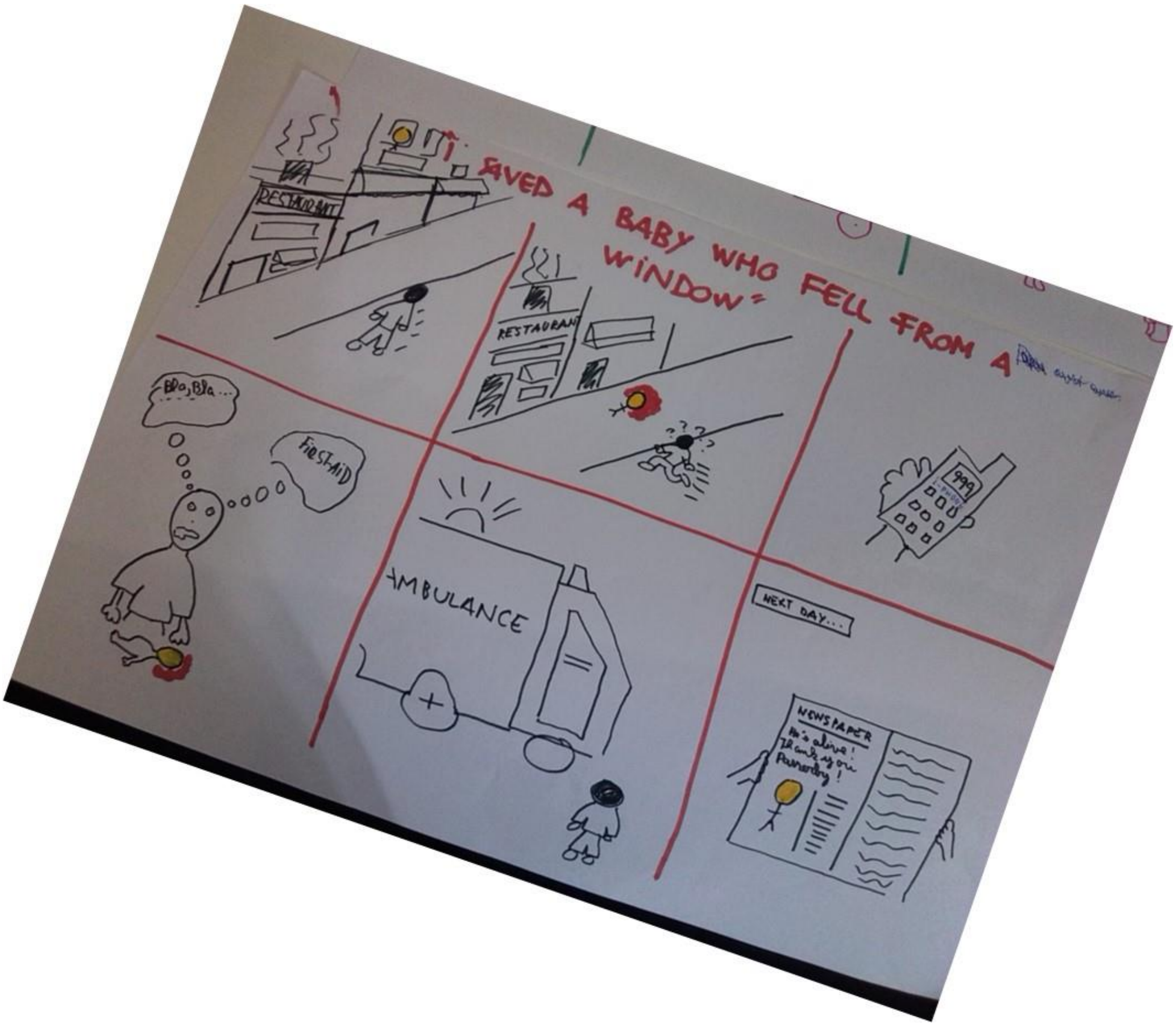
Date ____

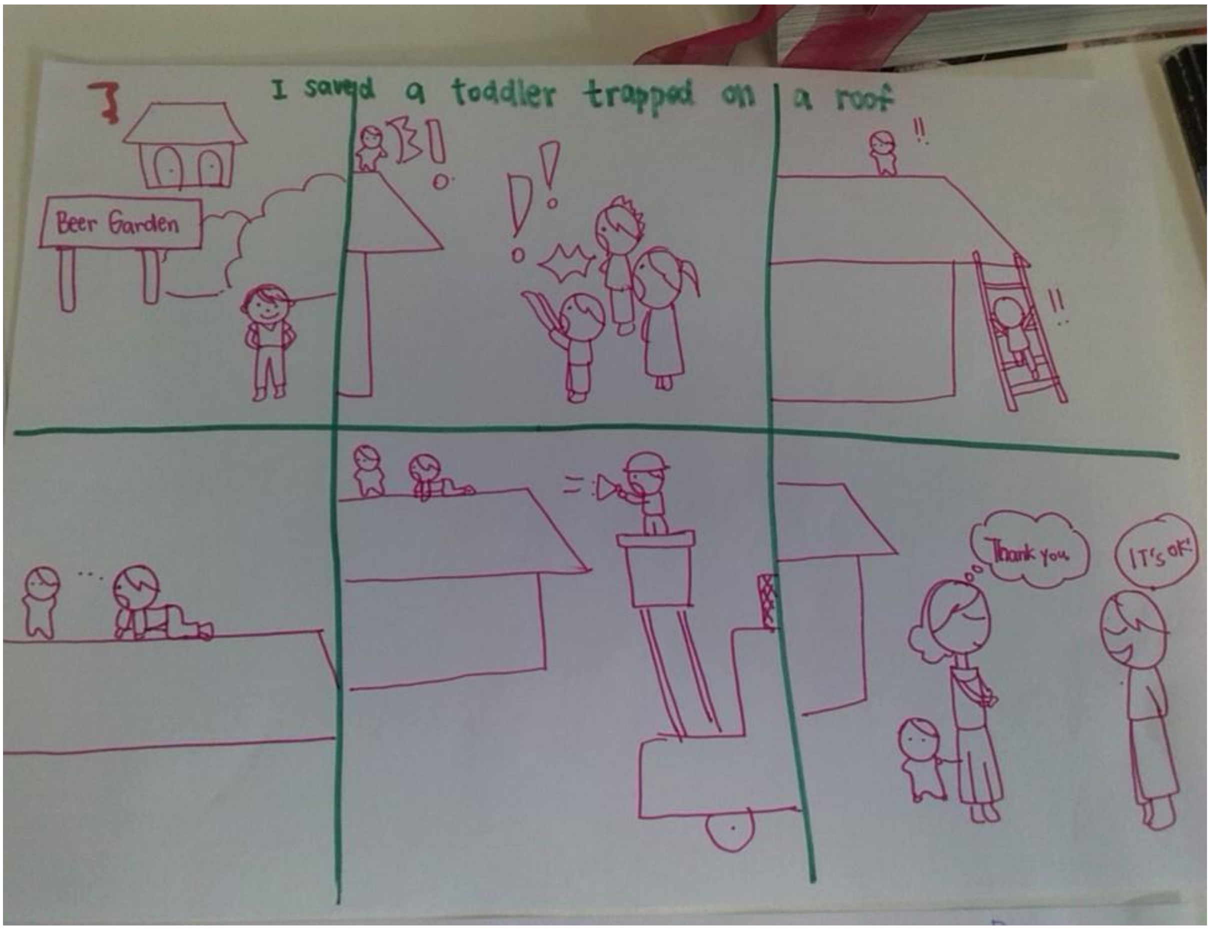
Email (optional)

Appendix 8 – PRAAT script for measuring fluency

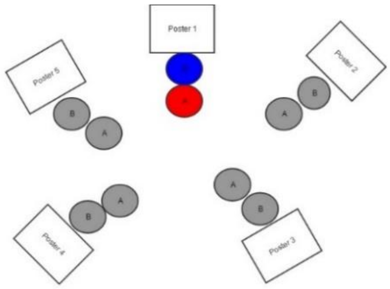
```
writeInfoLine ("This script has started.")
select all
Remove
tableID = do ("Create Table with column names...", "table", 0, "speaker task part tier interval label duration")
soundFileName$ = chooseReadFile$ ("Open a sound file.")
textGridFileName$ = chooseReadFile$ ("Open a TextGrid file.")
soundID = do ("Read from file...", soundFileName$)
textGridID = do ("Read from file...", textGridFileName$)
form Speaker + Task + Part
	real Speaker 1 (= default)
	real Task 1 (= default)
	real Part 1 (= default)
endform
select textGridID
numberOfTiers = do ("Get number of tiers")
tiersSafeSubstraction = numberOfTiers - 4
rowNumberCounter = 0
for tiersBeingAnalysed from tiersSafeSubstraction to numberOfTiers
	numberOfIntervals = do ("Get number of intervals...", tiersBeingAnalysed)

	if tiersBeingAnalysed == tiersSafeSubstraction or tiersBeingAnalysed == tiersSafeSubstraction + 1
		for intervalBeingAnalysed from 1 to numberOfIntervals
			select textGridID
			labelOfInterval$ = do$ ("Get label of interval...", tiersBeingAnalysed, intervalBeingAnalysed)
			if labelOfInterval$ == "0"
				# appendInfoLine ("This is zero...")
			else
				startInterval = do ("Get start point...", tiersBeingAnalysed, intervalBeingAnalysed)
				endInterval = do ("Get end point...", tiersBeingAnalysed, intervalBeingAnalysed)
				durationInterval = endInterval - startInterval
				select tableID
				do ("Append row")
				rowNumberCounter = rowNumberCounter + 1
				#Another option would be: "rowNumberCounter += 1"
				do ("Set numeric value...", rowNumberCounter, "speaker", speaker)
				do ("Set numeric value...", rowNumberCounter, "task", task)
				do ("Set numeric value...", rowNumberCounter, "part", part)
				do ("Set numeric value...", rowNumberCounter, "tier", tiersBeingAnalysed)
				do ("Set numeric value...", rowNumberCounter, "interval", intervalBeingAnalysed)

				do ("Set numeric value...", rowNumberCounter, "duration", durationInterval)
			endif
		endfor

	else
		for intervalBeingAnalysed from 1 to numberOfIntervals
			# appendInfoLine (intervalBeingAnalysed)
			select textGridID
			labelOfInterval$ = do$ ("Get label of interval...", tiersBeingAnalysed, intervalBeingAnalysed)
			if labelOfInterval$ == "v" or labelOfInterval$ == "vm" or labelOfInterval$ == "u" or labelOfInterval$ == "um"
				startInterval = do ("Get start point...", tiersBeingAnalysed, intervalBeingAnalysed)
				endInterval = do ("Get end point...", tiersBeingAnalysed, intervalBeingAnalysed)
				durationPauseInterval = endInterval - startInterval
				select tableID
				do ("Append row")
				rowNumberCounter = rowNumberCounter + 1
				do ("Set numeric value...", rowNumberCounter, "speaker", speaker)
				do ("Set numeric value...", rowNumberCounter, "task", task)
				do ("Set numeric value...", rowNumberCounter, "part", part)
				do ("Set numeric value...", rowNumberCounter, "tier", tiersBeingAnalysed)
				do ("Set numeric value...", rowNumberCounter, "interval", intervalBeingAnalysed)
				do ("Set string value...", rowNumberCounter, "label", labelOfInterval$)
				do ("Set numeric value...", rowNumberCounter, "duration", durationPauseInterval)
			else
				#Do nothing
			endif
		endfor
	endif
endfor
select tableID
View & Edit
speakerString$ = string$(speaker)
taskString$ = string$(task)
nameOfDataFile$ = "data" + "_" + speakerString$ + "_" + taskString$ + ".csv"
do ("Save as comma-separated file...", nameOfDataFile$)
appendInfoLine ("The script has ended.")
```

Appendix 9 – Examples of storyboards from carousel session

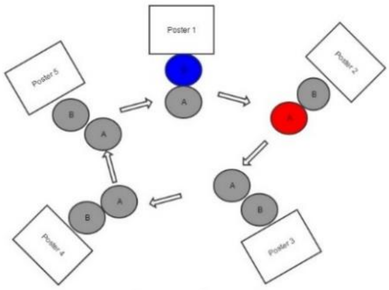

I SAVED A BABY WHO FELL FROM A WINDOW
RESTAURANT
RESTAURANT
Bla, Bla...
FIRST AID
999
AMBULANCE
NEXT DAY...
NEWSPAPER
He's alive! Thanks to you Passerby!


I saved a toddler trapped on a roof
Beer Garden
Thank you.
IT'S OK!

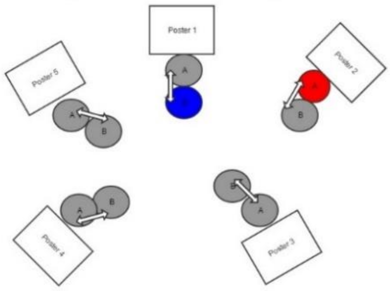
Appendix 10 – TR Carousel (diagram)



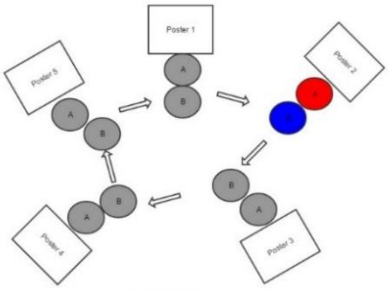
Students stand by their poster



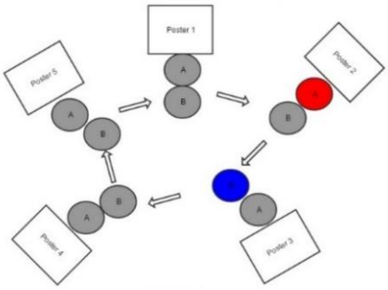
'A's move in a clockwise direction to the next poster. 'B's present their poster.



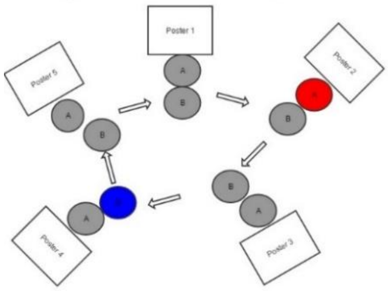
Once the 'B's have finished, 'A's and 'B's swap positions



'B's move in a clockwise direction to the next poster. 'A's present the new poster



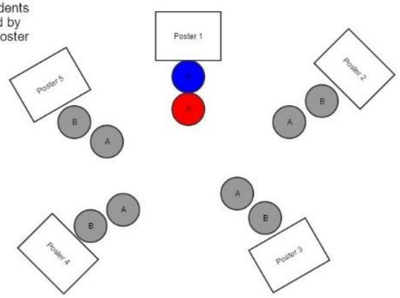
Once the 'A's have finished, 'B's move in a clockwise direction to the next poster. 'A's present the same poster again



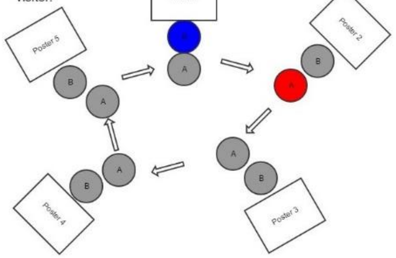
Once the 'A's have finished, 'B's move in a clockwise direction to the next poster. 'A's present the same poster for a third time

Appendix 11 – PR Carousel (diagram)

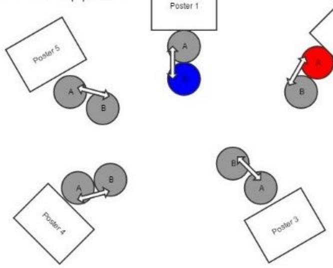
1. Students stand by their poster



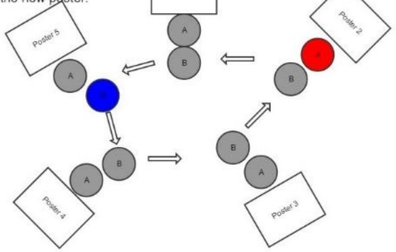
2. 'A's move in a *clockwise* direction. 'B's present their poster to the visitor.



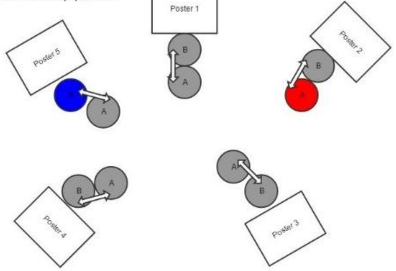
3. 'A's and 'B's swap places.



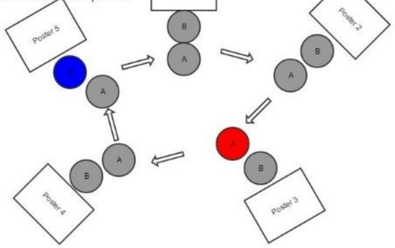
4. 'B's move in an *anticlockwise* direction to the next poster. 'A's present the new poster.



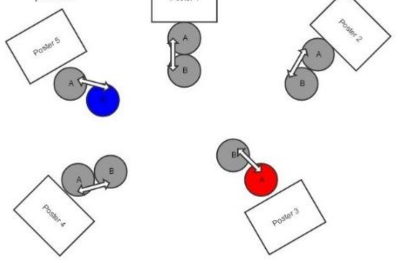
5. 'A's and 'B's swap places.



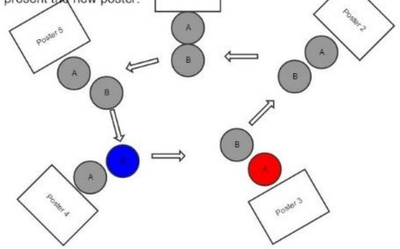
6. 'A's move in a *clockwise* direction to the next poster. 'B's present the new poster.



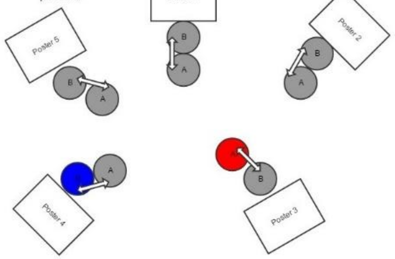
7. 'A's and 'B's swap places.



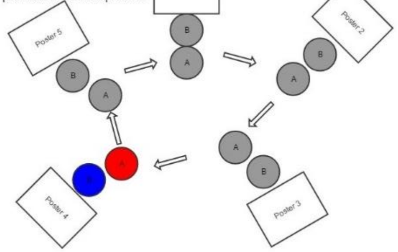
8. 'B's move in an *anticlockwise* direction to the next poster. 'A's present the new poster.



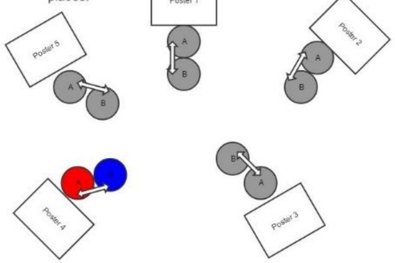
9. 'A's and 'B's swap places.



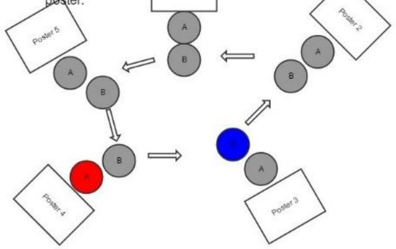
10. 'A's move in a *clockwise* direction to the next poster. 'B's present their new poster.



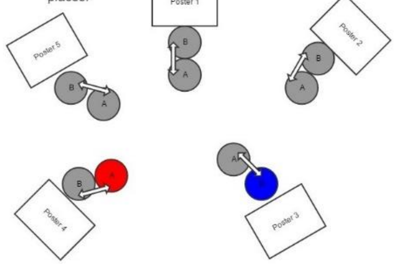
11. 'A's and 'B's swap places.



12. 'B's move in an *anticlockwise* directi on to the next poster. 'A's present their new poster.



13. 'A's and 'B's swap places.



14. 'A's move in a *clockwise* direction to the next poster. 'B's present their new poster.