A CLASSROOM-BASED, MIXED METHODS STUDY INTO THE INFLUENCE
OF TRANSCRIBING, REPORTING, AND TASK REPETITION: HOW DO THEY
IMPACT IN-CLASS STUDENT SPOKEN TASK PERFORMANCES

By

JASON MOSER

A thesis submitted to the

University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Centre for English Language Studies
Department of English
University of Birmingham
June 2012

Abstract

This study is a classroom-based, mixed methods study into the influence of transcribing, reporting, and task repetition on in class student oral task performances. The study investigates two questions. First do students in an intact classroom improve task performances when they repeat the same task in subsequent performances? In a previous exploratory study by Moser (2008) students did not take advantage of task repetition opportunities to improve a repeat task performance. It was concluded that the reason for this was that amongst many students there was a lack of perceived pedagogical rationale for task repetition. On this point and more specifically the study investigates does a more transparent pedagogical focus realized through a transcribing phase or a reporting phase prior to a repeat task performance result in improved subsequent task performances. Related to this, and the second question of this study, is does the more intensive transcription work result in improved task performances than the reporting work? The results of the study reveal no significant difference between transcribing or reporting on subsequent task performances; however, there was significant results for a task repetition effect on task performances. The classroom implications of these findings will be discussed.

TABLE OF CONTENTS

i

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction and background

1.0 Introduction

This study is the third module of a three module task repetition study. The first module involved a general literature review, and the framing of the exploratory study conducted for module two. In the study in module two I investigated the effectiveness of immediate task repetition in a large EFL classroom context. Module three, this module, is a follow-up study to module two which attempts to solve the problems uncovered in module two in relation to effectively using task repetition in the classroom. This thesis consists of seven chapters. In the first chapter I review the findings from the study in module two and explain the research agenda for study in module three, including a short literature review followed by an introduction to my teaching context and the basic design of the study. In chapter two the complexity, accuracy and fluency (CAF) constructs used to measure student oral task performance are explained, including discussion of codifying issues encountered and how the data was transcribed. The third chapter discusses the statistical tests used. The overall design was a three-way ANOVA with two factors varied between-subjects (transcribing versus reporting and four tasks) and one within-subject factor (the four times a task was done). Chapter 3 includes the findings for each CAF measure and their analysis.

Chapters four, five and six are the case study portion of this study. The fourth chapter explains my rationale for using a case study approach, including discussion of the approach adopted and how I chose the student. The first part of this chapter provides a short explanation on how to read student transcripts. This is crucial for understanding the proficiency level of the students. The second part gives an in-depth analysis of my

case study student's pre-task work and first performance. This analysis centres on questions formulated after analysing the quantitative data and transcripts. Chapter five analyses the transcription process and second task performance. Chapter six is the final chapter concerning the case study and examines the two final task performances. Chapter seven is the final chapter in the study. The first section summarizes the main findings, including comparison of my module three findings with my module two findings in order to provide an overall summary of the study. In the second section I discuss the limitations of this study, considering how the quality of my measurement tools, in particular CAF constructs, can be improved and my future plans for revisiting the research data. In the final part of this section I also describe the challenges of doing classroom-based research. The last portion of the chapter discusses the potential classroom implications of this study for my teaching context, providing specific ideas for realizing my findings in the classroom and across the larger language curriculum.

1.1 Revisiting module one and module two

This section briefly reviews module one and two to set the context for module three. This study originated from my regular use of immediate task repetition in my classroom, which involves asking students to repeat a task again with a different partner or partners. I regularly employed this pedagogic tool for three reasons. First, it expands the time students spend doing task work, an important point considering my students, like many students in EFL contexts, have limited opportunities to speak English, either inside or outside the classroom. In my classroom, for example, most pair work task performances rarely go beyond three minutes before students cannot continue, thus task repetition helps extend time on task. The second reason for using

2

task repetition is that it offers students the opportunity to work with different partners, and as my students have indicated in past surveys that one of the most enjoyable parts of their communication classes is working with different students, this hopefully positively impacts student motivation and engagement in the classroom. The third reason is it has been suggested that repeating a task frees cognitive resources, allowing students to perform better in subsequent task repetitions. In task-based learning research it is generally argued that the freeing of cognitive resources is realized through online planning, strategic planning, and task repetition, all three of which were reviewed in detail in module one (for a recent review see Ellis 2009a).

In brief, the first two, online planning and strategic planning, facilitate different types of language production. Bygate and Samuda (2005) note that online planning facilitates the monitoring of lexico-grammatical elements but hinders fluency, discourse planning, and lexical variety. Strategic planning, because of working memory constraints, predisposes students to plan general ideas (meaning) to the detriment of micro planning (form). In task performances strategic planning facilitates fluency and complexity to the detriment of accuracy. Bygate and Samuda (2005) have demonstrated that task repetition, also called integrative planning, is potentially a combination of both strategic planning and online planning and therefore, unlike the other two types of planning, can simultaneously impact articulation, conceptualization, and formulation (see Levelt 1989 for an explanation of these terms).

Module two explored task repetition in my classroom context, specifically examining my students' productivity in large communicative classrooms (35 plus students per

class) during pair work involving immediate repetition of the same open-ended topic-based task with different partners. The research design was inspired by a number of task repetition studies that have shown positive results (see Bygate 1996, 2001, Bygate and Samuda 2005, Essig 2005, Gass, Mackey, Fernandez, and Alvarez-Torres 1999, Lynch and Maclean 2000, 2001, Pinter 2005, Birjandi and Ahangari 2008, Ahmadian and Tavakoli 2011), research which has perhaps resulted in the recommendation in practical language teaching books like Edwards and Willis (2005), Thornbury (2005), Willis and Willis (2007) and Nation and Macalister (2010) of task repetition as an effective pedagogical tool to use in the classroom. However, as most of the research to date has been conducted in lab settings, involving monologue/narrative tasks, and including some time delay between task performances, it still remains unclear whether these same conclusions apply for intact classrooms, using open-ended tasks, with immediate task repetition. Thus my module two research explored whether, in my context, with intact classes using open-ended tasks, the general research findings that task repetition results in improved student production held true.

In my teaching, because of large class sizes, it was essentially the students' responsibility to take advantage of the learning opportunities made possible through immediate task repetition. To use Williams' (2005: 679) term, the "locus of responsibility" was on the students. In large classes during pair work there was little other option and I as the teacher was relegated to mostly an observer role once the task started. While I was confident students would take advantage of the benefits afforded by task repetition, particularly considering the positive benefits noted in the research cited above, two concerns remained. First was the possibility students would

4

become uninterested repeating the same task, particularly in light of Ohta's (2001) findings that her students became disengaged when repeating the same closed task, with their task performances becoming elliptical. While Ohta (2001) solved the problem by introducing new information each time to expand the task, in my context I hoped changing partners combined with an open-ended task would mitigate potential problems caused by repeating the task. However, while I was confident my concerns would prove unfounded, this assumption was mostly based on observations of the noise level in my classes and not from an in-depth analysis of student production. In module two I was able to answer these concerns directly through the research I conducted in my classroom.

My second concern regarded the quality of repeated task performances if students, rather than reworking a prior performance, talked about something new, especially because it was unclear how students would approach a repeat of a task performance. My hope was that they would rework their prior performance rather than starting anew, thus repeating what they had previously produced, and I encouraged students to do this, based on findings from task repetition research that if students do rework a prior performance, improvements in production can be expected. However, since the tasks used were open-ended and allowed students to pursue themes of interest guided by a general topic, it was unclear whether they actually did rework their prior task performances into subsequent repetitions. In summary, my research agenda for module two built on previous repetition studies by considering how students react in intact classes, where languages are taught, rather than in laboratory settings; my research employed open-ended tasks, rather than monologue/narrative tasks; and

finally students repeated the task immediately, rather than after some time delay between performances.

1.2 Exploratory research findings for module two

For module two I recorded 45 students in my regular classes doing the same open-ended task three times in a row with three different partners. I reviewed the raw data and concluded that the overall trend clearly demonstrated that students' task performances in terms of productivity and overall quality either did not improve or deteriorated. While they might have repeated a prior performance, there was no noticeable improvement, and often the repeat performances became elliptical, similar to the difficulties Ohta (2001) describes. Even when students changed topic, in most cases it was hard to consider later performances as improvements on the language from the first task iteration. There was a small group who did improve; of the 45 students no more than 12 appeared to make some use of the task repetition to improve their performance. Overall, these results were unexpected and thus deserved some additional analysis. To better understand the results, I chose to analyze in more depth four students' production, two students who had done poorly when repeating tasks and two who improved their task performance. The latter pair appeared to remain engaged in subsequent task work by reworking a large portion of their prior task performance. They seemed committed to improving subsequent task performances by retelling a specific story or "center of interest" (see Chafe 1980: 26 for a definition). By retelling a single story, students were able to improve and expand on it in repeat performances. They engaged in what I termed a "self-structuring task performance strategy" (Moser 2008:73). It also appeared that use of this strategy had little to do with proficiency level, as one of the students who employed this strategy was the

weakest of the four chosen for this more in-depth analysis. One variable that may have been overlooked in students' reworking prior performances was whether they were especially interested in telling a story based on that particular topic. However, the issue may be more than just having a story to tell; it may be more about students recognizing the purpose of repeating the task and thus creating a story to maximize the opportunities afforded by being able to rework their language during repeated performances, the pedagogical purpose of task repetition.

In contrast to the productive students, the students with poor subsequent task performances appear to have not recognized or acted on the pedagogical purpose of immediate task repetition. Reinforcing this conclusion, the two students used to further investigate poor repeat task performance wrote in their survey results that they saw repeating a task as a chance to talk to other students. More generally, in the same survey some of the students whose task performance did not improve indicated they understood the pedagogic benefits of task repetition, but the majority commented that they viewed task repetition as an opportunity to mingle and chat, which their task performances seemed to corroborate. I concluded most students became bored and disengaged repeating the same task as the social benefits of working with a new partner were evidently insufficient to sustain student task engagement. This result is unfortunate because in many cases this class was my students' only opportunity to speak in English, within my class the time for oral task work including repeating the task would not exceed 20 minutes in a 90-minute lesson, and within this 20 minutes the average student's productive talking time would in most cases not exceed ten minutes. Considering such speaking opportunities were so limited for most of the

students, the decrease in task performance during the repetition phase is particularly disappointing.

1.3 Setting the classroom research agenda for module three

My main finding from module two reflects Bygate and Samuda's (2009) warning that it is misguided to see communicative task work as automatically entailing language learning. They argue communicative tasks must involve some transparent pedagogical purpose (Bygate and Samuda 2009), which in the module two research design was intended to come after the task repetition in the form of students writing a report on their task work and then a few students (usually no more than three) were called on to report to the class and receive feedback and comments from me as the teacher. Unfortunately, the purpose of these reflection and reporting tasks must not have been pedagogically transparent to the student participants, as they did not appear to motivate the students to improve their repeat task performances. In hindsight it is questionable whether students viewed the report stage as of equal pedagogical importance to the task repetition phase, particularly since only a few students could be called on to report because of time limitations. There were also working memory limitations that prevented students from recalling substantial details of their task performances, resulting in reports that were often simple, basic summaries.

My focus for module three thus became an action research (Burns 2010) study intending to resolve the problem discovered with task repetitions in module two. With the module two task repetitions the *"locus of responsibility"* (Williams 2005: 679) for improving subsequent performances was placed solely on the students as I as the teacher did not offer any intervention or language advice between task performances. For module three I concluded that as the teacher I needed to play a more active role

before the task repetition phase and so decided to add teacher-directed practice and analytical activities between the task repetitions instead of students repeating tasks immediately. This included making the students' actual performances more visible to them, based partly on Thornbury's (2001: 70) contention that it is not enough for teachers to get students to talk, but that they must also "capture" or "pin down" their performances to make them available for analysis. I had thought the reporting stage as described in module two would have accomplished this, but in practice it proved insufficient. Thus in this module students were asked to work on recordings of their actual task language rather than summarizing previous performances. It was hoped this use of actual task language would overcome the short-term memory constraints (see Skehan 1998), which make recall of details of what was said during a task problematic, thereby limiting the efficacy of using students' oral summaries to improve their interlanguage. Additionally, it is likely that the language they cannot recall is also language where their interlanguage problems are most acute. In contrast, language that students can recall is likely to be already automatized and therefore easy to recall. Thus, if having access to the most challenging parts of students' production is necessary to facilitate noticing and automatization, then the pedagogical intervention used should assist students in accessing that challenging language. Such access to this more difficult language can then lead to reformulation work that can be consolidated in practice activities or attempted in a subsequent task performance work which should ultimately aid in the long-term acquisition of the language in question.

As spoken production is a key medium (Hughes 2002, Holliday 2005) for facilitating language acquisition, it is a central cornerstone of many language programs. Ironically, despite this preoccupation with speaking, in my experience, very few of the

teachers I work with have regular opportunities outside of tests to see or hear what students produce during task performances, and proposed solutions from the literature are often inadequate. For example, Nation and Macalister (2010: 59) suggest teachers should "stand by students" when they are doing a task. Looking for a solution that would result in greater student engagement with the task repetitions and inspired by Lynch (2001 2007), I decided to incorporate self-transcription into my revised lesson plan as a tool for capturing student language. Students were to self-transcribe their task performances, followed by corrective feedback and reformulation work, which would in turn set the stage for a subsequent task performance. In the next two sections I briefly review Lynch's (2001, 2007) studies to clarify how this study differs from them.

1.4 Out of class student transcription

This section offers an overview of Lynch's (2001) first transcription study. Lynch (2001) writes that the idea of having students transcribe their task production evolved out of his task repetition studies (Lynch and Maclean 2000, 2001), which concluded teachers need to create opportunities for students to analyze their performances. Lynch's (2001) first exploratory transcription study was out of class and involved eight volunteer subjects self-transcribing their task performances. Lynch's eight students, in four pairs, first performed a role-play task and then selected roughly two minutes of recording to transcribe. Next they were asked to edit, reformulate, and make grammatical and lexical corrections. Editing refers to removing dysfluencies while reformulation refers to making changes intended to make meaning more "precise" (Lynch 2001:129) and "inserting" (Lynch 2001: 129) new language into the original performance. The objective of the research was to investigate to what extent

students can notice and self-correct their errors or language problems and to analyze the types of changes students make to their production based on a self-analysis of their transcriptions. Lynch found that on average each pair made 28 changes to the transcripts with grammatical changes the largest single category of correction. Lexical changes were relatively small. Lynch concluded that the majority of all changes made by the students were better or correct. He was impressed that with less than ten minutes of spoken performance, nearly 200 language points in total were covered, with students noticing and correcting the majority of them. Lynch then engaged in a follow-up study to his 2001 research, which is described below.

1.5 In-class student self-transcription

Lynch's follow-up study (2007) investigated the feasibility of student self-transcription in an intact classroom. He points out that research to date suggests students do not effectively take advantage of negative feedback in the form of recasts or explicit feedback from teacher or peers (for a review of this literature see Lyster 2007). He therefore wondered if transcribing as "offline feedback" (Lynch 2007: 312) could facilitate student modification or reprocessing of their production. Lynch's thought was that since speech becomes "visible" (Lynch 2007: 312) through transcription, it might increase the probability of students noticing and reprocessing highlighted forms in a subsequent performance. This assumption is partially supported by Smith (2004) and Lai and Zhao (2006), who in their studies on noticing and text-based chatting found student noticing can be heightened by the "visual saliency" (Smith 2004: 372) of written speech. In Lynch's (2007) second study he established two groups. The task was a bank role-play task. In one group the students, after doing the task, self transcribed their recorded performance and with their partner

11

edited, corrected, and reformulated it. The students then submitted the transcript for further correction and reformulation by the teacher. In the final stage students discussed the changes with their teacher before repeating the task. In the other group the noticing was teacher-initiated. The teacher listened to the pairs' performances, selected incorrect parts, transcribed them, and then gave the transcript to the students to correct in pairs. They then discussed the changes with the teacher before listening to a native speaker performance. This was followed by a repeated performance of the task. In this group the feedback was teacher-initiated as opposed to student-initiated. In addition the feedback was limited and general in contrast to the detailed and specific feedback from the transcribing process.

Lynch (2007: 312) hypothesized that the transcribing group would demonstrate a "greater depth of processing" because of actively working on their performances. The teacher-directed group on the other hand would be "passive users" (Lynch 2007: 318) of the teacher's transcriptions. He found that the active transcribers, despite being the less proficient group, achieved a higher level of accuracy in a subsequent task performance (64% vs. 47%) on items that had been focused on in the transcription process. Lynch (2007) concluded that transcribing could be an effective tool for getting students to improve their accuracy in task performances and that transcribing could be used as a regular activity with intact classes. Student feedback regarding the transcription activity was also very positive.

The limitations of Lynch's (2007) study are worth mentioning because they provide another impetus for this study. First, the study was done with a small number of students all with a level of English proficiency sufficient to attend a British university. Their IELTS listening scores ranged from 4.5 to 7.5. Second, the transcription lesson

comprised only nine hours of a sixty-hour course, and so whether student transcription is effective as a regular part of class work remains unclear. Lastly, the task used for Lynch's (2001, 2007) studies, as with another transcription study (Mennim 2007), was a closed task. In Lynch's (2001, 2007) case it was situation specific and ensured the same language would have to be repeated in subsequent task performances. Thus it isn't clear whether an open task, where the language of the task performance can be changed more easily, would result in the same production benefits.

1.6 Research questions for module three

In this study there are two sets of research questions. The first are derived from the findings in module two while the second category of questions relate to questions that arose during this study. The first set of questions are presented below and investigated in the quantitative portion of the study in chapter three. Based on a synthesis of my module two research design and Lynch's (2001, 2007) self transcription studies, this module investigates two main questions in the quantitative portion of this study. The first relates to my module two findings, which concluded that when students immediately repeated a task performance their language production did not improve. Thus this module investigates whether an intervention between task repetitions can resolve this problem and encourage students repeating the same task to improve their production. To accomplish this, a set of learning phases which make the pedagogical purpose of task repetition more salient were inserted between task repetitions. I chose two strategies to make the purpose of repetition more apparent; student self-transcribing and reporting. While module two incorporated reporting, in module three the sequencing and nature of how the language from the task work was reported was

adjusted to include some language work between task repetitions. Section 1.9 provides a detailed description of the two different types of between-repetition interventions; self-transcription and reporting. Thus the first question for module three is: Do transcription and reporting between iterations within a task repetition sequence facilitate student task engagement and improved language performance?

As a supplement to this first question, like Bygate (2001), repeat performances after a longer gap in instruction is also included in the research design, in this case incorporating data from student production nearly a month after the initial task performances. My second question considers whether there would be any difference in results between the reporting and self-transcription groups. The problems encountered with task repetition in module two, assumed to have been caused by the pedagogical opaqueness of the report stage, and the positive findings of Lynch's (2001, 2007) studies, encourage the hypothesis that the self-transcription group will outperform the reporting group. This could be due to greater teacher and student involvement in the transcribing process and students being able to see and hear their actual performances resulting in better performances than simply reporting. Skehan (2007: 58) believes that, at least in the case of language accuracy, transcription is "more powerful" than a public performance, or reporting. Thus the second research question is: Does the transcription group outperform the reporting group in repeated task repetitions after pedagogical interventions? and Do any differences persist one month after the lesson? These two questions make up the main focus of the quantitative portion of the study, and are further addressed in the latter qualitative chapters. Starting in Chapter 3, I also ask supplementary questions that emerged from the study and the results. For example in Chapter 3 I address questions relating to the

significant results for task effect, and the significant results for interactions between task and a few of the CAF measures. In Chapters 4,5, and 6, which are the qualitative sections of the study, I formulate different questions. These questions again emerged from the results and findings of the study. At the start of each chapter I introduce the questions and the reason for asking them. The chapters correspond with the sequence of the lesson in this study. In Chapter 4 my focus is on pre-task work and the first task performance. In Chapter 5, I focus on the correction and reformulation process, and the second task performance. Finally, in Chapter 6 my questions centre on the third and fourth task performances. The different focuses mean that for the most part I ask different questions. Some of the questions as will be pointed out apply specifically to the focus of the chapter while a few questions apply to all task performances.

1.7 Participants and the teaching context

1.7.1 Participants

This study involved two intact classes of 20 students each at a women's university in Japan. The study was conducted in the first semester, covering fifteen weeks. The students were second year non-English majors from different departments. At the university students are divided into classes based on an English placement test and both were the top class in their respective departments, although my impression as their teacher was that almost all the students were low in language proficiency, perhaps due to the university's low academic ranking. The very top scores on the placement test are roughly equivalent to 3.5 on the IELTS test. Based on my own assessment, I would suggest none of the students in this study would be able to enter regular undergraduate programs in a British university without additional intensive

15

English study, thus they are of a considerably lower level of English proficiency than the students involved in Lynch's (2001, 2007) studies.

The demographic included in this research, students from a less academically competitive university, is important because a number of task-based studies from the literature investigating the Japanese university context originate from highly ranked universities (see Mennin 2007, Weaver 2007, Stillwell, et al. 2010) or with motivated, proficient students in laboratory settings (see Lambert and Engler 2007). The students who participated in my module two research followed this trend; they were from a well known public university for English language education. From talking to other colleagues in Japan one reason why they prefer not to do research in lower ranked universities is because the students are less proficient and often indifferent to English. This means collecting data from student task performances in such contexts, particularly oral data, is challenging and risky, especially since the bias in publishing is toward research that demonstrates statistical significance in the measures employed. Thus the student participants investigated here represent an under-researched demographic in Japan; students from a private, less academically competitive university, a context I believe typical of many universities in Japan. In these universities students often have negative experiences of learning English from junior and senior high school that carry over to their university study (See Murphey 2004). Furthermore, these students, like the majority of non-English majors studying English in universities, are taking English as a required subject and thus may be more interested in passing the course than in improving their proficiency. With the students in this study this is partially illustrated by the fact that once they finish their required English classes very few enrol in subsequent elective English courses.

1.7.2 The teaching and learning context

The students in the study were in the second year of a two year required English program for non-English majors. Students had two 90-minute classes a week with two different teachers. I taught them once a week. This is what Swan (2005:16) calls a "3hpw" (three hours per week) context. The limited aim of the English for general purposes program is to help students achieve a basic level of communicative ability. The task-based program emphasizes the development of procedural knowledge over declarative knowledge (see Johnson 1996) and maintains a weak "interface position" (Ellis 2009a:21) in that it assumes explicit knowledge can convert to implicit knowledge through exposure and practice if the students are developmentally ready. Experience shows that while our students lack procedural knowledge they do have sufficient declarative knowledge from their earlier English studies. The program premise, mostly untested, is that having students engage in communicative tasks builds their lexicalized language and can, through task repetition, improve their access to their declarative knowledge, moving that knowledge along the cline toward implicit knowledge. In this sense our program mostly focuses on what students are already familiar with and can potentially proceduralise.

Support for this pedagogical focus of prioritizing the development of procedural knowledge at the beginner stages through tasks comes from Ellis's modular approach (see Ellis 2002, 2003). In this approach unfocused communicative tasks are the main component of a syllabus that emphasizes the development of lexicalized language. This approach makes sense because a basic store of formulaic language is a precondition for syntactic development (Skehan 1998), a goal supplemented by the

17

additional target of developing a basic level of communicative ability among students, the stated focus of the program. From an affective standpoint we also believe developing formulaic repertoires of language through tasks builds self-confidence and motivation and renews student interest in learning English. Program feedback has supported these assumptions of the affective benefits of the task-based learning program.

At the start of the semester I divided the 20 students in each class into two groups of ten. Each group performed transcription twice and reporting twice. There were two reasons I divided the students this way. First, because the transcription was more labour intensive it was not possible to have one group of students do this for the whole semester while the other group did the less labour intensive reporting. I divided students into groups based on their English language placement test scores, the only information available describing their English proficiency prior to the start of the course. For each class the top, middle and weaker students were divided into two groups. For example, the highest scoring student went into group one and the second and third highest students went into group two with the fourth highest student going into group one. Table 1 illustrates the test scores and student placement into the treatment groups. The placement test did not have writing or speaking components but observation showed students with the top scores had more productive English ability than students with lower scores.

Table 1.0 Student placement into groups by class according to placement test scores

Class One	
Group one	Group two
238	241
235	232
226	232
221	216
212	213
211	210
208	209
208	208
207	207
206	206

Class Two	
Group one	Group two
270	275
221	209
205	208
201	200
190	195
189	188
185	184
180	183
179	178
175	177

After dividing the students into the two groups I ran a paired-samples *t* test and found no problems with the groupings. The second reason for dividing the students the way I did is that, as the placement test scores show, class one was stronger than class two. These placements scores only confirmed what I knew about the two departments. I felt the two classes could not be used as separate, intact comparison groups because class one was clearly stronger as one of the departments is more academically competitive than the other. This difference in student English proficiency between classes formed the motivation for creating two combined groups from both classes with each group doing both treatments (reporting and transcribing). It is for this latter reason that the study can be considered to be a pseudo between-subjects design (discussed in further detail in Chapter 7). What this means is that although all the

students contribute the data for each treatment, it is treated statistically as if the students contributed to only one of the treatments. Each group in an alternating pattern did each treatment twice.

1.8 Tasks

This research employed open-ended topic tasks, commonly used at the beginner/false beginner level in many EFL language programs (see Saito-Abbott 2004 and Weaver 2007). As Saito-Abbott (2004: 130) comments in her university's task-based Japanese language program, students start with a theme-based syllabus before being gradually exposed to more content-based learning in the latter part of the program. This progression corresponds with Estaire and Zanon's (1994: 21) "wheel theme generator," where themes start from immediate and surrounding life then move outward toward societal themes. In this study task is seen as a 'flexible pedagogical tool' (Samuda and Bygate 2007: 229) that is shaped to fit different teaching contexts. For this reason my selection of task may not entirely correspond with more prescriptive definitions of task (see Ellis 2003).

The key characteristic of task in my teaching context is that it must primary involve an attempt by students to be meaningful. The basic task for this study involved pairs taking turns asking each other a set of questions based on themes related to their immediate life and experiences. A task would start with the interviewer asking a question from the list provided. The interviewee was expected to respond, and in turn the interviewer was encouraged to push the interviewee's output by asking following up questions. This type of communicative task centred on a set of questions is very common in beginner textbooks as well. I have found that it is popular amongst students and teachers in my EFL context. The reason for its popularity is I believe

students find it more meaningful and motivating then closed or referential tasks. For this reason they seem to generate more output than other task types.   In this study one type of task was used, but it should be mentioned that I often use different task types to support a main task, which is usually open-ended communicative interview task.

Regarding the themes of the tasks used, as Ellis (2003: 120) and Willis (2004: 31) emphasize that choosing topics should be an important part of task-based lesson design, some thought was put into this part of the research. This is particularly important since in much task-based research the process of topic selection has been neglected (Ellis 2003). Thus the objective was to select task topics familiar to the students and intrinsically motivating. Considering this, I chose four task themes; *pets and animals*, *shopping and fashion*, *health and exercise* and *dating and marriage*. None of these topics were used in the students' other English class. In the first year of the program students had at least one lesson each on the second and third topics but the other two topics were first exposures for almost all the students.

I prepared a set of questions for each topic (see Appendix 1). This topic-based question task format is popular with textbooks and English classrooms. For example, Weaver (2007) used a number of task types including one that involved students asking each other five questions about their holiday. For this study I decided to make the questions in order to save in-class time for other activities. When making the questions I considered a number of important issues, including using questions that would maximize student production. This required choosing a breadth of questions so all students would have something to talk about. In this sense I visualized how students would respond to the questions. For example for the topic *pets and animals* I

knew that some students would not have pets, so I included questions about stuffed animals, pets they would like to have, and their favorite animals. I also included questions that pushed for monologic story telling, such as asking students to describe a scary experience with a pet or animal.

The questions provided the students with a basic schema for each topic, which is important because students, especially beginner students, may have the basic language resources necessary to complete the task, but they may be unaware of how to approach or talk about the topic in the target language. Thus the questions for a topic provide what Bygate and Samuda (2009: 96) call the "field" of a task. The field of the task contains conceptual, linguistic and discourse features, and the term field is used to suggest it is open rather than closed (Bygate and Samuda 2009). Another way to view field is as a schema. Widdowson (1983: 37) notes schemata are stereotypical patterns of cognitive structure, which in regard to propositional content are "frames of reference" and in regards to illocutionary activity are "rhetorical routines." In explaining the importance of schema in second language acquisition, Widdowson (1983: 38) writes:

> [What] I would want to suggest is that as the child abstracts his linguistic rules from the mass of language data, so he also abstracts contextual outlines from the recurrent circumstance of language used and associates these outlines with linguistic realizations. So the language that is learned retains a trace of situational provenance.

In summary, the four tasks designed followed four different themes, *health and exercise, pets and animals*, *shopping and fashion*, and *dating and marriage*, around

which questions were developed for students to ask one another, and which established a basic schema or field for the task. This hopefully made performing the task easier, pushed students to maximize their production, and helped developed awareness of the how to talk about the themes. I expected that all of these features would further be facilitated by the use of task repetition after intervening language work with the students, as described in 1.6.

1.8.1 Determining task roles

At initial trialling of the tasks a number of issues arose concerning the dialogic nature of the tasks and the language produced. As Jenks (2007, 2009) notes, task participatory structures influence student oral production. Thus in a dialogic task one student's performance is influenced by their partner's engagement (or lack thereof), so a student's poor task performance could be the result of their partner's unwillingness to participate or of their partner dominating the conversation. In addition to the influence of the partner's level of engagement on task performance in dyadic pairs, shared conversation management responsibilities also impact student production, an observation from my module one and two investigations. As Jenks (2007) observes, two-way tasks require a high level of communicative competence and thus a student's attention resources can be overtaxed from trying to manage the flow of a communicative task. This results in a "between contexts" (Batstone 2002: 11) problem, where the social aspects of the task prevent pushing output and maximizing risk taking. For this reason Yalden (1987: 96) suggests focus on sociolinguistic competence and discourse competence should come after students have acquired a basic level of language ability.

To ensure strong individual task performances I established clear interactional roles for the pairs, for example by setting out a one-way interview format similar to Bygate's (2001). Students had two roles; speaker and asker/supporter. The speaker's role was to be the interviewee and talk while the interlocutor was responsible for asking questions and providing support. These were role responsibilities that put pressure on the students to cooperate with each other. This established at the level of discourse what van Lier (1988: 105) calls "rules of turn distribution," in other words making clear who speaks when and about what. The downside is that these clear roles restrict participants' ability to shape the discourse. The process for each task performance thus involved each student taking a turn at each role. They were asked to perform the task in their assigned role for six minutes before switching roles with their partner. Students had their own recorders and microphones to create a record of the conversation.

1.9 Lesson sequences for transcribing and reporting.

This section explains the two lesson treatments used for the study. Rather than talking about the lesson plans, this section primarily describes issues faced teaching the classes and collecting data and should provide a satisfactory description of the actual teaching context in which this research was conducted. Overall one lesson on one theme spanned two classes of two weeks. A fourth and final task performance on the same theme occurred three to four weeks later. In relation to time gaps the second and third repeat task performances occurred exactly one week after the first task performance. The third task performance occurred immediately after the second task performance, and as mentioned the fourth task performance on the same theme took

place three or four weeks (depending on the schedule). Below is the first lesson for each treatment.

Figure 1.0 First class lesson sequences for transcribing and reporting classes

Transcribing class	Willis report class
First Class	**First Class**
1. Unguided pre-task planning (15-20) minutes	1. Unguided pre-task planning (15-20) minutes
2. Do task with partner	2. Do task with partner
3. Transcribe performance	3. Recall and write report as dialogue
4. Self-correct transcript	4. Read report with teacher feedback
5. Edit and reformulate transcript	5. Edit and Reformulate recalled dialogue
6. Answer self-reflection questions and hand in notebook.	6. Answer self-reflection questions and hand in notebook.

1.9.1 The first lesson

Pre-task planning was the same for both groups. A week before the lesson, students were given the set of questions for the upcoming topic. Students were required for homework to make notes under each question. Students did not consistently complete their homework, so 15-20 minutes of unguided pre-task planning was provided in the first of the two classes. Mehnert's (1998) research on pre-task planning suggests pre-task planning over ten minutes does not have a major affect on fluency during task performance, but in my context students feel that less than ten minutes is inadequate to prepare for a task, and in this research students seemed to spend a considerable amount of time simply trying to comprehend the questions. Once the pre-task planning was finished all students had to find a partner from their group who they had not worked with yet and if there was a problem I selected some of the partners. Once in pairs students decided with their partner who would start as the interviewee.

1.9.2 Borrowing

At the trialling stage, I had initially prohibited students from using their pre-task notes once the task started; however, they ignored my request. My situation was thus similar to Weaver's (2007) students where their willingness to communicate in his study was dependent upon them being allowed to look at their notes. Thus, in order to maintain student cooperation and keep the classroom atmosphere positive, I allowed them to consult their notes but not to read them verbatim. The one exception to this was in the fourth performance where students, outside of a ten-minute review prior to the task, were not allowed to keep any notes or papers on their desks. Equally, I tried to discourage students from using dictionaries during the task as I wanted them to rely on one another's linguistic resources rather than engage in what Prabhu (1987: 60) calls "borrowing." One reason why the students value notes and dictionary use is, as Prabhu (1987: 61) argues, "Borrowing is necessary for maintaining task-based activity," adding it probably has some "direct value" for acquisition. As an example, in my context few students feel confident doing a task without the help of support material. While borrowing is generally accepted in my classroom, it created a problem for the research because of the analysis tool I was using, the cognitive based AS-unit, and so student borrowing needed to be taken into account in my coding of my data. In particular I needed to be able to account for reading of notes when it occurred during task performances and exclude it from codifying. In Chapter 2 I describe how I managed the reading of notes by students including how I identified it in task performances.

1.9.3 The first lesson sequence

Individual task performances were timed for roughly six minutes (+/- 30 seconds), based on the amount of time students were able and willing to speak in trials of the lessons. The six minutes was the amount of time students had to do the task and did not equate to six minutes of production. Once the task was finished I instructed the transcribing group to start self-transcribing their own task performance. I told them to leave out false starts, repetitions, off topic language, Japanese, and so forth. With the exception of these features they were told to transcribe their performance as it was produced. Each student had their own digital recorder and microphone. The quality of recorders was excellent and there were not any cases of the students or myself not being able to clearly understand a performance.

While the transcribing group began their transcription work after the task was completed, the reporting group was instructed to try to recall their dialogues by writing them out in their notebooks after passing their recorders to me as the instructor. I also asked them not to reference their pre-task notes, but many regularly did. Again prohibiting students from using their notes was not possible because of the classroom context. The classroom was large enough that having two groups do different work at the same time did not seem to be a problem. The report-writing group always finished before the students doing the more time-consuming transcription work. Once the reporting group finished I then conducted a teacher-directed feedback session with them while the transcription group continued their work. I chose on average three students to report about their task production, with selected students reading their written, recalled dialogues, and as the teacher I

provided real-time corrective feedback and further pushed output. As the reports were mostly general summaries, my work predominately involved pushing student output through asking follow up questions, thus the report sessions were more meaning-focused than form-focused. When I did provide feedback the reporting group listened without taking notes, an issue I have investigated previously (Moser 2005). The end of the reporting phase usually coincided with the transcribing group finishing their transcriptions and initial self-corrections of their first task performances. Before starting the next phase both groups were asked to highlight language on their written transcripts which may correspond with text they read verbatim from their pre-task notes during the task performance.

In the next phase both groups were asked to start reformulating their performances. Roughly 15 minutes was allocated for this activity. During this phase my role involved walking around the class and helping students with problems, although it was rare for the students to ask for help. The first class ended with students having to answer self-reflection questions (see Appendix 2) in their notebooks. The purpose of these questions was to explore student opinions about the lesson and their impressions of their task performance language. To facilitate reflection the reflection questions were in Japanese and students were told they could write their answers in Japanese. These notebooks were submitted at the end of the class.

Once the notebooks were submitted I reviewed both groups' notebooks and listened to each of the task performances. This was done to mine their performances for language problems as well as offer topic expansion ideas to present to the class as a whole at the start of the second lesson. I found the students to be quite accurate in their transcriptions. For the reporting group, I did not give any feedback or comments in

their notebooks as they received feedback during the reporting phase in class. With the transcription group I corrected and reformulated language they missed or could not correct on their own. In addition I provided general suggestions for reformulating and words of encouragement.

1.9.4 The second lesson

Below are the lessons phases for both treatment groups.

Figure 1.1 Second class lesson sequences for transcribing and reporting classes

Second class for transcribing	Second class for reporting group
1. Teacher returns notebooks	1. Teacher returns notebooks
2. Teacher-directed whole class feedback session. Positive and negative evidence of language used plus new native speaker input introduced.	2. Teacher-directed whole class feedback session. Positive and negative evidence of language used plus new native speaker input introduced.
3. Students review transcript and teacher feedback, finish editing and reformulating their transcript.	3. Students review dialogue and finish editing and reformulating.
4. Students do a rehearsal of same task with first partner.	4. Students do a rehearsal of same task with first partner.
5. Students are assigned a new partner and do the same task again.	5. Students are assigned a new partner and do the same task again.
6. Students answer self-reflection questions in their journal and are assigned next week's homework.	6. Students answer self-reflection questions in their journal and are assigned next week's homework.
7. Hand in notebooks. Teacher listens to transcribed second and third performances and gives limited but updated feedback.	7. Hand in notebooks. No feedback provided for second or third performances
Three/Four weeks later	**Three/Four weeks later**
A review session is held and the students do the task again with a new partner from their group	A review session is held and the students do the task again with a new partner from their group

The second class started with 10-15 minutes of teacher-directed whole class feedback with both the reporting and the transcribing groups using positive and negative evidence of student errors and problems. Students rarely asked questions or took notes. After this students first listened to a recording of two native-speakers doing the same task and then were provided with the transcript of the task performance. I used this new input to highlight language and new ways to expand on the main topic. After

this input session was completed, students had 10-15 minutes to reformulate their work from the previous week in the case of the reporting group or to review my feedback in the case of the transcribing group.

In the next phase students repeated the task with the same partner as the first performance to give students an opportunity to improve their task performances and to clarify whether students could improve their performances after the intervention. In Chapter five I discuss this issue in further detail. Requiring students to work with the same partner was intended to ensure their subsequent performances were comparable to their original task language, as using the same partners eliminated the influence a new partner may have had on their performance.

My hope in having students repeat the task was not for them to repeat their first performance verbatim but to rework and build on prior topics to produce an improved performance. Once the second task performance was finished students immediately found a new partner from their group and did the task again. At the end of the second lesson, before submitting their notebooks, students again wrote their reflections based on a set of questions I provided. The basic themes of the second set of questions were the same as the first. Outside of class I again listened to the transcribed performances and provided more feedback and encouragement in their notebooks. This plan was not in the original study, but I considered it necessary to ensure students knew the second lesson performances were also important.

The final phase for each of the four lessons was to have the students do the task again three or four weeks later, primarily because scheduling conflicts prevented using one set interval. The students were told that this final task performance was a review and

not a test, although upon examining the data it is not clear whether the students felt that way. For this final performance students were not supposed to use dictionaries or their notes. As I will discuss in Chapter 6 some students did use their dictionaries but during the task performances all students had to remove their notebooks from the desk. Students were given ten minutes to review their notes prior to the task performance. In order to facilitate production during the task performance the interviewer was provided with the same question sheet used in the prior task performances. This sheet was new and therefore did not have any student notes on it.

In summary the first part of this chapter I reviewed module one and module two in order to explain the research agenda for module three. Module three was designed in response to the poor findings from module two. In the remainder of this chapter I discussed the teaching context for module three and the design of the study. In the next chapter I will discuss the quantitative design of the study.

Chapter 2 Quantitative background

2.0 Introduction

This chapter discusses the quantitative portion of this mixed-methods study, providing a transparent and detailed account of the transcription and codifying process. The first part of this chapter reviews how students' task performances were transcribed, in particular describing which aspects of their production were filtered into or out of the final transcripts. The remainder of the chapter discusses the complexity, accuracy, and fluency (CAF) constructs used to evaluate students' oral task performances, including discussion of how the spoken data was codified into AS-units (analysis of speech units, Foster et al. 2000) and some of the problems encountered in doing this. Foster et al. (2000: 362) in reviewing research using the T-unit, upon which the AS-unit is based, found none of the studies reviewed reported problems applying the unit to their data, leading them to wonder how much data ended up on the "cutting-room floor." This led them to criticize researchers for potentially avoiding important questions about the application of analysis units to spoken data (Foster et al. 2000), a criticism which motivates this chapter's detailed explanation of the coding system used for this research.

2.1 AS-units

Before moving into a discussion of the transcription process, it is important to first define the AS-unit (Foster et al. 2000). AS-units encompass independent clauses, subordinate clauses, and sub-clausal units (Foster et al. 2000). A sub-clausal unit is defined as a minor utterance or any utterance that can be made into a full clause by adding elided parts. One reason the AS-unit was chosen for this research is that it is a popular measure for spoken texts (see for example Tavakoli and Skehan 2005; Lambert and Engler 2007; Michel, Kuiken and Vedder 2007).

2.2 The transcription system

This section explains how student task performances were transcribed, an essentially "reductive exercise" (Swann 2001: 341), as the transcription process requires the researcher to focus on certain aspects of an oral performance while excluding or ignoring other aspects. As Richards (2003: 202) notes, a completely accurate transcript is "an illusion." Because of the inherent incompleteness of transcriptions it is crucial to explain the focus of the transcription system used and what was excluded from the transcripts in this research. Foster et al. (2000) recognize three levels of inclusiveness for transcribing spoken L2 data into AS-units. The first level includes all L2 use while the third level includes only complete AS-units. The level of transcription used here is the second level where back channelling was removed (*oh, okay, uh-huh, really*), and "one-word minor utterances and echoic responses" (Foster, et al. 2000: 370). The production of the student in the supporting role outside of the questions she asked was also excluded.

2.2.1 L1 use during task production

The largest amount of language excluded in the transcripts was L1 use and not reporting that students used their L1 extensively during task performances would be a substantial omission, as students relied quite heavily on Japanese. The abundance of L1 during task performances in this study is not unique in the research; Swain and Lapkin (2000: 268) reported being "taken aback" in a French immersion context by the amount of student L1 use during task work. They identified three interrelated uses of L1 by students: 1) task management, 2) focusing attention on language, and 3) interpersonal interaction (Swain and Lapkin 2000). All three purposes were evident

in the task performances in this study, with Japanese tending to be interwoven with English during task work, at times complicating the codifying of AS-units. Extract 2.0 illustrates how L1 use interspersed the unfiltered transcripts.

Extract 2.0
(L1 is in bold)

13P: It's very cute! Is there something that you want to buy right now?

14L6: /I want many t-shirts **(P: Un un)** and **nanka** cool bag *cool na kanji no*
                                                  something like   cool looking bag
(P: Oh oh) natsu dakara /Because {3.31} it is {hotter more hotter} **(P: Un)**
         because it's summer
hotter and hotter/**(P: Un un nan to naku wakaru)** *atsuku naru*
                Yes I kind of understand   becoming hot

(P: atsuku naru*) atsuku naru nanka* /Cool and light bag/**(P: Ah ii ne)**/
    becoming hot   becoming hot   It's good

15P: Do you often buy things you don't need?

16L6: /Yes {I think} I thought **(P: Un un)** it is (cu) very cute *de*/(I buy) I bought it
                                                  so/then
        but it is too small for me **(P: Ah ah)**/(7+10)

A prominent feature of the task language is the amount of L1 questioner back channelling, which was evident in every task performance. Its absence from the final transcripts hides how interactive the dyads actually were and how interpersonal communication was maintained. L1 was also used for negotiating task procedures and grammatical/lexical searches or problems. In Extract 2.0 there are four examples of L6 using Japanese for grammatical and discourse purposes. In 14L6 she uses it three times and then once more at 16L6. For the purposes of the quantitative AS-unit analysis I removed all the L1 from the transcript except for proper nouns that could be familiar to a non-Japanese English speaker. Lastly, I used L1 as secondary evidence for codifying AS-units. For example, in Extract 2.0, the bulk of the student's Japanese

use at 14L6 occurs where the two AS-units in the turn are demarcated. This is one case where L1 use appears to be a dysfluency acting as a boundary between runs. Using the L1 instances to inform conversational boundaries was a common strategy I used in analysing the transcripts.

2.3 Complexity, accuracy, and fluency (CAF)

This section explains how I adapted CAF to my data. Ellis (2009b), Skehan (2009), Pallotti (2009), Norris and Ortega (2009), and Larsen-Freeman (2009) provide updates on current research into CAF. In recent years several measurements for these three aspects of oral performance have been used in task research (see Ellis and Barkhuizen 2005). Norris and Ortega (2009: 575) warn researchers CAF it is not "some kind of universal construct" that can be blanketed across all contexts without adaptation and remind us that researchers must consider their context and what CAF is supposed to reflect. On this point one of the challenges of using CAF is to adapt the construct to the context in which the research is being conducted to accurately quantify task performances.

2.3.1 Complexity

As in my module two research (Moser 2008) I used the AS-unit (Foster et al. 2000) as the base unit for quantifying oral speech. One shortcoming of the measure is that it can give sub-clausal units or sentence fragments syntactical status, something which some researchers want to avoid (see for example Mochizuki and Ortega 2008), as including these minor speech acts can significantly influence the final quantitative results. To overcome this shortcoming of the measure, in this study I did not count single-word or two-word units in AS-unit totals. Additionally three-word AS-units

functioning as simple responses (*Yes I can, Yes I do, Yes I am*) were generally excluded unless the student's performance was very weak.

For complexity I used the following measures, largely drawn from my module two research (Moser 2008):

- Total words per task performance

- Words per AS-unit

- Total words of AS-units seven words and over

- Number of AS-units seven words and over

- Largest single AS-unit

Each of these is discussed in more detail below.

2.3.1.1 Total words per task performance

The measure 'total words per task performance' reflects student language productivity. Simply, the more words a student produces the more productive they are, which Dörnyei and Kormos (2000) consider a reflection of student task engagement. As the results will show, in this study the more words students produced the better their performances in relation to CAF measures.

2.3.1.2 Words per AS-unit

My module two extensively reviewed key issues for the average number of words per AS-unit (Moser 2008), but to summarize, an increase in words per AS-unit can be achieved in two ways. First, the number of words per AS-unit reflects student ability to build language around a simple clause structure by, for example, adding adjectives, adverbs, prepositional phrases, etc. The second way is through either subordination or

coordination. The problem with words per AS-unit is that there is no way to distinguish between these two basic ways of extending AS-unit length (Norris and Ortega 2009). Thus the main shortcoming of words per AS-unit is that, as with other length measures, it can only provide a global measure for linguistic complexity. Additionally, number of words per AS-unit is somewhat misleading because it suggests language processing is done word by word which is not the case for proficient language users. However, this may be the case for low beginner students such as those included in this research, as their language tends to be rule-based (Foster 2001). This means that using the number of words per AS-unit here is appropriate. However, as I discussed in module two (Moser 2008), and as Foster (2001) also recognizes, speaking involves more than single words strung together; it also includes formulaic language. In the case of the students in this research, their repertoire of formulaic language is very limited and, as Foster (2001: 90) describes it, "idiosyncratic" because it does not always correspond with L1 norms, making it difficult to identify. Despite acknowledging the use of lexical routines by students in my study, I nonetheless counted what appeared to be formulaic language as individual words. Thus one should assume longer AS-units in this study are partially comprised of formulaic chunks.

2.3.1.3 Total words of AS-units over seven words per unit and number of AS-units over seven words

The next two complexity measures are number of AS-units seven words and over and total word count of AS-units seven words and over. Both measures are an attempt to identify how much of a student's task performance was at the upper boundaries of their interlanguage. The idea is from Skehan and Foster's (2005: 198) accuracy

benchmark, which involved discovering a 'cut-off point" in terms of length of clause where students cannot produce correct clauses. In module two (Moser 2008) I explored the idea of a complexity benchmark and analysed at what word count an AS-unit started to reflect challenging production for the students in that study, determining the border was at six- and seven-word AS-units. At the six- and seven-word mark there was an increase in error-filled AS-units and an increase in conjunctions within the AS-units. The same benchmark of seven words was used for this study because the students here were noticeably less proficient than the students from my module two (Moser 2008), making this benchmark quite high for their level of English proficiency. Benchmark in this study then refers to a 'starting point' where I determined students' AS-units to reflect complex language use.

2.3.1.4 Largest single AS-unit

I chose to include the largest single AS-unit as well because with the total word count of AS-units seven words and over, some task performances may have fewer AS-units over the benchmark, but those AS-units may have high word counts, as one AS-unit could comprise up to 30 words. Thus the largest single AS-unit from a task performance was also included as a measure. While it refers to only one AS-unit in a performance, in this study a large AS-unit could represent a large percentage of a speaker's task performance. Examining the data, students appeared to have the highest level of task engagement when producing these long AS-units.

2.3.1.5 Quantifying AS-units

I followed the same procedures as in my module two (Moser 2008) for counting words in AS-units. All words were included minus Japanese words that a non-Japanese speaker would not understand. Inaudible words were not counted. Contractions were counted as two words as were compound words. Words repeated for rhetorical effect or for clarification due to noise were not counted. Dysfluent language was treated as Foster et al. (2000) recommend:

- False starts (placed in brackets and excluded from unit)
- Repetitions (placed in brackets and excluded from unit)
- Self-corrections (include final version in unit)
- Interlocutor interruption/Scaffolding (include if speaker incorporates the word or language into the unit)

Dictionary use and reading notes were excluded from the word count because of the need to have a cognitive quality to the AS-unit. The difference between these two types of borrowing and scaffolding was fairly clear in my data. As Foster et al. (2000) recommend I only counted interlocutor scaffolding if the speaker incorporated the word into the AS-unit, something the students often did. This contrasts with dictionary use where students often simply read the word verbatim from the dictionary without attempting to incorporate the word into the relevant AS-unit. I concluded the reason for the difference was that the speaker was often familiar with the grammar or language suggested by their interlocutor as scaffolding while dictionaries often supplied low frequency words or language well beyond the students' current interlanguage.

2.3.1.6 Problems codifying AS-units

Three main problems presented when codifying student production into AS-units. The first, the same from module two (Moser 2008), centred on applying Foster et al.'s (2000) criteria to my spoken data. The issue was that, as in Foster et al. (2000), I treated the AS-unit as syntactical and gave secondary importance to intonational and hesitational phenomena, with the rule interpreted as attached clauses consisting of syntactical markers, a pause of no more than 0.5 seconds, and no rising or falling intonation. These features are important as Foster et al. (2000: 356) believe they are necessary in order to reflect the psycholinguistic processes of what a student can do in a "single unit of micro-planning." However, Chafe (1980: 14) notes that all three are not necessarily present at unit boundaries and the presence of one all of them does not consistently signal a boundary, adding it would be "surprising" if one or all three together would be consistently present. Besides the three features above, L1 use, interlocutor interjection, and borrowing from notes, dictionaries, and partners also played a role in codifying AS-units.

One issue was that students, perhaps because of their low level of English proficiency, produced highly variable intonation and hesitation patterns, making intonational cues unreliable. Additionally, students relied heavily on an add-on strategy for speaking, or clausal chaining, which can be prevalent in open-ended theme-based tasks where unfolding real-time conversation often accommodates using this strategy. This made it challenging to determine which clauses should be codified together, a difficultly compounded by the fact that conjunctions serve multiple functions in speech. Both coordinators and subordinators, besides being grammatical, can also serve discourse functions as sentence starters or to end sentences. In addition the presence of subordinators does not necessarily mean clausal embedding is present. Carter and

McCarthy (2006) note that in conversation, subordinators like *because* and *so* function more like coordinating conjunctions. This meant grammatical cues were often unreliable in delineating boundaries between AS-units when students used clausal chaining.

The second, related problem, centred around difficulty actualizing Foster et al.'s (2000) suggestion that an AS-unit should reflect a single psycholinguistically valid micro planning activity, which the nature of the students' production challenged. It is difficult to treat AS-units using this rule with the data gathered for this research because of dysfluencies between fluent spurts of speech. Thus I argue fluent speech is a different cognitive activity from pausing or producing dysfluent language. In analysing the transcripts I decided length of run (a flow of speech between dysfluent language, pausing, and L1) reflected automated speech while pausing or dysfluent language reflected attentional focus on language. As well, Japanese interwoven with English indicates deliberate attentional focus on language. Extracts 2.1 and 2.2 provide examples of the problem as it presented itself in the data and demonstrate how the distinction between automated and attentional focused language was codified in the transcripts.

Extract 2.1

12L5: /It's Mystic/{Mystic is Mystic}[Its clothes is very expensive but]{3.56 they}
[they are very cute and]{many 5.09}[many kind of clothes](3+15)

Extract 2.2

6L3:/[I also go shopping to Umeda because]{there there ar}[it has many shops and]{it isn't it isn't not it isn't far from not so far it is not so far no so far}[is not so far (my) from my house]/(19)

Both extracts are from two different students and represent one turn each. In Extract 2.1 there are two AS-units (3words + 15words), and in the second extract there is only one AS-unit (19 words). The slashes demarcate AS-unit boundaries. The fifteen-word and nineteen-word AS-units are both comprised of multiple clauses/subclauses that are chained together. The first AS-unit in Extract 2.1 was easy to codify and distinguish from the second as it is syntactically separate and intonation and pausing isolate it from the second AS-unit, making applying Foster et al.'s (2000) criteria straightforward. In the second AS-unit the square brackets [ ] enclose a single length of run and the parentheses enclose { } dysfluent language. In the first extract there are two large pauses. The second extract contains dysfluent language. With both extracts there is a 'start and stop' quality to them. Essentially, both extracts demonstrate that after producing a spurt of fluent speech the students were forced to stop (breakdown fluency), or they deliberately stopped their speech (repair fluency) to plan their next spurt. Besides being arguably different types of processing than fluent speech based on the degree of attention required or present, the time during 'stop' periods is well beyond Foster et al.'s criteria for considering clauses or subclauses to be part of one AS-unit. If I had followed Foster et al.'s (2000) criteria then in many cases, as illustrated in Extracts 2.1 and 2.2, their intention in proposing the criteria would have been distorted.

Because of the difficulties outlined above, I moved from viewing the AS-unit as a cognitive-based micro planning activity to thinking about it as spurts of speech during a turn combined with dysfluent language, including L1 use, often reflecting macro

planning. Bygate (2005: 111) writes that when speakers start phrases or clauses they often know the semantic content and grammar which will follow, adding it is almost impossible for someone to start a phrase or clause without knowing "what is going into the later parts" of it. In Extract 2.1 and 2.2, despite the substantial dysfluencies, the students did eventually manage to produce well-constructed utterances. This overall coherence suggests they roughly knew what they wanted to say and substantial reformulation work, as in Extract 2.2, is evidence of this. Thus I concluded AS-units should represent predominately macro planning events, as students micro plan in increments but also tend to have a general plan of what they want to say and the language needed to say it. This interpretation of AS-units is supported by Chafe (1980) and Garman (1990), who both suggest speech contains both micro and macro planning. Thus in the end I kept Foster et al.'s (2000) criteria for codifying AS-units but de-emphasized the idea that the AS-unit is strictly a cognitive-based micro planning activity because my data did not seem to support this idea; there was too much evidence that a macro perspective of student production was more accurate.

2.4 Accounting for reading of notes

Due to the cognitive quality of the AS-unit the third issue involved students reading their notes during their first, second, third task performances. Students were forbidden to have access to their notes during the fourth task performance. As discussed previously, students' willingness to take part in task work often partially depends upon them having access to such notes and being able to borrow from them. The only option I had in this study was to try and wean the students off of reading their notes. Regardless of whether they used their notes or not, for my research it was crucial that I had records or identified in their task work production parts that were

read. In the case of keeping records students after each task performance were given a special coloured pen to underline parts that they were sure they had read. They could underline language in their reports, transcripts, and notes. Time was allocated for this, and I circulated around the class to check that students were doing this properly. During the task performance as well I made notes of students who I had identified as reading notes, and then checked while they were reviewing their performances for reading to confirm that had underlined parts that they had read. Students that did not underline anything I would often ask them if that was in fact the case. It was very rare for students who had read their notes not to underline anything. Transcribing students of course had access to their first performance recordings. To facilitate honesty, I told students that in no way would the reading of notes affect their final grade, unless they consistently throughout the semester just read without making any effort to speak at all. I told the students that since the class was an oral communicative class that they should try not to read their notes. For many students not reading or looking at their notes became a major goal for them during the course. On that point the reading of notes did decrease as students repeated tasks over a task session, and as we progressed through the semester long course.

2.4.1 Identifying readers and reading

While it was always a challenge throughout the study to try and discourage students from reading, in addition to having them properly acknowledge what they read; nonetheless, the actual identifying of parts of their recorded spoken production that they may have read and failed to acknowledge was relatively straightforward. First, I found the students to be very honest and quite accurate in acknowledging what they read. Since students were aware that reading did not impact their grades, and were

regularly reminded of it, there was no reason to mislead the teacher.   As I will show in Chapter 5 students were generally conservative when it came to giving oneself credit for speaking without borrowing. By this I mean that students would almost always underline a portion of language when they were unsure whether they had read it or not. Again, there was no incentive for them to suggest otherwise.

When students did not underline language they had read there was usually a combination of clues that helped me identify it as read. First, it should be pointed out that I knew who were the 'habitual readers', and, moreover, they were quite open about their tendency to read. This was a very small group of students. These students would often acknowledge reading their whole task performances. Second, as I will reveal later the largest portion of spoken production in task performances emerged from unplanned and unpredictable follow-up questions, which obviously made student notes irrelevant. In contrast I paid special attention to production that was derived from the questions that I had made for the tasks.

In relation to clues of reading that were recorded in the task performance one was the shuffling of paper when students were looking through their notes or holding their notes while they read. Students L1 use was also a clue or indicator. Students would often tell their partners in the L1 they needed to reference their notes. The habitual readers would often just tell their partner they were going to read parts because they could not say them in English otherwise.   Most of the L1 was not directed at their interlocutor per se, but rather private speech.   Another indicator was when a student's speech suddenly became substantially fluent for a short extended turn with no dysfluency present. This burst of speech was almost always traceable back to notes.

In addition it often sounded like quickly read notes. On this point the most common indicator was speech that clearly sounded like it was read, because it was usuall monotone and lacked any intonation. On top of this it was often slow and choppy.

When I did find read parts that were not underlined, as with most borrowing I excluded it from my data. If I was unsure whether it was read or spoken I always excluded it so as not to compromise my data. Overall I believe that students accurately reported what they read, and if they did not, then I was in most cases successfully able to identify substantial portions of read speech that would have distorted the numerical results of a student's task performance. Generally, when students read it was usually in the form of a monologue which was relatively easy to identify because of the indicators explained above. In addition it is also very difficult for students to mix speech and reading during a real-time task performance without providing fairly obvious clues and markers. The difficulty of it I believe dissuaded many students from try to do it. Generally, I feel that the reading issue was an important problem to be tackled in my study, but I believe that the way I managed it during class; and, furthermore, accounted for it in my raw data ensured that it did not compromise the quantitative results.

2.5 Accuracy

For accuracy I used four measures.

- Number of reformulations

- Error-free AS-units (as a percentage of total AS-units)

- Total number of error-free AS-units of seven words and over

- Total number of words from error-free AS-units of seven words and over

The measure number of reformulations per task performance indicates an orientation on the part of the student to speak accurately, with reformulations defined as changes to syntax, word order, and morphology as well as replacements or lexical corrections/substitutions. While lexical replacements may not reflect attention to form, considering the level of the students, I decided to include any orientation to self-correct or improve performance as an attempt by the students to be more accurate. In my data there were fewer lexical replacements than grammatical reformulations. Lastly I did not differentiate between whether the self-correction was correct or incorrect in counting reformulations.

2.5.1 Error-free AS-units

Tavakoli and Skehan (2005: 256) argue that while general measures like error-free AS-units are more "blunt" than specific measures, they nonetheless capture variance in performance more successfully than specific measures. Because of the beginner level of the students in this study I chose error-free AS-units rather than Bygate's (2001) errors per unit. During the codifying of the data it became clear that when students were being productive and engaging in risking taking by producing longer AS-units that these AS-units almost always contained multiple errors. Based on a preliminary analysis of the data I concluded that if I used this measure that incidence of error per AS-unit would increase with more productive task performances. This is consistent with Skehan's (1998) Trade-off Hypothesis where accuracy is compromised at the expense of fluency and complexity. In my study this measure would reflect student risk taking more than students being accurate in their production. I concluded that error-free AS-units would provide a clearer indication of student accuracy because it would not be overly influenced by productivity like errors

per unit. In summary I assumed almost every student in this study could not be accurate while they were risking taking or output pushing. I was interested if during such productive task performances if some level of accuracy could be maintained or improved on. In this sense I associate error-free AS-units with careful production in contrast to incidence of error per AS-unit which I associate with risk taking. In order to provide a finer measure of error-free AS-units in a task performance I used an accuracy benchmark. For the next two measures I counted the number of error-free AS-units of seven words and over and their total word count, hoping these two measures would provide a finer analysis of how accurate the best part of a student's performance was.

2.5.2 Identifying errors

In this study I considered errors to be errors in syntax, word order, morphology, and lexical choice using native-like accuracy as a comparison model. Like Bygate (2001) I took a conservative approach to accuracy. While Holliday (2005) is critical of relying on native-speaker models for language teaching and evaluation, I agree with Harmer (2007) who notes that even amongst foreign language students accurate language is generally considered more important than other performance aspects. My reasons for using a native-speaker standard were also for more practical purposes. First, an issue arose over whether or not to count articles and other grammatical words that are difficult for beginner students. I concluded that to start excluding certain grammatical words is problematic because if one type of error is excluded then why not another. Second, my study was partially investigating the assumption that transcription combined with task repetition is thought to allow students to improve on

errors that are often resistant to teaching as it allows students to better see typically non-salient language errors.

In codifying errors Lynch (2007) used Corder's (1967) distinction between slips (accidental mistakes), or what Loewen (2007: 102) refers to as "performance errors," and errors (systemically incorrect elements of interlanguage). Lynch (2007) contends that students can notice and correct mistakes but will miss errors. This seems reasonable but ignores the declarative/procedural dimension of language in that what students may not be able to produce or produce consistently they may be able to self-correct when it is presented in a transcript. Considering my students' English proficiency and the unstable nature of their interlanguage, Loewen's (2007: 102) term "performance errors" seemed appropriate for errors occurring during a task performance that may not be systemic. Loewen (2007) rightly comments that a performance error reflects a student's difficulty with the linguistic item that teachers will need to address, thus in this study I included performance errors as errors, meaning that anything that clearly did not look like a slip of the tongue was counted as an error. For codifying errors, any self-corrections students made in their transcriptions did not negate the initial performance errors; I still counted them.

2.6 Fluency

Fluency as a performance aspect in oral production can be categorized into three sub-dimensions; speed, repair, and breakdown (see Tavakoli and Skehan 2005). Below are the measurements used for each sub-dimension:

Speed

- Length of run

Repair

- Total number of dysfluencies (including verbatim and substitutive)

Breakdown

- Number of pauses

In my study length of run is a stream of speech measured in words sandwiched between clear pauses, dysfluent language, and L1 (Freed 2000) and not syllables per run as that measure was not appropriate for the data gathered, particularly since the students seemed to have trouble isolating individual English phonemes and thus unnecessarily created a number of extra syllables. Thus I was more interested in how many words students could produce in one spurt rather than the speed of that spurt. However, length of run presented one codifying issue, as initially two types emerged from the data. One was word by word where single words were divided by dysfluency boundaries and the other was what length of run is normally categorized as; spurts of words strung together. As student production during the task performances contained both, only counting length of run in AS-units of seven words and over helped to resolve this problem. In particular, I counted length of run in AS-units that had at least two clauses or subclauses and in the case of subclauses they had to be clearly elided. In summary length of run reflects the number of words a student can produce in one spurt of seven words or more which contains at least two potential full clauses. This reflects how fluent the best part of a student's task performance is. With some performances I did not codify any length of runs because they had no AS-units that met the criteria. In this case the length of run was the same as the number of words per AS-unit, which was below seven words in almost all instances.

For repair fluency I decided on total number of dysfluent words, including verbatim repetition (false starts, repetitions) and substitutive repetition (reformulations, replacements) (see Bygate 1996 for an explanation of these terms), which represent hesitation phenomena and reflect "moment by moment" production decisions by students (Foster and Skehan 1999: 229). I anticipated dysfluent language would be common in the data and thus potentially useful for detecting differences in performances. I counted both types of dysfluencies together because both are evidence of the student thinking and trying to act on their production, a concept explored in chapter four. In this study in almost all instances where students had productive task performances those performances included a considerable amount of dysfluent language.

The final measure of fluency was number of pauses (filled and unfilled), where pausing can either occur mid-clause or at the boundary of clauses. The latter is seen as a better indicator of fluent speech than the former (Freed 2000), as L1 speakers generally pause at boundaries rather than mid-speech, although for this study they were both counted together. In my data mid-clause pauses generally resulted in AS-units with small word counts which were almost always under the benchmark of seven words. I did not count pausing at AS-unit boundaries, pausing between AS-units, and only counted pausing once the student had clearly started to speak. Thus I only counted pausing that occurred within AS-unit boundaries.

Studies on pausing have debated the minimum length a pause should be before counting it as a dysfluency. Some studies use micropauses (see Kormos and Denes 2004), while Foster and Skehan (1999) used one second and over, although both

measures seemed inappropriate for my context. In Japan, Clennell (1999) found a five to ten second pause before continuing to speak was acceptable. Considering my students, the conversational nature of the task type, the relatively high tolerance Japanese have for pausing, and the recommendations of the literate, I decided to count pauses three seconds and over within an AS-unit boundary as dysfluent.

In summary in this chapter I presented the complexity, accuracy and fluency (CAF) constructs used to measure student oral task performance. In addition I discussed the codifying issues I encountered and how the data was transcribed.

Chapter 3 Quantitative results

3.0 Explanation of three-way ANOVA on repeated measures data

This section briefly explains the statistical tests used for this study. The sample size for this study was (n = 119). For example in task session one 13 students did the reporting treatment while 14 students did the transcribing treatment.

Table 3.0 Participants across treatments and task sessions

Task Session	Reporting	Transcribing	Total
1	13	14	27
2	12	18	30
3	15	18	33
4	14	15	29

Analysing student performances across the different groups, tasks, and task repetitions involved a variety of tests, which necessitates providing information on the functions of each. For the quantitative analysis a statistical advisor[1] was consulted to assist in designing the study and resulting analyses. This study used a three-way ANOVA with two factors varied between-subjects (transcribing versus reporting, and four task sessions) and one within-subject factor (four performances of each of the task sessions). Thus there were four different task sessions and each student performed those four tasks four times, for a total of 16 iterations of the open task type selected for this research. Every student did all four task sessions in the same order. The study as discussed previously was a pseudo between-subjects because all students did both treatments. Throughout this chapter the term task session will be used to refer to the four different task topics used in this research. In each task session there are four task performances. They are referred to as a first task performance, second task

---

1

　　　Dr. Allan White, The University of Birmingham

performance and so forth. In the case of examining within-subjects variables (four performances of each of the task sessions), an average of the measure from all four 'task sessions' will be provided, so the first task performance repetition numbers will represent an average of the first task performances of students on each of the four different task sessions; health and exercise, pets and animals, shopping and fashion, and dating and marriage; the second task performance repetition numbers their second task performance of the four task sessions, etc.

For repeated measures designs such as this one, certain assumptions have to be maintained, including "the homogeneity of variance of the differences between samples (sphericity)" (Hinton, et al. 2004: 152). In other words, the relationship across conditions and subjects must be roughly equal, or as Hinton, et al. (2004: 373) write, "Essentially we must assume the effect of an independent variable to be consistent across both conditions and subjects in these designs for the analysis to be appropriate" because there is the possibility subjects will have different responses to the treatment across conditions, such as the different task topics influencing student production. To confirm equal variance, Mauchly's test of sphericity was used; if results show a significant difference, or $p \leq .05$, then sphericity is violated. Conversely, if the test is non-significant, or $p > .05$, then equal variance can be assumed. When there was a sphericity violation the correction tests Greenhouse-Geisser and Huynh-Feldt were used to generate a corrected variance ratio (F-ratio). The statistical software used, SPSS 17.0, automatically provided these correction tests, which supply, "a correction factor that is applied to the degrees of freedom used to assess the observed F-ratio" (Field, 2009: 461). The statistical advisor also recommended including the less conservative Huynh-Feldt correction. In order to

provide a clearer picture of what this information means for analysis of the results of this research, I describe the process in greater detail in the next section.

3.1 Checking for sphericity and determining significance

Table 3.0 shows Mauchly's test of sphericity is non-significant (p = .225) for number of words per task performance. Sphericity can therefore be assumed and the corrections tests do not need to be applied to the data

Table 3.1 Mauchly's test of sphericity for number of words per task performance

Within subjects effect	Mauchly's W	Approx. Chi-square	Df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Number of words	.939	6.941	5	.225	.957	1.000	.333

The next step is to confirm whether or not the results for the measure are significant with sphericity assumed in within-subjects (across task performances) tests of probability using the ANOVA, the results of which are included in Table 3.2.

Table 3.2 Test of within subject effects for number of words per task performance

Source		Type III Sum of squares	Df	Mean square	F	Sig.
1. Number of words per task performance	ANOVA, sphericity assumed	40035.922	3	13345.307	27.252	≤ .001
	Greenhouse-Geisser	40035.922	2.870	13947.650	27.252	≤ .001
	Huynh-Feldt	40035.922	3.000	13345.307	27.252	≤ .001
	Lower-bound	40035.922	1.000	40035.922	27.252	≤ .001
2. Number of words per task performance * treat	ANOVA, sphericity assumed	615.395	3	205.132	.419	.740
	Greenhouse-Geisser	615.395	2.870	214.390	.419	.731
	Huynh-Feldt	615.395	3.000	205.132	.419	.740
	Lower-bound	615.395	1.000	615.395	.419	.519
3. Number of words per task performance * task	ANOVA, sphericity assumed	4053.110	9	450.346	.920	.508
	Greenhouse-Geisser	4053.110	8.611	470.672	.920	.505
	Huynh-Feldt	4053.110	9.000	450.346	.920	.508
	Lower-bound	4053.110	3.000	1351.037	.920	.434
4. Number of words per task performance* treat * task	ANOVA, sphericity assumed	3156.918	9	350.769	.716	.694

Source		Type III Sum of squares	Df	Mean square	F	Sig.
	Greenhouse-Geisser	3156.918	8.611	366.601	.716	.688
	Huynh-Feldt	3156.918	9.000	350.769	.716	.694
	Lower-bound	3156.918	3.000	1052.306	.716	.544

The column on the far left shows the measure number of words per task performance plus the different possible interactions. Treatment refers to the group doing transcription work versus the group reporting on their task performance. Only the ANOVA for number of words per task performance was significant ($p \leq .001$) with no significant results for any interaction effect. The next step reviews the polynomial trends to determine whether they are linear or quadratic. According to Field (2009: 372) polynomial trends should be checked when it is necessary to "order the categories of the independent variable," which in this case is the four task performances. Hinton et al. (2004: 158) notes polynomial trends are used to confirm the "means of the conditions" and whether they fit a trend, which for this research can be either linear or quadratic. A linear trend indicates a straight line, rising or falling, across the different task repetitions while a quadratic effect indicates a rising or falling curve. Table 3.3 shows the polynomial trend results for number of words per task performance and reveals a significant effect for linear and quadratic trends. The other interactions are irrelevant as the ANOVA indicated they were non-significant.

Table 3.3 Polynomial trends: number of words per task performance

Source	Complexity	Type III Sum of squares	Df	Mean square	F	Sig.
Number of words per task repetition	Linear	30509.655	1	30509.655	66.741	≤ .001
	Quadratic	6426.213	1	6426.213	11.348	.001
	Cubic	3100.054	1	3100.054	6.955	.010
Number of words per task repetition * treat	Linear	33.348	1	33.348	.073	.788
	Quadratic	567.741	1	567.741	1.003	.319
	Cubic	14.306	1	14.306	.032	.858
Number of words per task repetition* task	Linear	2989.163	3	996.388	2.180	.094
	Quadratic	635.985	3	211.995	.374	.772
	Cubic	427.962	3	142.654	.320	.811
Number of words per task performance * treat * task	Linear	426.909	3	142.303	.311	.817
	Quadratic	1389.246	3	463.082	.818	.487
	Cubic	1340.764	3	446.921	1.003	.394

Once this final test is checked the next step is to confirm the means for the measure.

3.2 Explanation of the presentation of the quantitative data

The following sections report the statistical results for the CAF measures. They will be reported in the following order: complexity, fluency, and accuracy. As an example, for complexity, I will first provide the transcription versus reporting results for all the individual complexity measures. This is done because all the results are the same as is the case with all fluency and accuracy measures. Next I will provide the task performance results for each individual measure. For each individual measure I will present any task or interaction effect. This means that for example with the measure words per AS-unit I will report and discuss the task repetition result including any task or interaction effects before moving to the next complexity measure. At the end of each of the three general aspects of production (CAF) I will provide a summary as well. I will conclude Chapter 3 by discussing the results in relation to my quantitative research questions. Because of the large amount of statistical data, Mauchly's test of sphericity, ANOVA tests of within-subjects effects (across task sessions with sphericity assumed, Huynh-Feldt), and tests of within-subjects contrasts (polynomial trends across task sessions) are not included here and can instead be found in the accompanying CD ROM. For each measure I report the results of these three tests.

3.3 Transcription versus reporting: complexity

Complexity was measured using the five measures described in section 2.3. Tables 3.4 to 3.8 present the results for comparing the transcription and reporting groups using these different measures. None of the measures reached significance, indicating neither transcribing nor reporting significantly influenced any of the complexity measures used.

Table 3.4 Transcription versus reporting: number of words per task performance

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	3.680E6	1	3.680E6	1634.016	≤ .001
Treat	1800.338	1	1800.338	.799	.373
Task	7078.088	3	2359.363	1.048	.374
Treat * task	6694.528	3	2231.509	.991	.400
Error	249966.475	111	2251.950		

Table 3.5 Transcription versus reporting: number of words per AS-unit

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	21909.219	1	21909.219	3630.176	≤ .001
Treat	.084	1	.084	.014	.906
Task	11.060	3	3.687	.611	.609
Treat * task	3.105	3	1.035	.171	.915
Error	669.919	111	6.035		

Table 3.6 Transcription versus reporting: number of AS-units seven words and over

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	12020.524	1	12020.524	952.367	$\leq .001$
Treat	5.393	1	5.393	.427	.515
Task	17.279	3	5.760	.456	.713
Treat * task	21.268	3	7.089	.562	.641
Error	1401.013	111	12.622		

Table 3.7 Transcription versus reporting: total words of AS-units seven words and over

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	1.270E6	1	1.270E6	601.937	$\leq .001$
Treat	538.625	1	538.625	.255	.614
Task	2963.660	3	987.887	.468	.705
Treat * task	2258.374	3	752.791	.357	.784
Error	234180.854	111	2109.737		

Table 3.8 Transcription versus reporting: largest single AS-unit per task performance

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	95046.695	1	95046.695	1438.675	≤ .001
Treat	.029	1	.029	.000	.983
Task	62.219	3	20.740	.314	.815
treat * task	39.811	3	13.270	.201	.896
Error	7333.261	111	66.065		

These results show that neither transcribing nor reporting had a significant effect on any of the complexity measures.

3.4 Complexity across task performances

When examining whether students' language complexity changed across the different task performances, there were significant results for all of the measures used. When examining for interactions, the only significant result was for number of words per task performance. Table 3.9 shows the means for words per task performance.

Table 3.9 Number of words per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	75.322	2.544	70.281	80.364
2	85.356	2.984	79.444	91.269
3	99.491	3.015	93.517	105.465
4	94.695	2.715	89.315	100.076

Table 3.9 reveals a gradual increase in total words spoken in task repetitions from the initial task performance task to the third task performance, with a slight drop in production on the fourth task performance, indicating students were most productive in the third task performance with an average increase of 24 words relative to the first task performance The confidence intervals reveal a number of significant pairwise comparisons as well. There were significant differences between task one and task three and task one and task four. The latter tasks performances have significantly higher means than the first performance. Next there was a significant difference between tasks two and three, again with the third performance being higher. The overall results for this measure suggest that the latter performances were stronger than the earlier performances.

3.4.1 Number of words per AS-unit

Number of words per AS-unit follows the same pattern as the previous measure with Mauchly's test of sphericity non-significant (p = .855) and the accompanying ANOVA significant (p ≤ .001, F = 9.554) with polynomial trends showing a significant quadratic effect (p ≤ .001, F = 16.250).

Table 3.10 Number of words per AS-unit per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	6.418	.135	6.150	6.686
2	6.957	.148	6.663	7.250
3	7.178	.155	6.871	7.484
4	6.830	.140	6.553	7.107

There is a significant pairwise comparison between the first and third task performances. The first performance was the lowest at 6.4 words per AS-unit, and the third the highest at 7.1.

3.4.2 Interaction between number of words per AS-unit and task session

The number of words per AS-unit also showed significant differences between task sessions ($p = .024$), meaning this measure depended on the task session and vice versa, with the ANOVA test showing significance ($p = .024$, $F = 2.164$) and a significant quadratic effect ($p = .009$, $F = 4.039$).

Table 3.11 Number of words per AS-unit and task session interaction

Task session	Task performance	Mean	Std. error	95% Confidence interval	
				Lower bound	Upper bound
1	1	6.837	.282	6.279	7.394
	2	6.917	.308	6.306	7.528
	3	7.161	.322	6.524	7.798
	4	7.136	.291	6.559	7.713
2	1	6.325	.272	5.785	6.865
	2	6.558	.298	5.967	7.150

	3	7.024	.311	6.407	7.640
	4	6.737	.282	6.179	7.296
3	1	6.456	.256	5.950	6.963
	2	7.011	.280	6.456	7.565
	3	6.814	.292	6.236	7.393
	4	6.616	.264	6.092	7.140
4	1	6.053	.272	5.514	6.591
	2	7.341	.298	6.752	7.931
	3	7.711	.310	7.096	8.326
	4	6.831	.281	6.275	7.388

The salient feature for the interaction between task session and number of words per AS-unit appears to occur in the fourth task session. There are significant pairwise comparisons in the fourth session between task one and task two, and then task one and task three are significant as well. The first performance in the fourth session started has the lowest value than the other first task performances in the other three task sessions. In addition the third performance in the same session has the highest value.

The two lowest first performances were from the fourth task session and the second task session. Both were comparable in how many students were below the seven-word benchmark. After reviewing the raw data I found that in the first performance in the fourth task session 20 students out of 29 had word averages per AS-unit below seven words. In addition seven students had a mean average of number of words per AS-unit below five words. In the first performance in the second task session it was 21

students of 30 who had an average below seven words.   In the first performance of the fourth task session only three students had a mean of eight words and over, but by the second performance it jumped to 11 students, and then to 13 students by the third performance. The second and third performances in the fourth task session were the best performances in all task sessions in regards to students producing AS-units seven words and over. The third highest number of students was 10 out of 27 students in the first task session's third performance.

3.4.2.1 The influence of task session on complexity: accounting for the fourth task session

In this study there were a number of task session effects on performance. In this case it is an interaction between task session (dating and marriage) and number of words per AS-unit. Because of the small groups used in this study I am cautious about making generalizations about task effects and interactions, but my qualitative understanding of the students' task performances supports the interpretation that the results in the fourth task session which produced a topic effect interaction is the result of topic difficulty. Table 3.12 illustrates this as, at the end of the course, students were asked to rank the four topics in order of difficulty. A majority of students ranked dating and marriage (fourth session) as the most difficult task.

Table 3.12 Difficulty ranking of task topics by students

Sequence in course	Difficulty ranking
1. Health and exercise	1. Dating and marriage
2. Pets and animals	2. Pets and animals
3. Shopping and fashion	3. Health and exercise
4. Dating and marriage	4. Shopping and fashion

Shopping and fashion and health and exercise were probably ranked as least difficult because they were familiar to the students and so they were comfortable with the English necessary to discuss them, as those topics are typical of communicative English language syllabi in Japan and students had covered both in their first year English classes. In contrast, the other two topics, pets and animals and dating and marriage were unfamiliar. According to student verbal feedback, they had not previously discussed either topic in an English class.

Dating and marriage, besides being unfamiliar, was also considered controversial by many students, with a number questioning whether this topic was too private to be discussed in class. In task recordings some students jokingly commented certain questions were private. In the classroom there was obvious awkwardness with the topic, but once the lesson started and students gained familiarity with the theme and navigating privacy issues, they appeared to enjoy themselves; lesson journal feedback and task recordings showed the majority of students enjoyed the task. There was lots of laughter and generally performances were lively, indicating the poor first performances were the result of unfamiliarity with the topic and its controversial nature. However, once students became familiar with the theme, the controversial topic appeared to have a positive and motivating influence on student performances relative to the other tasks, which lends support to Cook's (2000: 158) suggestion of a connection between controversial topics and language play.

3.4.3 The final three complexity measures

The final three measures (number of AS-units seven words and over, total words of

AS-units seven words and over, and largest single AS-unit) all produced similar results. All three results comparing the transcription and reporting groups were non-significant. For results across task sessions, there were significant main effects for the final three complexity measures but no interaction effect. For number of AS-units seven words and over, Mauchly's test of sphericity was on the edge of significance (p = .050), so I referred to the Huynh-Feldt correction test, which indicated significance (p ≤ .001, F = 26.117) and a significant quadratic effect (p ≤ .001, F = 17.373). Table 3.13 includes the means for the measure.

Table 3.13 Number of AS-units seven words and over per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	3.877	.186	3.508	4.246
2	4.900	.219	4.467	5.334
3	5.994	.253	5.492	6.496
4	5.511	.239	5.038	5.984

The third and fourth task repetitions had the highest number of AS-units seven words and over the first the lowest, with the third task performance on average 2.1 AS-units greater than the first and the fourth 1.7 AS-units more. The pairwise comparisons reveal a number of significant differences. Task performances two, three and four are significantly higher than the first. In addition there is a significant different between the second and third performance with the latter being higher than the former.

For the second measure, total words of AS-units seven words and over, Mauchly's test of sphericity was significant (p = .045) and the Huynh-Feldt correction test was

significant (p ≤ .001, F = 26.357). There was also a significant quadratic effect (p ≤ .001, F = 21.989).

Table 3.14 Total words of AS-units seven words and over per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	38.594	2.256	34.123	43.065
2	50.841	2.785	45.323	56.358
3	62.861	3.062	56.794	68.928
4	56.175	2.799	50.628	61.722

With total words of AS-units seven words and over the third and fourth task performances again had the highest values with the first performance the lowest. The pairwise comparisons reveal that there were significant differences between the task performances. Task one is significantly lower than the other three task performances. In addition task performance three's pairwise results are significantly greater than task performance two's.

It is useful from a descriptive standpoint to look at this measure in comparison to means for number of words per performance. To do this, the mean number of words per AS-units of seven words or more (x) versus the total number of words per task performance (y) is summarised as x/y below:

Task repetition 1: 39/75   (52%)
Task repetition 2: 51/85   (60%)
Task repetition 3: 63/100 (63%)
Task repetition 4: 56/95   (59%)

The third task performance had on average 24 more words belonging to the complex part of performances than in the first task performance. All the other task performances had more words in the complex part than the first task performance. In addition they comprised a higher percentage of the overall word total for each task performance when compared to the first performance. Perhaps most tellingly, the complex portion of the third performance was only slightly lower (12 words) than the total words for the first task performance.

The final measure, largest single AS-unit per performance, had a significant Mauchly's test of sphericity (p =.033). The Huynh-Feldt correction test was significant (p ≤ .001, F = 11.352) and there was a significant quadratic trend (p ≤ .001, F = 10.584).

Table 3.15 Largest single AS-unit per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	12.605	.452	11.709	13.500
2	14.059	.470	13.127	14.990
3	15.746	.551	14.655	16.837
4	14.623	.537	13.560	15.687

The smallest AS-unit in this measure was in the first performance and was close to 13 words. In almost all cases thirteen-word AS-units would have at least two integrated clauses. The pairwise comparisons reveal that the first performance was significantly lower than both the third and fourth performances.

3.4.4 Summarizing complexity measures

In summary, with respect to complexity, there were no significant differences between the transcription and reporting groups but there were significant results for all students from both the transcription and reporting groups across task performances with the lowest complexity in the first task performance. The means for all the complexity measures suggest the third performance was the most complex. In addition, four of the five measures suggest that the fourth performance was more complex than the second. The measure number of words per AS-unit, which indirectly measures linguistic complexity, reveals only slight differences between repetitions. This suggests that, despite the increase in word quantity, there was not much change in complexity between repetitions. Equally, the largest single AS-unit shows only a slight difference between the first and third repetitions, which could be interpreted as indicating little change in linguistic complexity. The other three measures, number of words per performance, number of AS-units seven words and over, and total words of AS-units seven words and over, indicate a difference between the first performance and the other three repetitions. The final two measures, number of AS-units seven words and over, total words of AS-units seven words and over, and largest single AS-unit per task performance, indicate that with the increase in quantity there was a corresponding increase in complexity. Specifically, there were more words/units that were over my benchmark in the latter three performances than the first. One potential confounding variable for these findings is that task two performance involved students working with the same interlocutor as the first task performance while the latter two performances involved a new interlocutor each time. While I do argue in section 5.6 and Chapter 7 that interlocutors did factor in the results between the

second and latter task performances the results where the first task performance is less than both the second performance with same interlocutor and the third performance with a new interlocutor suggest that the results are probably not attributable to this variable. Finally, the pairwise comparisons revealed a significant difference between some of the task performances. First on every measure the first task performance was significantly lower than the third task performance. In addition with exception of the measure number of words per AS-unit, the first task performance was significantly lower than the fourth task performance. What's more the second task performance on three of the five measures was significantly weaker than the third task performance. These significant pairwise comparison results provide evidence that the latter task performances, in particular the third task performance was stronger than the task performances that occurred before it.

3.5 Results for fluency

3.5.1 Transcription versus reporting results

For the fluency measures (length of run, number of dysfluent words, and number of pauses) there were no main effect between the transcription and reporting groups. Results are presented in Tables 3.16 to 3.18.

Table 3.16 Transcription versus reporting: length of run

Source	Type III Sum of squares	Df	Mean square	F	Sig.
Intercept	18795.546	1	18795.546	5351.536	≤ .001
Treat	1.210	1	1.210	.345	.558
Task	42.295	3	14.098	4.014	.009
treat * task	1.474	3	.491	.140	.936
Error	389.852	111	3.512		

Table 3.17 Transcription versus reporting: number of dysfluent words

Source	Type III Sum of squares	Df	Mean square	F	Sig.
Intercept	81583.999	1	81583.999	366.000	≤ .001
Treat	420.784	1	420.784	1.888	.172
Task	702.683	3	234.228	1.051	.373
treat * task	486.154	3	162.051	.727	.538
Error	24742.667	111	222.907		

Table 3.18 Transcription versus reporting: number of pauses

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	348.350	1	348.350	89.363	≤ .001
Treat	10.245	1	10.245	2.628	.108
Task	44.148	3	14.716	3.775	.013
treat * task	23.229	3	7.743	1.986	.120
Error	432.695	111	3.898		

3.5.2 Results for fluency across task performances

Mauchly's test of sphericity for length of run was non-significant (p = .636). The follow-up ANOVA test was significant (p ≤ .001, F = 12.761) There was also a

73

significant quadratic effect ($p \leq .001$, $F = 21.028$).

3.19 Length of run per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	5.954	.098	5.760	6.148
2	6.442	.115	6.215	6.669
3	6.631	.119	6.395	6.868
4	6.334	.111	6.115	6.554

For length of run the means (Table 3.19) show a recurring pattern of the first performance having the lowest fluency and the third the highest. The pairwise comparisons reveal that there are significant results between task performance one and task performance two, and then task performance one and task performance three. In both cases the first performance is significantly weaker.

There was a significant effect for task session ($p = .009$, $F = 4.014$) on length of run (Table 3.16).

Table 3.20 Task session effect on length of run

Task session	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	6.756	.180	6.399	7.114
2	6.063	.175	5.717	6.409
3	6.040	.164	5.715	6.364
4	6.503	.174	6.158	6.848

I would suggest task topic did not influence length of run. The salient features of

Table 3.20 are that there appear to be two pairs of task sessions; sessions one and four as a pair and two and three as a pair that are similar in length of run. There is a significant pairwise comparison between task performances one and three with the first task performance being significantly greater than the third task performance. It also does not appear that topic can explain the difference because in each pair one of the tasks was ranked as one of the top two easiest tasks while the other was ranked as one of the two most difficult tasks. In the case of task session two and task session three, task two (health and fitness) was ranked as the least difficult while the third (pets and animals) was ranked as the second most difficult task. For there to be an effect of topic on length of run, the student survey feedback should support this conclusion.

As the results were puzzling for task effect on length of run, I reanalyzed the data and found that in the second and third performance there was a large group of students who had high length of run counts (number of length of runs per performance). This difference raised a question about the consistency of the codifying process, particularly since length of run was the most difficult aspect of performance to code. For this reason, I kept a record of coding decisions and gradually built up a set of rules to ensure uniformity, but by the time the rules were complete I had transcribed half of the data. While I transcribed the second half of the data and then returned to re-codify the first half again, task two was the first session coded while task four was the last. Thus the two pairs of task were from different transcribing times. I also reviewed all the transcripts a number of times and found no major inconsistencies with how I codified the data. Thus it appears that this result could be due to the sample size (n = 119) which is not large (see table 3.0). One possibility is that the

results for length of run were caused by a group of students trying to be more fluent in task sessions two and three, which coincided with the middle of the semester, and where students, from my experience, settle into the learning requirements for the semester. At the start of the semester students are generally getting readjusted while at the end of the semester fatigue becomes a factor. In this sense the results may reflect the flow of a semester.

3.5.3 Dysfluent words per task performance

For number of dysfluent words per task performance Mauchly's test of sphericity was non-significant ($p = .264$) and the ANOVA test was significant ($p \leq .001$, $F = 11.523$). There was a strong linear effect ($p \leq .001$, $F = 23.288$), which means there was a substantial increase from the first performance to last the last performance. The results of the analysis are presented in Table 3.21.

Table 3.21 Number of dysfluent words per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	9.838	.807	8.239	11.436
2	13.585	.913	11.775	15.394
3	14.149	.953	12.261	16.038
4	15.268	.978	13.329	17.206

The means suggest that with each performance there was an increase in dysfluent words. The pairwise comparisons reveal that the first performance was significantly lower than the other three task times. From the first performance to the third, an increase in dysfluent words corresponded to an increase in overall language

76

production. However, in the fourth performance total word production was greater than the first and second performance but decreased relative to the third performance. The fact that there was a time interval of close to a month between the third and fourth performance would suggest this variable may possibly account for the increase in dysfluent words in the fourth performance.

When I reviewed the raw data of the 119 task session performances (four performances per session), 43 had the largest number of dysfluent words at the fourth performance. This increase of dysfluent words at the fourth performance despite the slight decrease in production relative to the third could be the result of the three or four week interval between the two previous task performances, and the fourth. Essentially, the task was familiar to the students, but yet the fact that they had not done it for a month or slightly less means they may have had some trouble recalling what they did or finding new topics to pursue. One final possible reason, discussed in the next section, is that over the four performances accuracy improved including an increase in reformulations. Generally, when students reformulate or self-correct dysfluent language is more prevalent than when they produce false starts or repetitions because self-correction can often involve repeating whole phrases or clauses. This may have contributed to the increase in dysfluency across performances. The dysfluent language produced from reformulations can be seen as a trade-off between accuracy and fluency.

3.5.4 Number of pauses

For the measure number of pauses there was only a significant result for task session effect (see Table 3.18). Number of pauses per performance was not affected by

77

repetition. There are two possible reasons for the results. First, students when speaking overwhelmingly produced dysfluent language rather than filled or unfilled pauses. Second, in coding I did not count pausing at AS-unit boundaries; instead the coding only included speaking phenomena within AS-units. In many cases the pausing at AS-unit boundaries was not pausing but shared silence, and there was a considerable amount of pausing/shared silence which may have yielded more detailed results had it been possible to distinguish between the two, but it would be difficult to do for between AS-units. Nevertheless it was typical for the speaker to be silent for some time before speaking after a question was asked, and since this was so common in the data, it may have be an effective measure for future analysis.

3.5.4.1 Task session effect for pausing

Task session effect on the number of pauses was the only significant result for this measure.

Table 3.22 Task session effect on number of pauses

Task session	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	.477	.190	.101	.854
2	.840	.184	.476	1.205
3	1.325	.173	.983	1.667
4	.810	.183	.447	1.174

The salient feature for task effect on number of pauses appears to be the large gap between the first task session and the third task session. The significant pairwise result between the two sessions supports this. While the third session does appear to have a bit more pausing than task sessions two and four, the results for task three were

distorted by one student's performance in the third task. She paused 29 times over the four task performances in task session three, an exceptionally large amount of pausing. In comparison most students paused well over less than half of this amount.

3.5.5 Summary of fluency measures

In summary, for fluency the means for length of run follows the same quadratic trend as complexity; namely, from the first to the third performance there was an increase in the measures before a marginal drop in the fourth. The pairwise comparisons reveal a significant difference between task performance one and task performance two, and then task performance one and task performance three. This would suggest that there was an improvement from task performance one in terms of length of run. The means for number of dysfluent words shows a linear trend upward with the fourth performance showing an increased level of dysfluent language, potentially explained by the one-month interval and an increase in reformulations. The pairwise comparisons reveal that the first task performance was significantly different from the other three task performances with a small number of dysfluent words. There was also an apparent task session effect on length of run, although the small sample size may account for this anomalous result, as some students had high length of run counts (number of length of runs) in two task sessions. I suggested the location of these tasks in the semester could account for this, as students tend to be most engaged during the middle of the semester and the task sessions with significant increases in fluency measures occurred in the middle of the semester. At times a low length of run average can be associated with risk taking, as it reflects students trying to produce long extended utterances that end up being comprised of spurts of speech. This is the case in those two task sessions where a group of students produced high length of run

counts (number of length of runs). Most were generally trying to be fluent, which resulted in lower length of run averages. Finally, there were non-significant results for pausing except for a task session effect, which may be the result of one student's performance skewing the data. The reason for the non-significant result for pausing was, I suggested, the result of what I excluded and included in codifying for pausing. I only codified pausing that occurred within AS-units and not between them, but students did not tend to pause once they started speaking, although they did pause before they spoke.

3.6 Accuracy results

Accuracy was measured through number of error-free AS-units, total number of error-free AS-units seven words and over, total number of words of AS-units seven words and over, and number of reformulations. As with the other measures there were no main effects for transcribing versus reporting.

Table 3.23 Transcription versus reporting: number of error-free AS-units

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	18177.533	1	18177.533	1052.920	.000
Treat	9.640	1	9.640	.558	.456
Task	369.633	3	123.211	7.137	.000
Treat*task	114.115	3	38.038	2.203	.092
Error	1916.296	111	17.264		

80

Table 3.24 Transcription versus reporting: total number of AS-units seven words and over

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	1540.418	1	1540.418	405.919	≤ .001
Treat	.844	1	.844	.222	.638
Task	71.136	3	23.712	6.248	.001
treat * task	6.230	3	2.077	.547	.651
Error	421.233	111	3.795		

Table 3.25 Transcription versus reporting: number of words from AS-units seven words and over

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	134411.803	1	134411.803	353.478	≤ .001
Treat	242.014	1	242.014	.636	.427
Task	7740.073	3	2580.024	6.785	.000
Treat * task	498.727	3	166.242	.437	.727
Error	42208.331	111	380.255		

Table 3.26 Transcription versus reporting: number of reformulations

Source	Type III Sum of squares	df	Mean square	F	Sig.
Intercept	1045.787	1	1045.787	298.799	.000
Treat	.238	1	.238	.068	.795
Task	22.800	3	7.600	2.171	.095
Treat * task	14.167	3	4.722	1.349	.262
Error	388.496	111	3.500		

3.6.1 Number of error-free AS-units across task performances

Results across task repetitions were significant for all measures used. For number of error-free AS-units, Mauchly's test of sphericity was non-significant (p = .120). The ANOVA test was significant (p = .002, F = 4.977). The polynomial trends showed only a significant linear effect (p = .003, F = 9.078), indicating the overall trend is for greater accuracy with no significant drop back.

Table 3.27 Number of error-free AS-units per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	5.628	.272	5.088	6.168
2	6.076	.302	5.478	6.674
3	6.886	.277	6.337	7.435
4	6.352	.270	5.817	6.886

The mean scores follow the main pattern observed throughout the different CAF measures. The first performance had the least number of error-free AS-units and the third the most. The pairwise data shows that there is a significant difference between task performance one and task performance three with the first task performance being significantly less accurate than the third task performance. As the data shows the third performance had on average 6.88 error-free AS-units per performance. The second highest was the fourth performance at an average of 6.35 units. As the third and fourth performances had the two highest error- free performances, the overall trend appears that students produced more error-free AS-units as the repeated the tasks.

3.7 Task session effect on accuracy measures

The next table shows the mean averages for task session influence on error-free AS-units. For task session effects I have been cautious about drawing inferences from the results unless they were consistent with other data. In this case the quantitative results reinforce students' survey feedback in regard to task difficulty. Task session one (health and exercise) and task session three (shopping and fashion) were identified as the two easiest while task session two (pets and animals) and task session four (dating and marriage) were identified as the two most difficult with dating and marriage considered the most difficult. The reasons for these choices appears to be topic familiarity.

Table 3.28 Task session effect on error-free AS-units

Task session	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	6.810	.400	6.017	7.603
2	6.125	.387	5.358	6.892
3	7.156	.363	6.436	7.875
4	4.851	.386	4.086	5.616

Table 3.28 reveals the two most difficult task sessions (task two and task four) exhibited fewer error-free AS-units while the easiest task session had the highest number of error-free AS-units. The difference between the third task session and the fourth is the most salient. I would suggest students had trouble being accurate when the task was unfamiliar and involved some affective discomfort. The pairwise comparisons showed significant differences between task session one and task session four, and task session three and task session four were significant. This means that in

particular the fourth task session which was identified as the most challenging showed significant differences in relation to two task sessions that students identified as easier. Specifically, task session four was significantly less accurate than task session and task session three.

3.8 Number of error-free AS-units seven words and over and number of words from error-free AS-units seven words and over per task performance

For both measures Mauchly's test of sphericity was non-significant (p = .107 and .106). Both ANOVA tests were significant (p ≤ .001 and p ≤ .001, F = 16.502 and F = 15.014). Polynomial results produced a significant quadratic trend (p ≤ .001 and p ≤ .001, F = 18.436 and F = 19.296).

Table 3.29 Number of error-free AS-units seven words and over

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	1.275	.097	1.082	1.467
2	1.746	.145	1.459	2.033
3	2.429	.154	2.123	2.734
4	1.811	.138	1.538	2.085

Table 3.30 Total number of words from error-free AS-units seven words and over

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	11.498	.971	9.574	13.423
2	17.000	1.510	14.008	19.992
3	22.700	1.583	19.563	25.836
4	16.624	1.309	14.030	19.218

The accuracy results for these two measures again reflect the main trend that has emerged throughout the whole study. With both measures the third performances were twice that of the first performance. With table 3.29 the pairwise comparisons show that the third task performance was significantly different in relation to the other three task performances. In other words on this measure it was significantly more accurate than any other task performance. In addition the fourth task performance is significantly different to the first task performance. In relation to the pairwise comparisons for table 3.30 there were significant results as well. As with the previous table and was common throughout the study the first performance was significantly weaker than the other three task performances. The other three task performances on this accuracy measure were significantly more accurate than the first. Finally, the third task performance on this measure was significantly more accurate than the fourth task performance.

There was also a task session effect on both measures (see tables 3.24 and 3.25). They exhibit the same pattern as number of error-free AS-units.

Table 3.31 Task session effect on number of error-free AS-units seven words and over

Task session	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	2.168	.188	1.797	2.540
2	1.399	.181	1.040	1.759
3	2.235	.170	1.897	2.572
4	1.458	.181	1.100	1.817

Table 3.32 Task session effect on total number of words from error-free AS-units seven words and over

Task session	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	20.323	1.878	16.603	24.044
2	12.410	1.817	8.810	16.010
3	21.563	1.704	18.185	24.940
4	13.527	1.812	9.937	17.117

The means from the tables show that more difficult task sessions/topics appeared to have resulted in students producing less error-free language. The pairwise comparisons support this. Task session one is significantly more accurate than task session two, while task session three is significantly greater than both task session two and task session four. For task session two and task session four the students produced fewer error-free AS-units seven words and over. In addition the total word count of these error-free AS-units was also lower. In the next chapter I will analyse the nature of accuracy and suggest that accurate student performances predominately involved the student producing partially proceduralized language in the form of multi-word

chunks. The results above would suggest that this also occurred when tasks were familiar to the students.

3.9 Number of reformulations

The final measure, which reflects an orientation towards being accurate, did not have a significant result for Mauchly's Sphericity test ($p = .193$). The ANOVA test was significant ($p = .002$, $F = 4.899$) and there was a significant linear effect ($p = .001$, $F = 12.395$).

Table 3.33 Number of reformulations per task performance

Task performance	Mean	Std. error	95% Confidence interval	
			Lower bound	Upper bound
1	1.116	.128	.862	1.370
2	1.485	.134	1.219	1.751
3	1.569	.158	1.256	1.882
4	1.813	.147	1.522	2.103

What can be seen from the means for each performance is that there was a slight increase in the number of reformulations from task performance one (1.1 reformulations) to the fourth task performance (1.8 reformulations). There is a significant pairwise comparison between task performance one and task performance four with the latter having significantly more reformulations.

3.10 Summary for accuracy results

In summarizing like other CAF measures, the means suggest that accuracy was lowest during the first task performance and highest in the third task performance. What is more this improved accuracy was characterized by longer more accurate AS-units.

On all the measures in relation to the pairwise results task performance one was significantly less accurate in comparison to at least one other task performance. For error-free AS-units seven words and over, and number of words from error-free AS-units seven words and over, the first task performance was less accurate than all three latter task performances. The results would suggest that students accuracy improves when then repeat familiar tasks.

Lastly, there was a task session effect on three of the four measures, with the two most difficult topics the least accurate. Significant pairwise comparisons confirmed that task session four on three of the accuracy measures was less accurate than task session one and task session three. In addition on error-free AS-units seven words and over, and number of words from error-free AS-units seven words and over, task session three was also significantly more accurate than task session two. These result support the suggestion that task session or topic influenced student accuracy. For more difficult tasks it impacted accuracy negatively.

3.11 Overall summary of quantitative results

At the end of each CAF section I provided specific summaries of the quantitative data results. In this section I will briefly summarize these results in relation to the questions I have asked in Chapter 1. In the following qualitative chapters I will provide a more thorough analysis of the quantitative results through a case study approach. The first question involved investigating if transcribing and reporting in task repetition facilitate student task engagement and improved language performance. This question was asked in relation to my module two results where students' repeat task performances did not show improvement. The quantitative

results do suggest that these interventions provided a clearer pedagogical focus which appears to have contributed to improved repeat task performances including task performances three to fours weeks after the last prior task performance. In the next three chapters I will specifically address how directly or indirectly the two treatments contributed to the results. The second question I considered was would the more intensive transcribing process result in better task performances than the reporting process as reflected by CAF measures. The results for this question as reported were very clear as neither group outperformed the other on any measure of CAF. In the following qualitative chapters I will analyze this result and provide possible reasons for it.

Chapter 4 The case study

4.0 A qualitative perspective

As explained earlier, this research incorporated two parts; the quantitative analysis of all of the student data presented in Chapter 3, and a qualitative case study of one student, which will be presented here and in Chapter 5 and 6. Incorporating a qualitative aspect to this research was partly inspired by recent calls for more mixed methods research in applied linguistics (see Dörnyei 2007). This case study intends "to show the kinds of language hiding behind the numbers" (Bygate and Samuda 2005: 54) and to highlight the teaching and learning variables that appeared to influence student production during their task work.

At the start of Chapter 4 I will discuss the qualitative methodology for the qualitative chapters. This will focus on why I chose a case study approach including how I selected the student used in this study. In addition in this chapter I will provide an overall descriptive account of the spoken characteristics of most students in this study. I have done this in order to provide the reader with contextual information on how to read and understand the transcripts of this study. In each of the three chapters I ask key questions that emerged while I was doing this study. In Chapter 4 I ask two questions that relate to how effective pre-task work was, and, second, how important the anchorperson's role was in facilitating their partner's spoken production. The first question relates directly to the first performance while the second question applies to all task performances. For Chapter 5 I will analyze the error correction and reformulation processes used in this study as well as the second task performances. The main question I ask in this chapter is how effective the error correction and reformulation processes were in this study. This analysis should provide important

information on why there were no significant results for the transcribing versus reporting portion of the study.   The second question involves first asking how productive were the second performances; and, moreover, how much 'carry-over' was there from the first performance to the second performance. By this I mean to answer how much language did students repeat from the first task performance in the second task performance. I will use my case study as an example.   In Chapter 6 I address the third and fourth task performances. The questions I ask with both performances are the same as the second question in Chapter 5. Namely, I want to investigate how productive these latter performances were and what were their chief characteristics, including how much carry-over or language was repeated from the second task performance to the third task performance, and then finally from the third task performance to the fourth performance.

4.1 Case study defined

Duff (2008) describes case study research as characterized by the following principles: *Singularity*, *in-depth study*, *multiple perspectives*, *particularity*, *contextualization*, and *interpretation.* Critical questions for organizing case studies include how many subjects to select, which ones, and why. Furthermore, a rationale must be given for choosing to sample data intensively or extensively (Duff 2008). In this research I decided to focus on one student, or intensive data sampling, to help reveal more than just the language behind the numbers; my intuition was that the language students produced was embedded in classroom processes and interactions and thus focusing on one student would allow those interactions to come to the fore while sampling extensively would have resulted in a focus on language instead of the social processes behind the language. Thus this case study is intended as an in-depth

analysis that as a secondary focus includes a picture of my involvement in the class, showing how I as the teacher participated in the lessons and allowing for an evaluation of my teaching, addressing Samuda and Bygate's (2008) criticism that task-based studies have a tendency to leave out the teacher's role. Since this study's ultimate purpose is to inform classroom pedagogy, including my teaching in the analysis is essential.

4.2 The student chosen for this case study

For this study I chose D5 after data collection was complete because she was one of a number of students who I had taught previously, and I was therefore familiar with her attitude towards English and her English proficiency level. I hoped this familiarity would help me interpret her motives and performance. She was a pleasant student but did not like English and like most of the students involved in this research was low in English proficiency. While her motivation and willingness to do English varied from week to week if she enjoyed a particular activity and thought the goals achievable she would put in the effort necessary to complete that particular exercise, in general she was not a particularly motivated learner. Her attitude and behaviour was typical of most students including that she was very honest about her feelings toward the classwork, which I felt would assist in interpreting the implications of her task discourse. Additionally, based on transcripts and notes about common features of student performances, such as how they improved across task performances, D5's language fit the norms for student language production. In short how she improved on aspects of her spoken production is consistent with the underlying processes I believe most students demonstrated. Finally, in this study I use only one of D5's task sessions. This was done to provide a complete and comprehensive account of the

learning and teaching. I selected the task session on health and exercise because a number of the measures roughly fit the trend for quantitative results where the first performance was the weakest while the third performance was the strongest. For example, D5 with total words per task performance produced the following levels of production:

Task performance 1   70 words

Task performance 2 139 words

Task performance 3 167 words

Task performance 4 80 words

Overall though my choice for picking this task session is that it best encapsulates the key processing features of most of the students in this study which help explain the quantitative results.

4.3 D5's first task performance and spoken production characteristics

4.3.1 Non-fluency

The purpose of this section is, using D5's performance, provide a summary picture of most of students' language production and how I interpreted those performances. This background is important for reading the transcripts and understanding what level of language proficiency the transcripts are trying to reflect. In this section I am using the term 'fluent' to reflect overall proficiency, or as Segalowitiz (2010: 30) notes the "ease and efficiency with which the underlying planning and production processes function". In this sense all CAF measures fall under this definition. After listening to all of the student performances, I concluded the majority of students, including D5, were not fluent. This was especially obvious in the first performances of task

sessions, students tended to rely on their declarative knowledge, as their production was more controlled than proceduralised or automatic (for a definition of these terms see DeKeyser 2007b). However, my non-fluent students were still capable of becoming temporarily fluent during a task performance, or artificially fluent, which became possible when pre-task work and task repetition eased language processing pressures. To illustrate what temporary or artificial fluency means as it applies to student production, I have focused on the parts of D5's transcripts that demonstrate her reliance on controlled processing, using her first performance as an example because this is where all aspects of her spoken production were consistently the weakest of the four task performances, a characteristic consistent with other students' task language. Before moving on to analyse D5's transcripts in the next section it is important to note that orthographic transcripts, such as those used here, have limitations in reflecting spoken production, particularly when prosodic features are absent, as in this study. Nevertheless the transcripts as they are presented should help to make the discussion of their contents more salient for the reader.

4.3.2 Characteristics of low fluency: slowness of speech

One salient feature indicating a lack of fluency in D5's first repetition was the slowness of her speech. She produced some idiosyncratic formulaic chunks that made her production less controlled. However, a large part of her production throughout most task performances was spoken one word at a time and even one syllable at a time on occasion. Also, when using apparently formulaic language, she tended to enunciate that language word by word. To demonstrate the slowness of her speech, the number of seconds it took her to produce each AS-unit are included in Extract 4.0 below.

Extract 4.0

Shared Silence 14.52

1P: How often do you exercise?

2D5: /{I}[I play tennis]{5.70}{two}[two times a week]/**(7)(15.00 seconds)**

3P: Only tennis?

4D5: /Yeah/

5P: Then what sports or exercise you used to do? Tennis or (in Japanese *anything*

       else?)

6D5: /[I used to {used to?} play to basketball]{2.40} for {2.56}[junior high school

       and high school]/**(12)(21.70 seconds)**

7P: So six years

8D5: /Yeah/

9P: Why do play tennis in college?

10D5: /{I I started I want to start}{6.03}{I} I want to started new sports (in Japanese

       new)/**(6)(23 seconds)**

11P: So which do you like basketball or tennis?

12D5: /{I}{5.55}[I like basketball and tennis but]{3.79}[now I play tennis twice a

       week so]{I}{3.07}][my best sports is tennis]/**(19)(34.19 seconds)**

Shared Silence 13.00

13P: Do you do any other sports?

14D5: /{I I I I I have played} I used to do swimming/**(5)(28.22 seconds)**

15P: Can you swim butterfly?

16D5: /Yeah/

17P: Can you swim well? Fast?

18D5: (In Japanese *One more time)*

19P: Can you play fast? (in Japanese *Fast?)*

20D5: /So so/

21P: Do you have a lot stress?

22D5: /{I} I don't have a lot of stress/**(8)(4.52)**

23P: Don't you have some problems?

24D5: /{If} If {I}[I have problems or stress but {I}]{4.31}{I}[I forget {8.23}{while}

while sleep]/**(11)(32.5 seconds)**

With the exception of line 22, all of D5's AS-units took a substantial amount of time to complete. For example, at line 24, even if the nearly 13 seconds of pausing is subtracted, the 11-word AS-unit and five dysfluent words still take about 20 seconds to produce, which averages 1.25 seconds per word (including dysfluent words).

4.3.3 Characteristics of low fluency: dysfluent words

In this study an increase in word count during a task performance almost always coincided with an overall increase in dysfluent language. The amount of dysfluent words in a performance reflects the fluency of a student. Generally, the fewer dysfluent words a student produces the more fluent his/her performance is considered (see Ellis and Barkhuizen: 2005). My results are probably related to the low proficiency level of my students where any increase in production will be accompanied by an increase in dysfluency. In this sense the number of dysfluent words per performance can at time be a positive indicator of improved production including an attempt to be more fluent. Current task-based research tends to use subjects that have already acquired a degree of proceduralization or automaticity in their oral production. This in turn creates a certain perspective on how fluency is

96

described and understood. For example in task-based research, fluency is commonly measured by breakdown fluency (filled and unfilled pausing) or repair fluency (reformulations, repetitions, false starts, replacements). However, the terms repair and breakdown suggest students are already capable of fluent language, or can produce fluent speech, which at times breaks down or needs to be repaired when their processing capacity is overtaxed. This means current task-based research seems designed around students who already have a degree of proceduralisation or automaticity in their oral production, which in turn influences how fluency is described and understood in the literature. Yet with the students in this study, who lack such automatic language processing capabilities in English, the usual ways of measuring fluency are less useful or less revealing.

I would argue D5's dysfluencies, as with most students in this study, were the result of controlled processing and that in her case fluent speech emerged through dysfluent speech, and not vice versa. Thus in this study repair and breakdown fluency need to be thought of differently. If it is argued in this study that the students were primarily non-fluent then repair and breakdown fluency need to be understood from a non-fluent performance perspective. Breakdowns are more about 'struggling to speak' than speech breaking down. The difference is that with the former the students are struggling just to produce basic utterances while with the latter the students are capable of producing extended utterances which at times results in breakdown of speech as their language processing is taxed by real-time processing pressures.

With repair it is not necessarily the student deliberately trying to repair language, but rather struggling holistically with production. On this point one of the main characteristics of struggling to speak (discussed in module two and also see Kumaravadivelu 2007) is that students do not often distinguish between meaning and form. They process language holistically. The processing pressures of speaking real-time and students' lack of proficiency prevent this. The processing of language holistically is important to recognize especially for repair fluency, in that researchers see self-corrections as a deliberate 'attentional shift' from meaning focus to form focus. In my study I would suggest that what appears to be deliberate self-correction is more often than not the student struggling to produce speech. Below are two extracts that demonstrate the above points about viewing breakdown and repair fluency differently in my teaching context.

In D5's first task performance, 10D5 and 14D5 (Extracts 4.1 and 4.2) are good examples of D5's struggle with basic production, and how fluent speech emerged from non-fluent speech. These two AS-units comprised 51 seconds of her two minutes and thirty-nine seconds of task production time, or 32 percent of her whole first task performance time. This means that while these two AS-units were small, they were a substantial focus of D5's attention during her first performance.

Extract 4.1

9P: Why do play tennis in college?

10D5: /{I I started I want to start}{6.03}{I} I want to started new sports (in Japanese

        new)/(6)**(23 seconds)**

Extract 4.2

13P: Do you do any other sports?

14D5: /{I I I I I have played} I used to do swimming/(5)**(28.22)**

At 10D5, the dysfluent language could be categorized as a self-correction. However, I would suggest D5 was not focusing on form any more than she was struggling to be meaningful. Thus it was not a deliberate focus on form that resulted in the dysfluency, but rather D5's real time struggle to make meaning. With 14D5 the AS-unit could be viewed a self-correction, repetition, or false start. If it was a self-correction, then it has to be assumed she deliberately created the dysfluent language as she addressed the form of the language. The term false start (as, for example, in a race) suggests there is a quick start followed by abandonment and then a restart. However, the multiple repetitions of *I* suggest there was nothing quick about her language. This would suggest D5 did not switch from automatic to controlled production or even delayed automatic production to focus on form. Instead D5 depended on controlled processing as she tried to be accurate and meaningful at the same time.   In concluding this section, I have tried to demonstrate D5's production, like most students in this study, was controlled, non-fluent, and she produced language holistically with little differentiation between meaning and form. These three characteristics should be kept in mind when reviewing her task production. Through task repetition and other pre-task work students like D5 could improve their task performance and achieve temporary states of fluency, accuracy, and complexity, but overall their task performances involved struggling to speak.   In other words as students become familiarized with the field of the task through preparing and repetition the aspects of their production and processing become temporarily inflated. It is at this state where

students are arguably working at the limits of their interlanguage. The perceived benefit of this is that it is arguably a catalyst for language restructuring.

4.4 Research questions

After reviewing the quantitative results and students' first task performances, I had two main questions to address in analysing D5's pre-task work and first task performance. One question was regarding the influence of pre-task work on production and the other concerned the quality of the first task performances. These questions are explained in greater detail below.

4.4.1 Investigating the influence of pre-task work on first performances

Task repetition research has demonstrated a first task performance is generally poorer in comparison to subsequent task repetitions. This held true in this study despite a substantial amount of unguided pre-task work prior to first performances. On this point one of the questions I have with regard to the first task performance was:

1. *What influence if any did unguided pre-task work have on the first task performance?*

Students at the pre-task stage completed unguided strategic planning by writing down their responses to questions I provided. The questions were intended to better frame the topic for students and provide them with a basic schema of what to talk about, intentions consistent with research that shows strategic planning tends to facilitate macro-planning over micro-planning (Bygate and Samuda 2005). Considering the low proficiency of the students, I had assumed they would use language from their pre-task work in their first task performances. Thus here I investigate whether D5 used

language from her pre-task work, how much she used, and what benefit if any it had for her first task performance.

An important variable in explaining the influence of pre-task work on the first task repetition was the role of the asker. The asker as the anchorperson in the dyad decided what questions to use from the pre-task. If the anchorperson primarily asked pre-task questions then hypothetically the speaker should perform better than if asking different questions that the answerer had not prepared for. However, productive performances in the data incorporated the anchorperson asking a limited number of pre-task questions throughout the task, which were supported by follow-up questions in each instance. In contrast, poorer performances were characterized by the anchorperson relying solely on the pre-task questions and not asking original follow-up questions, revealing how critical the anchorperson's role is to the quality of the speaker's performance. I illustrate specific instances of this using extracts from D5's first task performance in Extract 4.5 below.

4.4.2 First performances: how did students perform?

The second question I investigated in regards to the first task performance was:

2. *How poor were first performances?*

The quantitative results indicated the first performances were the weakest of all four of the performances over the four task sessions in terms of the measures used. Using D5's first performance, I have highlighted key production characteristics of her first performance. I would argue that her production generally reflects the first performances for most students. In her first performance D5 produced a total of 70

words, ten less than in her fourth performance and less than half of the second and third performances. In her first performance only three of her seven AS-units were error-free, but two were seven words and over. Six of her seven AS-units were seven words and over. Her average number of words per AS-unit was ten words, which was quite high relative to other students. Her length of run was 8.8 words. Overall D5 did not produce a lot of language and was not very accurate, but her performance was fairly complex in regards to the number of words per AS-unit relative to the other students' first task performances in this study, meaning she performed slightly better than average. Nonetheless, the overall features of her performance, including non-fluent speech (Extracts 4.1-4.3), were characteristic of most students in this study during their first task performances.

Looking at the quantitative results a legitimate concern is whether the first performance results involved minimal language use or students not engaging with the task. This was not the case for D5 or for the majority of students in the study. Rather the results are indicative of students at a very basic level of language ability performing a new task for the first time. At 12D5 (Extract 4.3) and 24D5 (Extract 4.4), D5 engaged in output pushing. In Extract 4.3 D5 assembles her longest AS-unit from the first task performance.

Extract 4.3

11P: So which do you like basketball or tennis?

12D5: /{I}{5.55}[I like basketball and tennis but]{3.79}[now I play tennis twice a week so]{I}{3.07}][my best sports is tennis]/(19)**(34.19)**

Shared Silence 13.00

The substantial pausing at clausal boundaries demonstrates this long AS-unit was not easy for D5 to produce, and in this study a student producing 19-word AS-unit, especially in the first performance, was excellent. Extract 4.4 shows D5's most grammatically challenging AS-unit in the first task performance, involving clausal embedding.

Extract 4.4

23P: Don't you have some problems?

24D5: /If {I}[I have problems or stress but {I}]{4.31}{I}[I forget {8.23}{while} while sleep]/(11)**(32.5 seconds)**

Here D5 stretched her production to the point where it resulted in the longest pausing of this task performance, somewhat successfully producing a conditional relationship between two clauses. Extracts 4.3 and 4.4 demonstrate D5's willingness to take risks during even this first task performance, and other students' first task performances were similar. Thus rather than seeing these first performances as worse than the other repetitions, I would in most cases argue they represent the necessary precursors to the more complex performances in later task repetitions.

4.5 D5's pre-task work

This section explores how the pre-task work influenced D5's first task performance language. Figure 4.0 shows D5's pre-task work answering the provided questions on this task session topic.

Figure 4.0 Pre-task notes

Lesson for May 21st

Health and Exericse

- Is your lifestyle healthy or unhealthy? Give least 4 reasons why

Because
1. My lifestyle is healthy
2. I take a lot of sleep time.
3. I like vegerables very much.
4. I go to school every day .

- What do you do to stay healthy or what do you do that is unhealthy? Give 4 examples

1. don't have breakfast
2. go to bed too late.
3. do not to do anything.
4. not excise

- How often do you exercise? What do you do? What sports or exercise you used to do?

I often to play tennis two times a week.

- Is your diet healthy or unhealthy? Give 5 reasons.

1. diet myself on vegetable
2. Body become confortable move when loseweight
3. It is become beautiful skin.
4.
5.

- How often do you eat junk food? When do you eat it and what is your favorite junk food?

My favorite junk food is potatochips and chocolate.

- Do you have a lot of stress? Please explain why you do or why you don't?

No. I don't. Because I forget to problems when I'msleeping.

- How many hours of sleep do you get in a night? How about recently? Why?

I sleep for 4 hour
Because I have to many send in subject.

- How can you improve your healthy and lifestyle? Give 4 reasons
1. I get up early.
2. I eat decrease junk food.
3. I have more vegetables.
4. I go to bed at early.

D5 did not read her pre-task notes when answering questions in the first task performance, unlike many other students in their first task performances when struggling with speaking. D5's partner asked her three questions from the pre-task sheet, illustrated in Extracts 4.5 to 4.7 below, along with D5's answers. The extracts

also include what D5 had written in her pre-task notes prior to the first task performance.

Extract 4.5

Question 1

1P: How often do you exercise?

2D5:   /{I}[I play tennis]{5.70}{two}[two times a week]/(7)**(15.00 seconds)**

3P: Only tennis?

4D5:   /Yeah/

Pre-task notes: I often to play tennis two times a week

Extract 4.6

Question 2

5P: Then what sports or exercise you used to do? Tennis or (in Japanese *anything else?)*

6D5:   /[I used to {used to?} play to basketball]{2.40} for {2.56}[junior high school and high school]/(12)**(21.70)**

Pre-task notes: *I often to play tennis two times a week*

Extract 4.7

Question 3

21P: Do you have a lot stress?

22D5: /{I} I don't have a lot of stress/(8)(4.52)

23P: Don't you have some problems?

24D5: /If {I}[I have problems or stress but {I}]{4.31}{I}[I forget {8.23}{while} while sleep]/(11)**(32.5 seconds)**
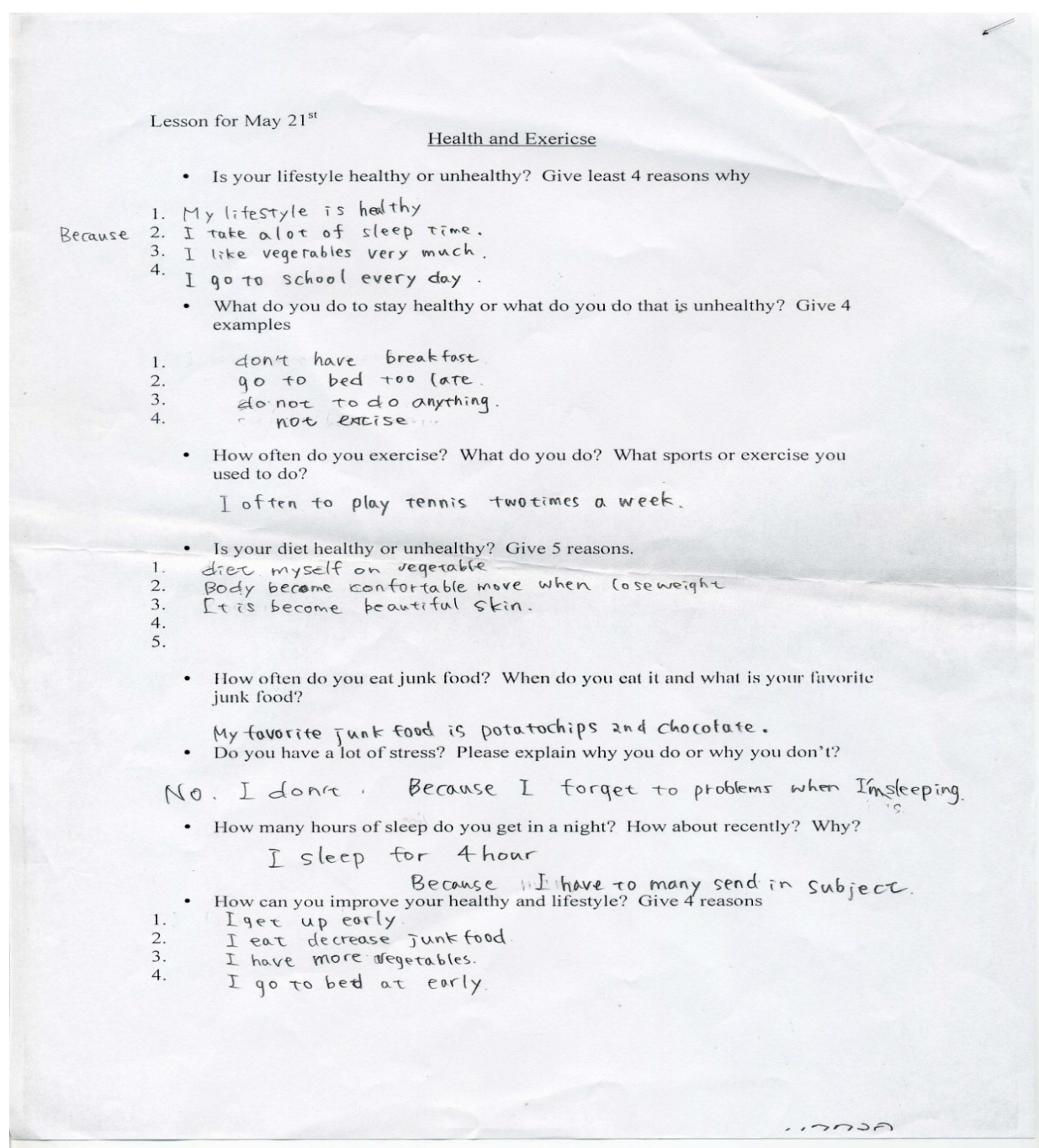
D5 responded to pre-task questions one and three (Extracts 4.5 and 4.7) relying on the language she had prepared in her pre-task notes, although in answering question two (Extract 4.6) she used different language, perhaps because she wrote the same answers for questions one and two in her pre-task notes. In other words she used the response she had prepared in her pre-task notes for question one and probably because she did not want to reuse the same response she thus used a new response for question two. D5's responses to questions one and three (Extracts 4.5 and 4.7) were not verbatim from her notes, though; both responses represent expansions of her pre-task notes. Her response to question one, unlike in the pre-task work, was grammatically correct, showing she was still composing her answers during the task work and not simply relying on her pre-task preparations for this first task performance. Furthermore, in Extract 4.7 D5 engages in risk taking, borrowing language from her partner's initial question (incorporation strategy) and connecting it to a clause from her pre-task work.

Pre-task work was used in most students' first task performances and was more common with less proficient students and less motivated students. D5's first task performance was typical of most students in that she relied on multi-word chunks from her pre-task work. The fact that language from her pre-task work appeared almost verbatim in her answer to the first and third questions suggests she was memorizing or recalling them as multi word units, leading me to refer to this language as chunks, which in most cases were clauses or phrases. Despite being produced as whole phrases and therefore being partially formulaic, this does not mean they were produced rapidly. Especially in the first task performances they were often produced

word by word. In other words, while they appear to have been stored as chunks, when it came to speaking the word by word nature made me loosely interpret them as only partially proceduralised. In D5's performance these chunks either stood alone as single AS-units or like in 12D5 and 24D5 they were adjoined to other clauses, increasing the complexity of her performance. The AS-units in turns 12D5 and 24D5 were two of her three largest AS-units in her first task performance.

4.6 Importance of an active anchorperson

Extracts 4.5 to 4.7 also highlight a key characteristic of all good task performances; the use of follow-up questions by the anchorperson. This was particularly important for the first task performance because in these first performances anchorpersons tended to depend more heavily on the pre-task questions than in later repetitions, making the asking of follow-up questions less common here. However, the anchorperson's willingness to push their partner's production through follow-up questions became an important trait of all good speaker performances, meaning there were fewer productive performances in first task performances than in later repetitions. Extract 4.8 illustrates the importance of follow-up questions to improving the complexity and fluency of task language with respect to D5's first task performance. Her partner starts with a pre-task question at 1P, followed by an expansion question at 3P. She then asks another pre-task question at 5P based on D5's response, further pursued with expansion questions at 9P and 11P.

Extract 4.8

Shared Silence 14.52

1P: How often do you exercise?

2D5:   /{I}[I play tennis]{5.70}{two}[two times a week]/(7)(15.00 seconds)

3P: Only tennis?

4D5:   /Yeah/

5P: Then what sports or exercise you used to do? Tennis or (in Japanese *anything*

       else?)

6D5:   /[I used to {used to?} play to basketball]{2.40} for {2.56}[junior high school

       and high school]/(12)(21.70)

7P: So six years

8D5:   /Yeah/

9P: Why do play tennis in college?

10D5: /{I I started I want to start}{6.03}{I} I want to started new sports (in Japanese

       new)/(6)(23 seconds)

11P: So which do you like basketball or tennis?

12D5: /{I}{5.55}[I like basketball and tennis but]{3.79}[now I play tennis twice a

       week so]{I}{3.07}][my best sports is tennis]/(19)(34.19)

Shared Silence 13.00

This exchange was the best part of D5's first task performance and included the
longest, 19-word AS-unit (12D5), of the whole task session. The key to D5's longer,
more complex language is her partner's follow-up questions; after learning D5 used to
play basketball and now plays tennis, her partner asks why she changed sports and
which sport she prefers in 11P, a question not included in the pre-task questions. Thus
the longest AS-unit in this task performance is a result of a spontaneously generated
answer and not a prepared answer to a set question. Furthermore, D5's most
linguistically complex production at 24D5 (Extract 4.9) was also the result of her

partner's willingness to ask unscripted questions. In the Extract 4.9 at 22P D5's partner responds in a suspicious tone with a confirmation check of D5's reply which ends with D5's response at 24D5.

Extract 4.9

21P: Do you have a lot stress?

22D5: /{I} I don't have a lot of stress/(8)(4.52)

23P: Don't you have some problems?

24D5: /If {I}[I have problems or stress but {I}]{4.31}{I}[I forget {8.23}{while} while sleep]/(11)**(32.5 seconds)**

Not all follow-up questions resulted in a higher quality of production. The next Extract 4.10 includes an extension of D5 and her partner's exchange on tennis and basketball. Following some silence, D5's partner tried to continue the talk on sports with her own question at 13P.

Extract 4.10

Shared Silence 13.00

13P: Do you do any other sports?

14D5: /{I I I I I have played} I used to do swimming/(5)**(28.22)**

15P: Can you swim butterfly?

16D5: /Yeah/

17P: Can you swim well? Fast?

18D5: (In Japanese *One more time)*

19P: Can you play fast? (in Japanese *Fast?)*

20D5: /So so/

21P: Do you have a lot stress?

D5 responds positively at 14D5 and then after this struggle to speak, apparently unable to respond to her partner's follow-up questions at 15P, 17P, and 19P. At 21P, D5's partner seems to decide D5 did not have much to talk about on the subject of swimming and changes the topic, going back to a verbatim pre-task question at 21P. In contrast to D5's interlocutor who was active in her follow-up questions, in many other dyads it was common for the anchorperson to simply read the pre-task questions without including follow-up questions. This was particularly true for first task performances, resulting in an appearance of students just going through the motions in the tapescripts rather than having an interactive conversation, as the speakers would respond with minimal responses read verbatim from their pre-task work. Looking through the scripts, I found that a committed anchorperson could almost always motivate a reluctant speaker to speak but that the opposite was not necessarily the case; if the anchorperson was uncommitted to asking questions, speakers did not necessarily try to supply more information than required to answer the questions asked, indicating students viewed the anchorperson as a kind of conversational leader of the dyads. Additionally, proficiency did not seem to be linked to whether the questioner asked follow-up questions, as many weaker students did ask follow-up questions, particularly, "Why?" Possible explanations for not asking follow-up questions seemed to stem from lack of familiarity with the task or partner and apprehension about how pushy to be in asking follow-up questions. With regard to the first two, lack of familiarity, in later repetitions increasingly familiarity with the questions and partners meant more follow-up questions were asked. In regards to apprehension about output pushing, this relates to the issue of the learning agenda of

110

pushing output conflicts with the interpersonal agenda maintaining social cohesion and avoiding face threatening situations (see Batstone 2002), frequent conflict among students with low motivation to learn English, particularly at the beginning of the course. Over the course of the semester the problem of only reading pre-task questions without follow-up questions decreased, likely because I focused my post-task feedback on this problem in comments in student notebooks.

4.7 Summary of pre-task and first performance

In summarizing the first performance, I first contextualized D5's transcripts, arguing she was non-fluent, which means she depended on controlled processing during her task production. This made speaking a struggle for her and meant her language was produced holistically as she lacked attentional resources during real time communication to shift back and forth between meaning and form. D5's slowness of speech and her dysfluencies were both examples of her controlled processing of the language she produced. Regarding breakdown fluency and repair fluency, I argued in non-fluent production such as D5's (and many other students involved in this research), breakdowns do not represent a breaking down of fluent speech so much as they represent a process of starting up a chunk of language production. Furthermore, repairs do not represent a shift between meaning focus and form focus but instead attention to holistic language production, meaning both dysfluency indexes are evidence of students struggling to speak. A general lack of fluency was a feature of all task performances, but it was more pronounced in these first performances. In subsequent repetitions students became temporarily fluent more frequently from eased processing pressure resulting from repeating the tasks, but their production tended to involve struggling to speak throughout all of the task repetitions. Regarding pre-task

111

work, I demonstrated D5 made limited use of her pre-task work, using specific chunks of language from her notes and employing them almost verbatim. In two cases she combined a pre-task chunk with other language to increase the complexity of her performance (Extracts 4.5 and 4.7), typical of many other students. This language is referred to as chunks because she either memorized it that way or she, I would suggest, had already partially proceduarlised the language and so accessed it as a unit, as most of the language students wrote in their pre-task work was likely language they were already familiar with.

The final point of note in the first task performance and all performances was the importance of the anchorperson in asking follow-up questions. First performances where students appeared to engage with the task and produce more language were characterized by the anchorperson asking a limited number of pre-task questions and building off their partner's responses with unscripted follow-up questions, and D5's first performance was not exception. With D5's first performance and other repetitions where more language was produced during the task, the pre-task questions appeared to serve as framing questions as they often did not lead directly to answers with substantial language. Instead the longest AS-units tended to be the result of unscripted follow-up questions rather than the scripted pre-task questions. However, as demonstrated with the topic of swimming (Extract 4.10) not all follow-up questions led to an increase in quantity of production. The next Chapter discusses how convincing students to ask follow-up questions became a key focus of my feedback in the students' notebooks, as it became clear to me that the influence of the interlocutor on the task performances was crucial, a conclusion supported by the general TBLT literature (Dörnyei 2002, Dörnyei and Kormos 2000, and Dörnyei and Teng 2009b).

Lastly, while the quantitative data showed these first task performances were the weakest of the four, most first performances did contain positive engagement, with weaknesses stemming from students depending on controlled production as a result of doing the task for the first time, as demonstrated through analysis of D5's first task performance language above.

Chapter 5   The case study: transcription and second performance

There are two questions I will investigate in this chapter. They correspond with the error correction and reformulation processes and the overall second task performance. First, I will investigate: How effective are the error correction and reformulation processes in this study? The quantitative results or lack of them suggest that neither was very effective in that the transcribers who engaged in a more intensive reformulating process did not outperform the reporters on any CAF measure. By analyzing D5's reformulation notes and my feedback, I hope to provide a clearer picture of some of the factors that contributed to the non-results for the transcribing versus reporting. I will start this question off by first reviewing the self and teacher correction processes. This will be followed by an analysis of the reformulation process including a teacher-led feedback and input session I conducted.

The second question I ask in this chapter is: How productive were the second performances and what were their main characteristics? As a supplementary question I want to investigate how much 'carry-over' was there from the first task performance to the second performance? By this I mean how much language did students repeat or recycle from the first performance and following-up sequences to the second performance. This question applies to all the task performances after the first task performance, but is especially important in relation to the second task performance, because students were told to try and repeat the language in the second task performance they had first used in the first task performance and had subsequently worked on in the follow-up language focus sequences prior to the second task performance.

5.0 Correcting errors through transcription

This section describes D5's language production in the second task performance, which came after she had transcribed, self-corrected, and received my correction feedback on her first task performance. As D5 was assigned to the transcription group for this task session, the next stage in the lesson sequence involved transcribing and self-correcting her performance through editing and reformulation. Editing involved leaving out dysfluent language. Reformulating involved improving the preciseness of an utterance by inserting additional explanation (Lynch 2001). Figures 5.0 and 5.1 show D5's transcription and editing work, along with my corrections after she had completed her self-editing work. She made her corrections in pencil and also did the pink underlining herself. My corrections are in red. The blue brackets were added later and represent D5's attempts at reformulation, coded by me. Similarly, the pink stars represent general non-language related advice on what she should try to expand on in her second performance of the task.

Figure 5.0 Transcription work left page

Today's partner is Aya Simonaga.　May 20th Wredneseday

A. How often do you exrcise?

K. I play tennis two times a week *+(plus)* with tennis circle's member.

A. Only tennis? ← good

K. yeah. *+d* but I used to join the basketball circle too, but I ~~can't~~ *couldn't* contined because I was very busy.

A. Then what's sports or exercise you used to do.

K. I used to ~~plays~~ *play* basketball ~~for~~ *in* junior high school and high school.

A. So. six years?

K. yeah

A. Why do you play tennis in college?

K. I ~~want~~ *wanted* to start new sports ← I wanted to join ~~a~~ some circle and started to ~~I used to not to~~ play sports.

A. So. which do you like basketball or tennis?

K. I like basketball and tennis, but now, I play tennis *twice* a week ~~two times~~ So I ~~My~~ best sports is tennis. ×
   [I like tennis very much.]

A. ~~Do you~~ *Did* *to do* any other sports.

K. I used to ~~swimming~~ *swim*

A. Can you swim batterfly!

K. yeah. { I liked to batterfly }

116

Figure 5.1 Transcription work right page

A. Me too.
A. Can you swim well?
~~I swim is fast?~~
Can you *swim* *play* fast?

K. So so.

A. Do you have a lot of stress?

K. I don't have a lot of stress.

A. So, Don't you have some ploblems?

K. If I have problems or stress but. I forget to while sleep.
                                        I forgot for some stress while sleep.

✳ I forget the stress when I sleep

— food —
I like vegetables and snacks too.
I have a breakfast everyday any time.
I eat vegetable every day.

I love basketball and tennis, basketball is my blue spring. but ? I play tennis now. two times a week, and I want to play tennis every time. So I lover tennis very much.

✳ Kanko good job - you tried hard

✳ Next time talk about the new ideas you wrote

✳ Also talk about you eating habits

117

D5 made five corrections to her first performance transcript, which are summarized in Extract 5.0.

Extract 5.0

6D6:   *I used to plays* **play** *basketball for junior and high school.*

10D5: *I want* **wanted** *to start new sports.*

12D5: I play tennis twice **two times** a week.

12D5: My best sports is tennis **I like tennis very much**

24D5: *If I have problems or stress but I forgot to while sleep* **I forgot for some stress while sleep**

Students were told to transcribe exactly what they heard and then afterwards make corrections, and they generally followed these instructions, accurately transcribing their performances. Here D5 was largely accurate in her transcription, although she incorrectly transcribed 6D6 and 10D5. For 6D6, she misheard the presence of *s* on *play,* as it was not present in the recording. In the case of 10D5, she left out the *ed* on *start,* which was present on the recording.

D5 only managed to successfully self-correct AS-unit 10D5 in her notebook. 12D5 was not incorrect, with the change representing more reformulation than correction. At the end of class I collected the notebooks and made additional corrections on points students missed. This included a number of minor corrections to D5's work in red ink, but I also missed a number of errors. For example, I did not correct: *I liked to batterfly, I eat vegetable everyday, I have a breakfast anytime.* In addition I only corrected half of the reformulated unit *...but I can't (couldn't) continued because...* where I missed the *ed* error. Such misses were unfortunately common throughout my

118

feedback on her different task performance transcripts and by extension were likely common on my comments on the other students' work as well.

D5's transcripts reveal her most common errors involved tense, subject verb agreement, and grammatical words (prepositions, determiners, articles, pronouns, and auxiliary verbs). The selection, absence, or unnecessary presence of grammatical words was the most common type of error for all students. Grammatical words are notoriously difficult to learn and as they do not impede meaning are also widely recognized as resistant to corrective feedback. Despite this, it remained to be seen if the proximity of the second performance to the first and the individualized nature of the self-focus on form by writing and correcting the transcript combined with personalized feedback from me as the teacher would result in improvements in her accuracy in the second task performance. However, one issue in comparing first and second task performances is that overall there was not a lot to correct, as D5's transcripts reveal, which was the norm in this study. Of the eight corrections I did make, three were from reformulated work and not the first performance. This lack of language to correct may be one contributing factor to the lack of significantly different results between the transcription and reporting groups in this study.

5.0.1 Substantial errors

I also provided feedback on substantial errors, where a hole or gap in student interlanguage was revealed. Normally students produced only a few of these errors per task performance. 24D5 was D5's only substantial error in her first task performance, with her self-corrected language in bold below:

119

Extract 5.1

24D5: *If I have problems or stress but I forgot to while sleep **I forgot for some stress while sleep***

I wanted to focus on these more substantial errors in my feedback for two reasons. First, I reasoned they would be more regularly noticed and acted upon by the students, which would lead to reattempts in subsequent performances. Second, these errors provide specific interlanguage problems to focus on in class. I hoped the saliency of the errors combined with my addressing them explicitly would increase the chances students would use the corrected language in subsequent task performances.

With 24D5, D5 only attempted to self-correct the latter half of the AS-unit. She might have wanted to say something like: *If I have problems or stress I forget about them while I sleep.* Here I had a fairly clear idea of what she was trying to say, although in other cases it was common for me to have to correct incomprehensible AS-units. This problem of the comprehensibility of students' intended meaning was just one of a number of issues I faced trying to provide feedback on these errors in students' notebooks. Not being able to talk to the students when I provided my corrections meant I could not be completely certain what they were trying to say or how much they were aware of the language issues they were having. This lack of direct contact with students during composition of my feedback made it difficult to decide what to correct and how to correct these errors in many cases. Looking at the example above of what she might have wanted to say, I needed to address a number of interconnected language points. The problem with these errors was that they usually did not represent one language problem, but rather multiple language issues within one AS-unit. Even when there was one gap in intended meaning, it was comprised of a mix of

120

language issues. Unfortunately I did not anticipate this problem when designing this research.

Trying to address all the language issues with these utterances would have required considerably more space in students' notebooks to provide examples and explanation, and it was not clear whether students would understand these examples. Considering the constraints of the notebook medium and the complexity of these errors, I opted for corrections that were slightly more appropriate or more accurate than the students' without going into considerable detail about the nature of the corrections, as I felt this would increase the chances students would employ the corrections in subsequent task performances. Thus in correcting 24D5 I suggested: *I forget the stress when I sleep.* There are a number shortcomings with this suggestion; if you look at the transcript, it does not address the whole AS-unit and the conditional relationship between the two clauses. Moreover, I substituted *when* for *while,* which could be confusing for D5. Nevertheless this minimalist strategy for providing suggested changes seemed the most practical option to follow, as it was not clear that more detailed feedback would benefit the students more.

5.0.2 Summary of self-correction and teacher correction

In summarizing the correction part of the transcription process, I have argued my corrections in students' notebooks were often haphazard and perhaps not in forms students could utilize in subsequent performances. This is because there were few corrections for me to make, partly because the first performances had the least amount of language, and also because the most common type of error was grammatical, the least resistant to quick treatment. Finally, the feedback provided was extensive, which

121

may have made it of limited value for the students since extensive comments can be overwhelming—students literally have to focus on everything at the same time.

Furthermore, more substantial errors created problems in terms of providing corrective feedback because in many cases students' intended meaning was incomprehensible and where I could infer intended meaning, the language issues involved were often multifaceted. Finally, providing appropriate feedback within the medium of students' notebooks and without talking to them directly was problematic. These problems often resulted in feedback that was less than ideal, as illustrated above.

5.1 Student reformulations: making transcripts more complex

Besides self-correcting, students were asked to reformulate or add more detail to their transcripts, following Lynch's (2001) concept of reformulation, which involves producing subordination and using grammatical metaphor. Unfortunately, this more advanced reformulation work was not appropriate for my students because they tended to rely on clausal chaining to extend their discourse, a common strategy for less proficient learners (Norris and Ortega 2009). Thus for my students making their transcripts more complex involved adding rather than inserting, or parataxis instead of hypotaxis. To highlight D5's reformulations I have enclosed them in blue brackets in Figures 5.0 and 5.1. When D5 made her additions she used these three symbols: $\Leftarrow$ + <. 12D5 is where she made the largest addition to an AS-unit (see also Figures 5.0 and 5.1.).

Extract 5.2

11P: So which do you like basketball or tennis?

12D5: /{I}{5.55}[I like basketball and tennis but]{3.79}[now I play tennis twice a week so]{I}{3.07}][my best sports is tennis]/(19)**(34.19)**

Added information

I love basketball and tennis. Basketball is my blue spring but now I play tennis two times a week and I want to play tennis every time. So I lover tennis very much.

D5 added more information to clarify why she likes tennis, but the challenge for her was retaining the additional language in working memory (for an explanation of this term see Baddeley: 2007) while waiting for the opportunity to use it in her second performance of the task, which was difficult for most students to do. Unfortunately when planning this research I had not realized expanding utterances would be so challenging for students.

I provided general non-language related advice regarding what students should try to expand on from their first performances (the pink starts in Figures 5.0 and 5.1), which were intended to orient students to talk about topics they had missed, avoided, or should explore more. I would also write follow-up questions where necessary to facilitate topic exploration. For example, if a student said *I sometimes use online shopping*, to encourage her to expand this AS-unit, I would write a follow-up question such as, *What do you buy online?* This encouragement to expand answers was further reinforced through whole class input and feedback in the subsequent lesson, the topic of the next section.

5.2 Whole class input and feedback session

Before the second performance I conducted a whole class language input and error feedback session for about 15 minutes at the start of the lesson. Students also listened to and read a scripted dialogue of me being asked the same questions on the topics for the different sessions. The contents of that listening exercise are reproduced in Figure 5.2 below.

Figure 5.2 My sample performance of the task session topic

Is your lifestyle healthy or unhealthy?

I think it is both. I usually go for a walk each night for 30 minutes after dinner. It is very relaxing and it helps me release my stress.

Do you walk alone?

Yes, I usually go for a walk alone because my wife works, but sometimes we go for walks together.

What else is healthy about your life?

Well I eat lots of vegetables and fruit everyday when I am at Osaka Shoin and I don't each much fried food or meat. I love vegetables. My wife also cooks very healthy Japanese food. Her father is a fisherman and grandfather grows delicious vegetables for us. His vegetables taste different from the supermarket.

Is your wife a good cook?

Yes she tries to cook new dishes each week and cooks a lot of fish. When she cooks a lot I lose my weight. She often uses Bistro SMAP's cookbook.

124

What about the healthy part of your life?

Well I don't smoke and drink only a little amount of alcohol. My problem is I love snacks and when I go for a walk I often buy snacks like ice cream and chocolate. There are convenience stores everywhere!

So going for a walk is not healthy!

Yes, I try not to buy snacks, but they are my weak point. If my wife sees me eating snacks she gets very angry. I am trying to stop, but it is difficult. I think that my life is mostly healthy. I ride my bicycle, drinks lots of green tea and do iaido, and take vitamins. Plus I try to control my stress by playing video games and relaxing, and I always sleep 8 hours a night. I like going to be early.

Do you play a lot of sports now?

No I don't I just do iaido twice a week. I want to try Kyudo in the future.

Did you used to play a lot of sports?

Yes in Canada young people play lots of sports. I used to play rugby, soccer, and lacrosse in the spring and summer. In the winter like most Canadian boys I played ice hockey. I used to be really good at all the sports. My favourite sports were rugby and hockey. My father in the winter used to make a skating rink in our backyard and I used to play hockey with all my friends. My younger brother used to play with us and we would always fight.

In performing the task I aimed for comprehensive language with follow-up questions in order to encourage their use among the students, particularly since use of follow-up questions resulted in better task production throughout the different task performances. Once students had read and listened to this example, I briefly highlighted useful language points from it. It was rare for students to ask questions or show interest in this stage of the class.

For the whole class corrective feedback session I highlighted common student errors or mistakes by providing positive evidence. These errors and mistakes were mined from the students' notebooks and presented on an overhead projector. Figure 5.3 includes the language points from D5's class's first task performance.

Figure 5.3 Feedback for D5's class on first task performance

I want to **exercise**. I want to **exercise** more. I **exercise** twice a week. I need to exercise. I don't

exercise enough/very much. What kind of exercise do you do?

Do you exercise? I don't exercise enough.

Do you eat vegetables? I don't eat many vegetables.

What **junk food** do you **like**? I like chocolate.

What **food** do you **dislike**? I **dislike** sushi. I **hate** sushi. I **don't really like** sushi

Do you eat a lot of vegetable**s**? How often do you eat vegetable**s**?

I eat **a lot** of vegetable**s**. I eat **a lot** of fish

I don't eat **many** vegetables. I don't eat **much** fish.

How **much** fish do you eat in a week. How **many times** do you eat **fish** in a week? I eat fish four times a week.

I eat **too much** junk food

I eat **too many** potato chips

126

Did you have/eat breakfast yesterday? I **had/ate** cornflakes

When **did** you go to bed last night?

Did you have breakfast today?

What sports **did** you **used to** do? I **used to** play soccer. I **used to** belong to the tennis club.

I go to bed **late**. I get up **late**. I eat snacks **late at night**

I eat snacks **twice a week/ four times a week/ five times a week**

As I presented positive feedback visually, I verbally provided negative evidence of common incorrect forms from the first performance. For example, Figure 5.3 includes *I each too much junk food* and as that phrase was displayed I pointed out many students said *I eat many junk food*. When necessary I also provided Japanese translations of key words and phrases.

Both teacher-directed sessions may have complicated the study by introducing other variables; however, because this study was classroom-based it was not possible to leave out learning phases for experimental purposes. In particular, the reporting group did not receive corrective feedback outside of this whole class feedback, so removing this feedback session would have been pedagogically problematic, even if it may have made the experiment more robust. Nevertheless my feeling is these feedback sessions had little impact on students' subsequent performances, particularly since they tended to lack interest in the input and feedback sessions, generally not taking notes or asking questions, an issue I have explored previously (Moser 2005).

5.3 Second task performance

In the original trialling of the study a repeat task performance was supposed to occur only once before a final performance three to four weeks later, as the research design originally called for only three performances instead of four. However, it became apparent during initial trailing that in the follow-up to the first performance students were not focusing on prior corrections or reformulations with their new partners. Since the second performance was meant to be a kind of language consolidation phase, in order to encourage students to attend to their prior work, I felt it necessary to designate one task performance for this purpose. This would check if in fact students would focus on the correction work in the study's design, a key question for this research. Thus students were told that in the second performance they were to retry their first performances, incorporating corrections and changes from their notebooks. In order to ensure they did this, students performed the second performance with the same partner from the first.

5.4 D5's second task performance

D5's second performance totalled 139 words, one word short of double the total word count of her first performance. Her word count of AS-units seven words and over was the same as for the first and fourth performances. Her average of number of words per AS-unit (6.3) was the lowest of the four performances. Her average length of run was only 6.0 words, the lowest of the four performances. As in the first performance, D5 did not use very much Japanese. The transcript of this second performance is included in Extract 5.3, below.

Second task performance

1D5: /My lifestyle is healthy/And a little unhealthy/Because I like junk food very much/Especially I like potato chips and {4.41} juice/{I I eat junk food}[I eat junk food {for}]{7.39}{more}[more than four times a week]/And I drink juice everyday/(4+4+7+6+11+5)

2P: What flavour of juice do you like?

3D5: /{I} I like fruit's juice/Mixed juice and fresh juice/(4+5)

4P: /100 percent/ (in Japanese)

5D5: /So {I I stopped} I stopped eat junk food but {I} I can't stop eat junk food/{But} but {I}{3.34} I think healthy life because {I have a} I have a breakfast anytime/I have a vegetables and fruits/(14+6+6)

6P: What vegetables do you eat for breakfast?

7D5: /{I} I eat breakfast is Japanese food/Miso soup and rice/So miso soup is in very many vegetables/{Onion} onion potato carrot/(6+4+8+3)

8P: What fruits do you eat for breakfast?

9D5: /{I eat necessary} I eat for strawberry/And my mother's friend send to {3.11} orange {4.94} full box/(4+9)

10P: Mikan? (orange? in Japanese)

11D5: /Hasaku/

12P: I like

13D5: {So I} I eat hasaku {every} after dinner/(5)

14P: Do you exercise?

15D5: /{I} I play tennis {for two week} for two times a week/And yesterday I play basketball in {Sho} Shoin basketball club/(8+9)

129

16P: Who do you play tennis with?

17D5: /{I I play} I joined to Kinki University and Shoin University's tennis circle/So

   circle member these days/(10+5)

The first salient feature of D5's second performance is the monologue at 1D5, consisting of six AS-units and comprising one minute twenty-seven seconds, 20% of the task time. Mini-monologues at the start of a task performance were common in the second and third performances where students tried to employ a large amount of memorized language on single questions from their pre-task notes or reformulation work, an observation reinforced by the fact that monologues always occurred at the start of performances. This seems an effective strategy as only a limited amount of pre-planned discourse can be held in students working memory at one time (Bygate and Samuda 2005) so it needed to be employed in the first few minutes of a performance before it was forgotten. Students also occasionally produced mini-monologues when they wanted to avoid doing the task, particularly at the start of the semester, although thankfully by the end of the semester this practice stopped.

The monologue 'technique' was how students incorporated large amounts of reformulated language on a single question into a repeat task performance. Very few students were able employ substantial portions of their reformulated work without reading from their notes. Students approached reformulated language in four different ways; first, many students memorized portions of their reformulated work on a specific question and produced it as a monologue near the beginning of a task performance. A second approach, the most common, was to not use the reformulated language at all, a tendency which correlates with a decrease in transcription reformulation work by students as the semester progressed. Third, they sporadically

recalled chunks of language. Finally, they read their reformulated work from their notes, also a very common approach. For many students trying not to read their notes became a self-stated goal for them during the semester.

D5's monologue at the start of the performance (1D5) was not from her reformulation work, but rather her pre-task notes. However, at 5D5 she produces another monologue consisting of three AS-units plus one read unit. The final two AS-units were from her reformulated work. D5 indicated in her notebook by underling them in pink that she read these two AS-units. The recording does not suggest that she really read them, nor does all that she underlined in her notebook appear verbatim in the performance. Thus I concluded that in her notebook she should have only underlined the AS-unit underlined in Extract 5.3. While it was rare this was one example where I concluded that the data did not support her assessment and thus I permitted it for codification. The reason why she underlined more than she should have may be because she looked at her notes during the task performance and was just trying to be as honest as possible even to the extent of downplaying how much she did in fact speak. As discussed students were often overly conservative in acknowledging spoken production. If they were unsure about whether they read or spoke a portion of their production they overwhelmingly would identify it as read. One of the accompanying characteristics of a task performance with a lot of memorized language from pre-task notes and reformulation work was a high number of error-free AS-units, with the bulk of these at the start of the dialogue. D5 is no exception, with seven of her 12 error-free AS-units in the first eight AS-units. Of these, six belong to the monologue at 1D5, and of these six, five are based on her pre-task notes, with the first verbatim from her notes and the second a slightly varied version of the first AS-unit.

131

The next three AS-units refer to junk food, a term D5 used twice in her pre-task notes in addition to writing she likes potato chips. The expression of frequency was from her pre-task notes as well. The final AS-unit was the only one not based on her notes, although *everyday* does appear in her pre-task notes.

The second performance was D5's most accurate performance for the task session, with 12 of 22, or 55% of her AS-units error-free. Four of these were seven words and over. A high count of error-free AS-units often involved students memorizing them from their notes and then employing them at the start of their task performance. Another main feature was students using simple verbs such as *like* or *have* in short AS-units. I concluded from these three features that students who had very accurate performances, drastically more accurate than their previous performance (as in D5's case), were employing language they were familiar with and therefore arguably partially proceduralised. This means they used language they could memorize or recall because it was familiar to them and thus already part of their interlanguage. This familiarity probably included knowing the language was grammatically correct, meaning these drastic increases in accuracy were not the result of improved syntactic processing, but rather being able to access or recall more familiarized language. In contrast, new AS-units untraceable to previous work did not involve simple verbs, were generally error-filled, and tended to emerge through follow-up questions.

In addition to the monologues, most of D5's individual turns had at least two AS-units while in the first performance there were only two turns with more than one AS-unit. Except for one turn, all her other turns had at least one AS-unit over eight words. While not 'within AS-unit complexity' these turns with multiple AS-units

demonstrated output pushing through clausal chaining. In this paper I refer to within AS-unit complexity as a single AS-unit that contains some form of clausal integration. Besides an AS-unit just having a single clause present, within AS-unit complexity can be also contrasted with 'between AS-unit complexity,' where the complexity for the latter is the result of separate AS-units chained together, but still cognitively separate (see Chapter 2 for a more detailed discussion). D5's commitment to complexity was not just through clausal chaining, but was also demonstrated through grammatical risking taking. For example at 9D5, D5 struggled with both AS-units (Extract 5.4).

Extract 5.4

9D5:   /{I eat necessary} I eat for strawberry/And my mother's friend send to {3.11} orange {4.94} full box/(4+9)

Both AS-units in Extract 5.4 included new meaning, with the second involving risk taking to the point where there was a substantial gap in D5's interlanguage. This is consistent with other dyads, where when students produced new meaning it almost always resulted in substantial errors.

5.5 Carry-over from the first performance

The final question of the second performance regards what D5 carried over from the first performance. Students were told to improve on their first performance, which meant acting on self and teacher error corrections and incorporating reformulated work into their performances. Despite this stated focus, D5's answers in the second performance were different from the first. This is partially my responsibility because I gave her mixed messages, suggesting in her notebook that she should talk about her diet as well, which ended up as her main topic in the second performance. In terms of

133

correction none of the AS-units corrected in her transcript were used in the second performance. The only carry-over is included in Extracts 5.5 (first performance) and 5.6 (second performance), below, a clause D5 first wrote in her pre-task work. In this sense it was not language carried over from the first performance and then reworked in the second performance; it was simply a verbatim performance of language originating from her pre-task work.

Extract 5.5

11P: So which do you like basketball or tennis?

12D5: /{I}[I like basketball and tennis **but now I play tennis twice a week** so {I}{3.07}][my best sports is tennis]/(19)**(28.26 seconds)**

Extract 5.6

14P: Do you exercise?

15D5: /**{I} I play tennis {for two week} for two times a week**/And yesterday I play basketball in {Sho} Shoin basketball club/(8+9)

16P: Who do you play tennis with?

Except for this carry-over answer about tennis, none of D5's second performance specifically related to production from her first performance; however, as we saw with the monologues, D5 did use specific language from her pre-task notes and reformulation work in her second performance. Thus D5's second performance was quite different from her first, typical of most students, especially the more proficient ones.

5.6 Summary of second performance

D5's second performance was consistent with the central tendency of the quantitative data in that on CAF measures her second performance was better than her first performance. She engaged in output pushing through grammatical risk taking and clausal chaining. She also employed memorized chunks of language from her notes, producing them as single AS-units or in chains as monologues. Because the chunks were from her pre-task and reformulation work, I would interpret that this language was familiar to D5 and thus partially proceduralised, accounting for her improvement in accuracy in her second performance.

D5 was productive in her second performance, but did not follow the pedagogical aim of the second performance. Outside of employing some reformulated work, she did not attempt to improve on the language used in her first performance, which would have involved repeating the same themes from her first performance and trying to employ corrections from her notes. Rather, she talked about different themes from her first task performance, for which she even instigated a monologue at the start of her second performance. My comments on her transcript may have partially oriented her to explore new themes in her second performance. Nevertheless, there is little continuity between her first performance and transcription work and the second performance. Like D5, most students who engaged with the task work did not follow the intended pedagogical focus of the second performance or de-prioritized it. They were in most cases more proficient students who had the ability to pursue new information content. The few students who tried to improve on their first performances did not tend to improve and in many cases their performances were

shorter and less complex. These students tended to be less motivated students or committed to following teacher instructions.

It was surprising to see the quantitative results for the second performance were so much higher than the first. As discussed above there were a lot of problems and disappointments with the transcription work and task performance, and I felt all the students struggled here, partly because they seemed to have conceptual trouble understanding the purpose of the second performance. There was uncertainty among pairs about the focus of the second performance; I often saw students checking with each other about the task instructions and I regularly had to explain what students were supposed to do. This is not surprising as this was the only form-focused performance of the task sessions.

There are at least two possible reasons for the struggle and confusion. First, with task performance encodings become "semantically reduced" (Widdowson 2003: 118), which can lead to boredom. This repetitiveness was compounded by the fact that students had the same partners. Thus perhaps D5 avoided boredom and disengagement by talking about different topics. In fact, students who repeated language from their first performances became disengaged and the task lost meaning. In the second performance I sensed students battled how to maintain meaningfulness in their interaction while repeating and reformulating what they had said previously. As meaningfulness was critical to successful interactions, from the students' perspective it was key to maintaining a friendly interaction more than fulfilling the teacher's intentions for the task performance. When students did follow instructions and tried to repeat and build on their first performances, meaningfulness was reduced and interaction between pairs sometimes became awkward.

I had envisioned students maintaining meaningfulness by recalling and expanding on the language from their previous performance by including ideas from their reformulation work, one of the key reasons for introducing transcription in the first place. However, the idea of having students recall previous work and expand on it with reformulated work in a real time performance was cognitively challenging, a second reason for students' struggle and confusion during this second performance. Students had to be able to retain large amounts of language from their previous performance and reformulation work in their working memory while simultaneously reworking it during a real-time task performance. They coped with this difficulty by memorizing or reading their previous work, including D5, who did a bit of both. When students did not read, memorize, or pursue a different topic, they struggled to recall in detail what they had said or prepared previously, and thus produced less language, leading to disengagement with the task. This may explain why student transcript reformulation work over the course of the study decreased. Ironically, I feel the constraints of the second performance had an impact on the third performances being the most quantitatively complex and fluent in the study, as students saw the third performance as a return to meaningfulness because they were paired with a new partner. The third performances were also free of the requirements of the second performance to revisit and attempt to expand on previous language work.

Chapter 6 The case study: third and fourth task performances

In the final chapter of the qualitative portion of this study I will analyze the third and fourth task performances. I want to investigate how productive these latter performances were and what were their chief characteristics, including how much carry-over or language was repeated from the second task performance to the third task performance, and then finally from the third task performance to the fourth performance. On all measures the third task performance had the highest means while the fourth task performances on numerous measures had the second highest mean. The main question I try to answer is what are the key features of each task performance, and how do they contribute to the results. The other question is the same question asked with the second performance and that is how much carry-over occurs in these latter task performances? This question is important for the fourth task performance since it occurs three to four weeks after the previous task performance. In addition the immediate proximity between the second and third task performances begs the question how much carry-over occurred.

In this chapter I will first analyze the third task performance focusing on its main features which I argue mostly involved adding new content or expanding on the same theme. I will also analyze how much language was carried-over from the second task performance. For the fourth task performance, I will analyze the main features including demonstrating that the issue of struggling to speak remerged as a feature in many fourth task performances. I will end this by chapter by briefly summarizing the qualitative portion of the study.

6.0 D5's quantitative measures for the third performance

D5's key complexity measures and fluency measures were strongest in the third performance. 12 of her 17 AS-units were seven words or over, 144 words out of a total 167 words. These particular complexity results are at the top end of all student performances. Her average number of words per AS-unit was 9.8, which was well above the seven-word AS-unit benchmark. As will be discussed (section 6.9) there were also a large variety of verbs produced in the third performance relative to other performances. Her accuracy measures in the third performance were either the highest or second highest of the task session. Regarding accuracy, six of her 17 AS-units were error-free and three of the six were seven words and over, for a total word count of AS-units seven words and over of 40 words, her highest for this task session. Her fluency measures were the highest in her third performance with an average length of run of nine words.

6.1 D5's third performance: repeating her monologue

D5 repeated her monologue from the second performance on eating habits at the start of her third performance, although the language she used varied slightly. Extract 6.0 includes the monologue from her second performance and Extract 6.1 reproduces D5's total third performance (her monologue is at 2D5, in bold).

Extract 6.0.

1D5: /My lifestyle is healthy/And a little unhealthy/Because I like junk food very much/Especially I like potato chips and {4.41} juice/{I I eat junk food}[I eat junk food {for}]{7.39}{more}[more than four times a week]/And I drink juice everyday/(4+4+7+6+10+5)

Extract 6.1.

1P: Is your lifestyle healthy or unhealthy?

2D5:   /My lifestyle is maybe healthy but a little unhealthy because {I} I like junk food/And {I I have a junk food}[I love junk food very much {I} I eat junk food]{3.97}[more than {two} four times a week]/And I drink juice everyday/(14+17+5)

3P: What kind of junk food do you like?

4D5:   /{I} I like chocolate and potato chips/{It's} it is {li} little unhealthy but I can't stop junk food/[But {I} I think my lifestyle is maybe healthy because {I} I have a breakfast anytime and I have vegetables everyday]{and} and vegetables and fruits is my favorite food]/(6+10+27)

5P: What kind fruits do you like?

6D5:   /{Necessary I necessary necessary} especially I like orange/My mother's friends send to hasaku full a box/So {I} I eat hasaku every after dinner with family/(4+9+9)

7P: What is different between orange and hasaku? (In Japanese)

8D5:   /{Orange is}[orange is almost mikan but hasaku is]{5.02}[taste grapefruit and orange]/(11)

9P: Do you like to play sports?

10D5: /Yeah {I like} I like watching sports and I play sports/{I} I play sports every two times a week/{I} I join to tennis circle for Kinki University and Shoin University {10.00} combinations circle/(9+8+13)

11P: Where do you play?

140

12D5: /Tennis/{I play tennis court is I I} we play tennis court in near the Kinki University's tennis court/{It's} it's {in} indoors tennis court/So rainy's day we can play tennis/(11+5+7)

13P: What do you like sports?

14D5: /I like basketball/(3)

D5's monologue in her third performance, unlike in her second performance, was produced in response to a question. Additionally, the average number of words per AS-unit varied between the two monologues. The 14-word AS-unit in the third performance monologue was derived from the first three AS-units in the second. Also, the 17-word AS-unit in the third performance monologue included verbatim performance of the 10-word AS-unit from her second performance. These longer AS-units mean D5's monologue in the third performance was more complex according to the constructs used for this research, although I feel more fluent would better describe the difference between the two monologues as there is not a substantial difference in linguistic complexity between them. Instead D5 processed the same language faster, which meant shorter AS-units were now codified together to produce longer AS-units. D5's production of larger AS-units in the third performance did not come at the expense of accuracy, as both the second and third performance monologues are error-free. This is of note because while the second performance monologue was likely memorized, and possibly held in the working memory prior to producing it, the student in third performance monologue did not have access to her notes nor would have been able to retain the monologue in her working memory for such an extended period of time. For this reason the third performance monologue may have in relation to the second task performance involved more challenging retrieval processing.

The immediate proximity of the second and third performances made carry-over between the two more common than with other performances. Most carry-over was verbatim-like performance of previous language, similar to D5's performance of her two monologues. In most cases there was no increase in linguistic complexity or accuracy. Instead the most common feature was larger AS-units created from smaller AS-units, with coordinators and subordinators going from sentence starters and enders between AS-units to serving integrating functions within longer AS-units.

6.2 D5's third performance: reworking content

D5's third performance had one example of reworking of prior content. In the second performance at 5D5 she said, *So {I I stopped} I stopped eat junk food but {I} I can't stop eat junk food*, which was reworked in her third performance at 4D5. The reworking of this content centered on her like of junk food and how she cannot stop eating it. In both turns 5D5 in the second performance and 4D5 in the third performance she progresses from talking about junk food to her breakfast and then vegetables. This same basic sequence makes it clear that 4D5 in her third performance is a repeat of 5D5 from her second performance. Below are the extracts.

Extract 6.2. Junk food: second performance

2P: What flavor of juice do you like?

3D5:   /{I} I like fruit's juice/Mixed juice and fresh juice/(4+5)

4P: /100 percent/ (in Japanese)

5D5:   /So {I I stopped} I stopped eat junk food but {I} I can't stop eat junk food/{But} but {I}{3.34} I think healthy life because {I have a} <u>I have a breakfast anytime</u>/I have a vegetables and fruits/(14+6+6)

In the third performance (Extract 6.3) she starts with *{I} I like chocolate and potato chips*, a repeat of 1D5 from her second performance; *Especially I like potato chips and {4.41) and juice* (Extract 6.2). This AS-unit can be traced back to her pre-task notes: *My favorite junk food is potato chips and chocolate.*

Extract 6.3. Junk food: third performance

3P: What kind of junk food do you like?

4D5:   /{I} I like chocolate and potato chips/{It's} it is {li} little unhealthy but I can't stop junk food/[But {I} I think my lifestyle is maybe healthy because {I} I have a breakfast anytime and I have vegetables everyday]{and}[and vegetables and fruits is my favorite food]/(6+10+27)

The next AS-unit at 4D5 is *{It's} it is {li} little unhealthy but I can't stop junk food*. The first part, *it's a little unhealthy* is new language not present in prior performances or language work. The pronoun at the start creates a cohesive link between the two AS-units at 4D5. In addition the clauses in the second AS-unit were cohesively linked with a conjunction *{It's} it is {li} little unhealthy but I can't stop junk food*. These slight improvements in cohesiveness do not appear to be occurring randomly. Instead, the relative complexity of the cohesiveness seems to suggest the changes involved conscious thought by D5. This contrasts with the rest of the content between the two extracts where the language in her third performance was mostly verbatim from her second performance. The changes D5 made with regards to junk food between the two performances stand out from anything else she did in all four task performances for this task session. It represents the only example I could find of a clear improvement over her previous performance that was arguably deliberate, and not just

simply the result of variation caused by the fact that no two task performances can be exactly the same. It appears D5 at some point before or just after the first AS-unit in the third performance knew she wanted to repeat what she had said previously about junk food and remembered what she had said previously was contradictory or confusing. In fact in the second performance at the end of AS-unit she produced a hesitator in Japanese expressing uncertainty about what she had just said. This reworking of language from earlier performances was uncommon in the research, with the exception of reformulations (self-corrections, replacements), which constitute one-off, real-time lexico-grammatical improvements.

6.3 D5's third performance: adding new content

Adding new information, similar to other students, was a key feature of D5's third performance. D5's main expansion of meaning was on the topic of sports, in particular tennis, which she talked about from the first through the third performance, illustrated in Extracts 6.4 to 6.6. In the first performance (Extract 6.4) D5 explains tennis is her favorite new sport, why she started it, and how often she plays it.

Extract 6.4

Shared Silence 14.52

1P: How often do you exercise?

2D5:   /{I}[I play tennis]{5.70}{two}[two times a week]/(7)**(15.00 seconds)**

3P: Only tennis?

4D5:   /Yeah/

5P: Then what sports or exercise you used to do? Tennis or (in Japanese *anything else?)*

144

6D5: /[I used to {used to?} play to basketball]{2.40} for {2.56}[junior high school and high school]/(12)**(21.70)**

7P: So six years

8D5: /Yeah/

9P: Why do play tennis in college?

10D5: /{I I started I want to start}{6.03}{I} I want to started new sports (in Japanese *new*)/(6)**(23 seconds)**

11P: So which do you like basketball or tennis?

12D5: /{I}[I like basketball and tennis but now I play tennis twice a week so {I}{3.07}][my best sports is tennis]/(19)**(28.26 seconds)**

In the second performance D5 repeats that she plays tennis twice a week, adding at 17D5 (Extract 6.5) that she joined a joint university tennis circle.

Extract 6.5

14P: Do you exercise?

15D5: /{I} I play tennis {for two week} for two times a week/And yesterday I play basketball in {Sho} Shoin basketball club/(8+9)

16P: Who do you play tennis with?

17D5: /{I I play} I join to Kinki University and Shoin University's tennis circle/So circle member these days/(9+5)

While not carried into the third performance, she added new information in her second performance about playing basketball at 15D5 (Extract 6.5). In the third performance she includes previous content with some variation about how often she

plays sports. The new content she adds at 12D5 (Extract 6.6) is in relation to where she plays tennis and what type of facility it is.

Extract 6.6

9P: Do you like to play sports?

10D5: /Yeah {I like} I like watching sports and I play sports/{I} I play sports every two times a week/{I} I join to tennis circle for Kinki University and Shoin University {10.00} combinations circle/(9+8+13)

11P: Where do you play?

12D5: /Tennis/{I play tennis court is I I} we play tennis court in near the Kinki University's tennis court/{It's} it's {in} indoors tennis court/So rainy's day we can play tennis/(11+5+7)

D5 made an effort at each performance (including the fourth) to include new information on this centre of interest. The challenge of this risk taking means very few or often none of the new AS-units are error-free across performances.

6.4 Repeating substantial errors

The final point to address regarding D5's third performance concerns her repeating an AS-unit with a substantial error from the second performance to the third, a common feature of many students' third performances. Perhaps because of the proximity of the two performances to each other, when substantial errors were repeated in a task session it was mostly between the second and the third performances. In Extracts 6.5 and 6.6, for example, D5 says *I join to* in both performances, which was also repeated in the fourth performance. A second, more substantial error, is included in Extracts 6.7

146

and 6.8. She first makes the error at 9D5 in the second performance (Extract 6.7), and then repeats the error at 6D5 in the third performance (Extract 6.8).

Extract 6.7

8P: What fruits do you eat for breakfast?

9D5:   /{I eat necessary} I eat for strawberry/And my mother's friend send to {3.11} orange {4.94} full box/(4+9)

10P: Mikan? (orange? in Japanese)

11D5: /Hasaku/

12P: I like

13D5: {So I} I eat hasaku {every} after dinner/(5)

Extract 6.8.

5P: What kind fruits do you like?

6D5:   /{Necessary I necessary necessary} especially I like orange/My mother's friends send to <u>hasaku</u> full a box/So {I} I eat <u>hasaku</u> every after dinner with family/(4+9+9)

7P: What is different between orange and hasaku?(In Japanese)

8D5:   /{Orange is}[orange is almost mikan but hasaku is]{5.02}[taste grapefruit and orange]/(11)

D5 probably wanted to say *My mother's friend sent (us) a full box of oranges.* The repeated error suggests D5 was unable to produce the grammatical construct *subject + verb + indirect object + direct object,* which represents a gap in her interlanguage. Because of the prevalence of such substantial errors in the third performance, I

corrected them after students submitted their notebooks, a step which was not part of the original research plan, but I felt providing feedback on these was necessary. Figures 6.0 and 6.1 contain my final feedback to D5, with Figure 6.0 including my specific feedback on her error from extract 6.8.

Figure 6.0 Student survey comments and teacher feedback on third task performance

Figure 6.1 Teacher feedback on third task performance
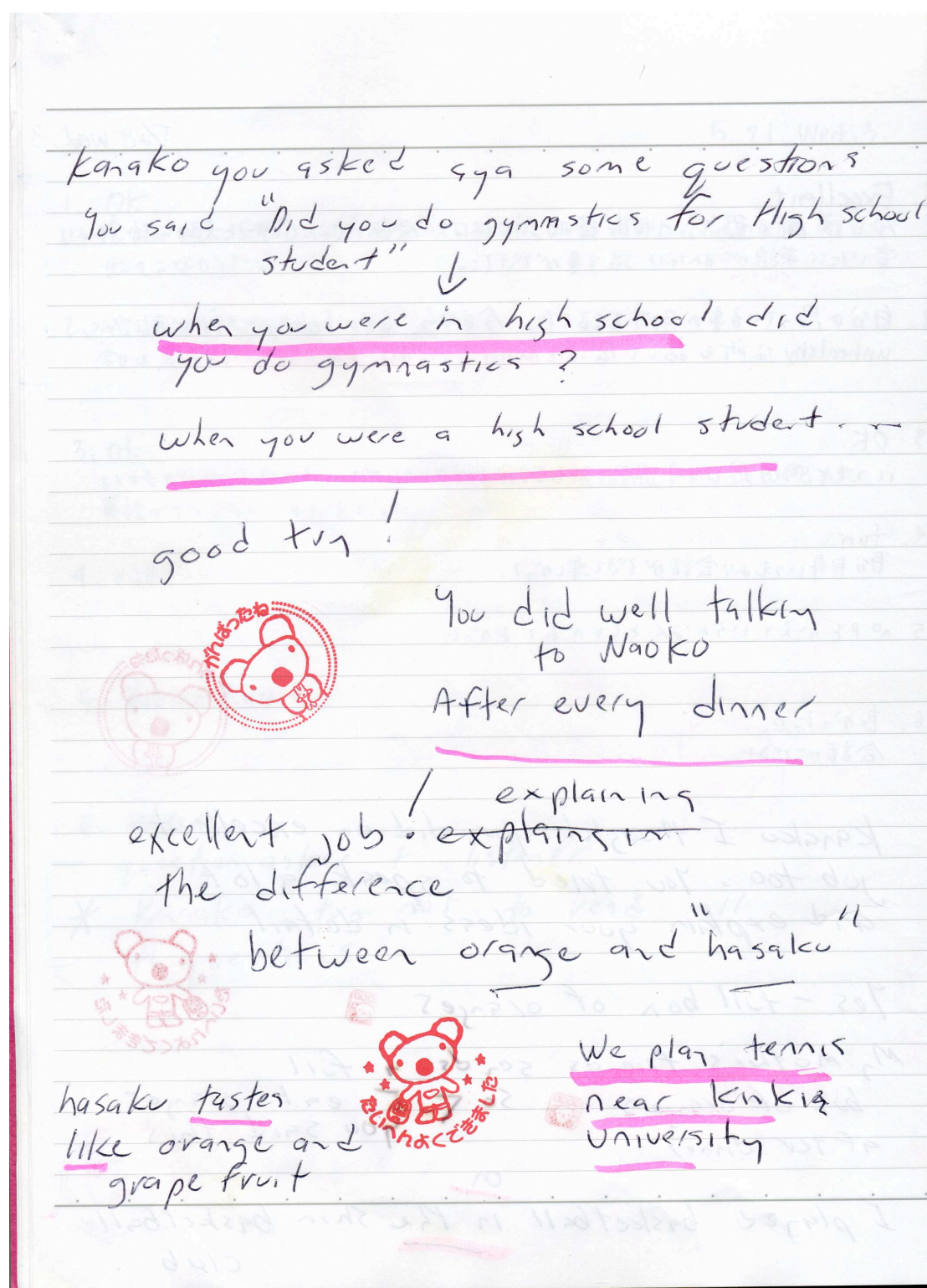
My feedback in the figures is messy and includes two mistakes (see bottom of figure 6.0) on my part; the plural on *friend* and *sends* rather than the correct *sent.* These mistakes highlight how difficult it is for a teacher working through a lot of students' work in a short space of time to pick up every error. Additionally, there was no guarantee students would act on the feedback and in most cases they did not, as was the case here; D5 did not produce the original incorrect or corrected language in her fourth performance.

6.5 Summary of third performance

As the quantitative data showed, the third performances were the best performances in all key CAF measures. D5 was no exception, with her most complex and fluent language occurring here, although her number of error-free AS-units decreased. One feature of D5's third performance was her verbatim repeating of language from her second performance, such as in her repeated monologue. By repeating language, D5 chained smaller AS-units together to form larger AS-units, a common feature of most students' third performance language. This was perhaps common because of the proximity of the second and third task performances. Creating longer AS-units did not so much represent an increase in linguistic complexity as an increase in fluency. Complexity as measured quantitatively increased in that clausal chaining occurred within AS-units in the third performance rather than between AS-units as in the second performance.

In her third performance D5 also reworked content, a tendency of the more proficient students. However, outside of self-corrections, deliberately reworking language from earlier performances was rare and if it did occur it was in this third performance,

perhaps stemming from the proximity of the second and third performances. In D5's case, there was only one example of this reworking across all four performances in this task session and it occurred in this third performance. The reason for this lack of recycling language across performances may relate to the difficulty of recalling previous performances in sufficient detail to be able to reuse the language in a deliberate and productive manner. If students could not recall large amounts of specific discourse then they could not rework and improve on it.

Another characteristic of D5's third performance was the addition of new propositional content on an earlier centre of interest. If topics were expanded from one performance to the next, it was usually from the second performance to the third where proximity between the performances made it easy to do. D5 added new information from her initial task performance in both the second and third performances and as I will show later, in the fourth performance as well. Such continuity from the first performance through the fourth was rare among students. In fact, a student producing new content on a new topic untraceable to previous performances represented the majority of the increased language production for most students in the third performance. This trend continues from the second performance; students generally produced more language when discussing new themes and less when expanding on earlier themes, although D5 did not talk about any new topics in her third performance. Adding new information by expanding a previous theme or introducing a new theme came with trade-offs, as very little of the new language was error-free. Additionally, for many students the new content revealed gaps in their interlanguage and, like D5's new content, contained substantial errors. The nature of

the errors indicates they were not the result of speaking too fast or carelessness but instead emerged from the difficulty of producing new content in real-time.

Regarding accuracy, D5's third performance was less accurate than the second. However, D5's total words over the accuracy benchmark were the highest here across all four performances. However, accuracy increases in the second performance were the result of the students memorizing chunks from pre-task or reformulated work and producing them at the start of performance while third performances did not involve as much memorization; instead students recalled either what they memorized for the second performance or had said for the first time in the second performance. In the third performances accurate language generally came near the start of the performance as monologues. For example, D5's error free AS-units were all located at the start of her third performance (four of the six) and were part of a monologue. The other influence on accuracy was students reusing familiar verbs such as *like* and *have* in smaller AS-units. Finally, it was also common to see substantial errors from the second performance repeated in the third performance, a common enough phenomenon that I felt obliged to provide corrective feedback at this stage as well. This performance of errors may be due to the proximity of the two performances.

6.6 The fourth performance

The final (fourth) performance was three or four weeks after the third. For D5's class it was four weeks later. Students were told the final performance was a review and they worked with new partners, as in the third performance. Before starting students were given ten minutes to review their notes. In the trial I tried to have students do the fourth task performance without letting them review their notes but they complained,

153

so I allowed ten minutes of review time in the experiment. Students were not allowed to look at their notes during the performance (all material was cleared from their desks), except for the anchorperson who could use the questions from the pre-task sheet. I also handed out a new copy of the questions so students could not reference notes that have been written on their own pre-task question. If I had not allowed students the time to review or access to the original questions the quantitative results may have shown a decrease in the measures used. In particular anchorpersons may have struggled to ask questions without a memory guide.

6.7 D5's quantitative results for the fourth performance

The quantitative results for most students and most measures revealed that the fourth performance was always more complex, fluent, and accurate than the first and never more complex, fluent, or accurate than the third (except for dysfluent words), and on almost all key measures more complex, fluent, and accurate than the second performance. This increase in complexity, fluency, and accuracy from the first to the fourth performance suggests the previous performances influenced this performance. While D5's quantitative results do not reflect the general quantitative results of students on their fourth performances (her measures are roughly comparable to her first and second performances), her performance is still useful for exploring the results of this fourth performance in more detail and highlighting key characteristics of students' fourth performances.

D5 spoke 81 words in her fourth performance, ten words more than her first performance and close to half of her second performance. Her complexity was essentially the same as her first and second performances. Her average number of words per AS-unit was nine, higher than her second performance. Her number of

error-free AS-units was three out of nine, equal to her first performance. Only one AS-unit was over the accuracy benchmark.

6.8 Salient features of D5's fourth performance

D5 had two features in her fourth performance unique to her, but still generally reflective of the struggles many students had here. First, she had two long pauses, her longest pauses for all four performances in this task session and evidence of her struggling to speak. In turns 16D5 and 20D5 (Extract 6.9), each pause is over 17 seconds (pause times are in bold for these).

Extract 6.9

1P: Is your lifestyle healthy or unhealthy?

2D5:   /Maybe my {lifestyle} lifestyle is healthy because I almost don't dislike food/(12)

3P: /Great/

4D5:   /And {I} I have a lot of {sleepy} sleep time/(8)

5P: What do you do to stay healthy?

6D5:   /{I I have I} I have a breakfast everyday and/As as (In Japanese: I eat breakfast)/(6)

7P: /Nani?/ (in Japanese: What?)

8D5:   /As dekiru dake as/(In Japanese: 'What do I put in?')

9P: /Possible?/

10D5: As as nani/(In Japanese: What?)/

11P: /As/

12D5: /{As dekiru dake} as walk <u>as possible</u> (P: Okay okay) for station/(4)(1 minute 20 seconds)

13P: Which do you eat rice or bread at morning?

14D5: /[Almost {I have} I have rice and miso shiru at morning][{but holiday is I have}[but I have bread in holiday]/(14)

15P: Do you do any exercise?

16D5: /{Now now now}{**17.96**} now {I I}[I don't play tennis because test time but]{5.59} [usually I play tennis so two times a week]/(18)

17P: Do you join to tennis circle?

18D5: /Yes I join tennis circle for Shoin and Kinki University tennis circle/(12)

19P: How often do you eat junk food?

20D5: /Yes/I like junk food/{I like I like}{**17.87**} I like potato chips and/(4+5)

With questions 15P and 19P, D5 needed time to think about how to respond, hence the long pauses. After the first pause in 16D5, she eventually produces new content about tennis, the main theme for all four performances. With the second question, at 20D5, she does not appear to understand the question and responds by partially recalling language from her pre-task notes which has already been repeated twice, in her second and third performances.

The second feature of D5's fourth performance that is unique to her is the one minute and twenty second negotiation episode that starts at 6D5 and ends at 12D5. At the end of 6D5, D5 wants to use the phrase *as much as possible*, but is unable to produce it. She seeks assistance from her partner who offers help at 9P, but most of this episode is silence because D5 is checking her dictionary, indicated by the underlining at 12D5. Close to two minutes of task time out of the total of six minutes was comprised of the

two pauses and the stoppage, impacting D5's total production for this task performance. Regarding dictionary use, before each performance students were asked not to check their dictionaries during the performance but to instead make a note of their problem and move on. For the fourth performances they were also asked not to use their notes. However, D5's stoppage was common for students throughout the study. Analysing these long stops I often concluded they were less about language searches and more about the students taking up task time. Based on the recordings, I concluded D5 was trying to use up her task time in her dictionary search. In fact her partner's tone at 12D5 with *Okay okay* indicates some annoyance at the time D5 is spending on her dictionary search. Along with the pausing, this event indicates she is struggling with producing enough language to fill the available task time.

6.9 D5's language carry-over to her fourth performance

Seven of D5's nine AS-units from the fourth performance can be connected to prior work. Four originate from pre-task work, two from her second performance, both repeated in the third performance as well, and one from her transcription work, repeated in the second and third performances. This carried over language is included in Extract 6.10, below.

Extract 6.10

2D5: /Maybe my {lifestyle} lifestyle is healthy because I almost don't dislike food/(12)

My lifestyle is healthy. (pre-task notes, second and third performance)

4D5: /And {I} I have a lot of {sleepy} sleep time/(8)

I take a lot of sleep time. (pre-task notes only)

157

6D5:   /{I I have I} I have a breakfast everyday and/(6)

I have a breakfast anytime (reformulated transcripts, read in second performance, third performance)

16D5: /{Now now now}{**17.96**} now {I I}[I don't play tennis because test time but]{5.59}[usually I play tennis so two times a week]/(18)

I often to play tennis two times a week (pre-task notes, first, second, third performances

18D5: /Yes I join tennis circle for Shoin and Kinki University tennis circle/(12)

I joined Kinki University and Shoin University's tennis circle (second and third performances)

20D5: /Yes/I like junk food/(4)

I like junk food very much (second and third performances)

20D5: /{I like I like}{**17.87**} I like potato chips and/(5)

My favorite junk food is potato chips and chocolate (pre-task notes, second, third)

Three of D5's four longest AS-units come from the AS-units in Extract 6.10. 14D5 was one of her longest AS-units and one of two AS-units containing new information. 18D5, another of her longest AS-units, was a verbatim-like performance of her second and third performances, an AS-unit that first appeared in her second task performance. She recalled this AS-unit from memory because there was no written record of it in her transcription or pre-task work. The next two largest AS-units, 2D5 and 16D5, were both combinations of previous and new content, with the first clause of the first

158

AS-unit at 2D5 taken verbatim from her pre-task notes: *My lifestyle is healthy.* She also produced it in the second and third performances as part of her monologues. At 16D5 the second part of the AS-unit, after the pause, is from her pre-task notes and was also used in her first and second performances. 16D5 is interesting because it represents a continuation of her center of interest across all of the task performances for this task session. The first part is an account of her tennis hobby, the only theme expanded on from her first through her fourth performances.

Despite the long pauses, the four large AS-units in her performance suggest that D5 was fairly productive in this fourth performance. The fact that three of these four AS-units involved previous work suggests her prior performances increased the complexity of these AS-units. However, there is other evidence suggesting a more lexically impoverished performance relative to the second and third performances. Specifically, she used only a limited number of verbs; *join* and *be* once each, and *dislike*, *like*, *have*, and *play*. In contrast in the third performance with the exception of *dislike* she used all the same verbs as in the fourth performance and *eat*, *drink*, *love*, *can*, *taste*, *think*, and *send* in addition, along with *be* on five occasions. An argument could be made that because her performance consisted of fewer words in her fourth performance she used fewer verbs, but these verbs, especially *like*, *dislike*, and *have* are often over-used by students when struggling to speak in real-time. D5's reliance on a limited number of verbs is reflected in her accuracy results; she only produced three accurate AS-units. The eight-word AS-unit at 2D5, and then a four-word AS-unit and a five-word AS-unit at 20D5 (Extract 6.10). All three were from her pre-task notes and used the verb *like*.

159

6.10 Summary of the fourth performance

Like D5, students generally struggled to replicate the quality and quantity of the third performance in their fourth performances. Two long pauses, one major stoppage, and an over reliance on a limited number of common verbs all point to this struggle for D5. My impression of this performance is students struggled more than the quantitative data suggests. The one-month interval between task performances appears to have negatively impacted student performances. This influence of delays between task performances was also evident between the first and second performances which had a one-week interval. In contrast the proximity of the second performance to the third appeared to boost student performance.

The main characteristic of the fourth performance for most students, not apparent in D5's performance, was students introducing new information not directly related to their earlier performance language. This is partially attributable to the fact that as in the third performance, students had new partners. While the questions may have been the same, they talked about different things, with more proficient students tending to use new information rather than trying to recall language from their previous performances. It was in the fourth performance with the one-month interval between the previous performance that new information seemed to represent a larger proportion of the overall performances than in the second and third performances. However, less proficient students like D5 appeared more dependent on recycling their previous work, perhaps because of the difficulties of producing new information in real time. In D5's case this leads to a struggle to speak because she was unable to easily produce new content and could not recall in much detail her language from a month earlier.

On the positive side, D5 produced five AS-units of seven words or more in this fourth performance, with four being well over seven words. Many students' performances mirrored D5's in that they used partially proceduralized language or chunks from their pre-task notes, their reformulated work, or prior performances. They recalled this language as single AS-units or combined it with new content, with this combining of language indicating a positive benefit for task performance even after a month-long interval. In general, compared to the first performance, students in the fourth seemed to employ language more easily, perhaps because after three task performances and the accompanying reformulation work they had more language options to recall. It should be noted here that fourth performances occurred throughout the semester. Except for the very last one, the other three fourth performances were followed by a new task session with a first performance, and the first performances were the weakest of all the performances. This sequence lends support to the idea that task familiarity contributed to the results of the fourth performances. It may also be the case that a repeat of the same task at an even longer interval would result in lower complexity, fluency, and accuracy than these fourth performances, with students likely returning to their first performance levels, indicating improvements in task performances are not maintained over long intervals.

6.11 Overall summary of qualitative results

In each of the qualitative chapters there are detailed summaries. In this summary I will just briefly restate the general findings together. First, as discussed the at the beginning of Chapter 4 the analysis of the data led me to conclude that almost all students were at a very basic level of proficiency where production is a constant

struggle. This was a continuous throughout all task performance, but task repetition eased processing pressures and allowed students to improve temporarily aspects of their spoken production.   The final main finding that was discussed in Chapter 4 was the importance of the anchorperson in productive task performances.   If the anchorperson was active in asking follow-up questions then the speaker usually responded with a positive task performance. In Chapter 5 I analyzed error correction and reformulation processes.   I reported that generally their impact on latter task performances appeared to be minimal.   Related to this finding is that when students did 'carry-over' language from previous task work including pre-task work or reformulation work it was often memorized and reproduced at the start of the new task performance. It was not entirely memorized verbatim as students would produce variations of the original chunk, and in many cases attach it with a different chunk. Nonetheless as reported throughout the study there was very few examples of students deliberately reworking prior work real-time during a subsequent task performance. A final important finding is that forcing students to try repeat their prior task performance and task work language had a demotivating effect on the students which was compounded by having them work with the same partner from the prior performance.   In Chapter 6 I analyzed the third and fourth task performances. The third performance was the strongest of all task performances on all aspects of language. The key feature saw students being able to talk about new content as well as expand on previous themes. This appears to be the result of eased processing pressure that was the result of task repetition, and proximity to the second task performance, which was immediate. Simply students appear to have had better lexico-grammatical access. Because of the proximity to the second task performance and the immediate repetition third task performances also demonstrated more carry-over than

other task performances.   The fourth task performance compared to the first performance appeared to be more productive.   While students seemed to forget language from their previous task, which made their fourth performances less of the third performance, nonetheless there was carry-over present. This could be interpreted to mean that some of what they had done three weeks or a month earlier remained. The main feature though for strong fourth performances was again students talking about different content then they had previously. This was particularly the case with more proficient students. Overall the main finding of the fourth performance was with students struggling to speak. In other words they had forgotten their previous work or had greater difficulty recalling what they had done previously.

Chapter 7: Summary, discussion, and conclusion

7.0 Structure of the concluding chapter

This final chapter is divided in to three sections. In the first I will discuss the differences between my research in the second and third modules. In the second section I will discuss the limitations of this study and how it could have been improved, providing ideas for future research. In the third and final section, I discuss the classroom implications of this study, beginning by reviewing the results of weekly surveys and one end of semester survey. This information provides student opinions of their weekly lessons and the course overall. The final part of the third section offers advice on using task repetition and transcription in similar classroom contexts to mine. In addition I discuss the overall relevance of task repetition for EFL contexts, including suggestions on how task repetition can be incorporated into the larger language curriculum.

7.1 Section one: comparing results between modules two and three

This section considers why the second module research results were different from the third. As an action research study I hoped there would be a difference, but there was no assurance going into module three that this would in fact be the case. To review, my module two results indicated immediate task repetition did not lead to improvement in student language, which became the impetus for this module three research. From an action research perspective the study in module three was a planned intervention into my classroom practice to solve problems uncovered in the findings in module two regarding the use of immediate task repetition. In the study reported in module two, students' repeat task performances stayed the same as or declined in quality compared to their initial task performances. My conclusion was that this

decline in performance in later task repetitions was the result of not providing a clear pedagogical focus for the students at the repetition stage, consistent with Bygate and Samuda's (2009) warning that communication does not automatically result in learning. I concluded most students viewed the task repetitions as a chance for mingling and not for learning, as most students did not recognize the pedagogical rationale for task repetition and so student task engagement declined with increased repetitions of the same task.

In the study for module three I investigated whether a set of learning phases interlinked with task repetition would encourage students to improve their production across task repetitions. Transcribing or reporting combined with teacher-directed feedback and input were intended to make the learning objectives more transparent. Other features added to highlight the pedagogical purpose of task repetition were specific task roles for the pairs that encouraged output pushing and a task performance where students were encouraged to rework language from an earlier task performance into a subsequent repetition. From an action research based perspective, I feel my intervention resolved the disappointing second module research results, a conclusion supported by the quantitative results presented in Chapter 3.

The success of the intervention in action research reported in the third module is tempered by the fact that students did not act on the specific language feedback they were provided. While there were some instances of recycling and improving on language, as was the case with D5 in the case study presented in Chapters, 4, 5, and 6, this was quite rare. Despite this I feel my teaching quality improved between the second and third modules because in module three I had a stronger teaching presence

(Garrison and Vaughan 2008: 18) with a more integrated and transparent set of learning sequences that included more teacher involvement and engagement with the class as a whole and individual students through the transcription feedback, which appears to be the crucial difference between the two modules. The term "teaching presence," (Garrison and Vaughan 2008: 18), while related to online learning, appears to apply here as well, with successful online learning environments requiring a *social presence*, *cognitive presence*, and *teaching presence* (Garrison and Vaughan 2008: 18). While social presence and cognitive presence are already discussed in task-based research, it has been only recently teaching presence has gained prominence (see Samuda and Bygate 2008; Bygate and Samuda 2009). As relatively new pedagogies, proponents of online learning and task-based learning have often assumed little teacher intervention is needed for students to take advantage of learning opportunities. The thinking was that teachers would provide the learning tools and the students would create the learning outcomes. Samuda and Bygate (2008: 35) in regards to task-based learning write:

...there appears to be a belief that second language pedagogical tasks engage learning processes as if by magic simply through interaction, and any attempt to shape the way learners carry out a task, whether though design or implementation , is to intervene in those processes.

My own education about task-based learning and teaching started with the assumption tasks automatically facilitate productive learning and so in designing tasks or task phases I carefully considered social and cognitive issues but failed to account for my role beyond lesson designer and implementer. With task repetition specifically I assumed students would enjoy the opportunity for additional social interaction and the

166

task themes would be stimulating, two features which combined with eased processing pressure from repeating the task would create an ideal learning environment for the students.

In contrast to the findings in module two, a key feature of my teaching presence in my module three is my voice, as expressed through the feedback and comments I gave the transcription group students each week in their notebooks. This included corrective feedback, reformulation, ideas about where to attempt reformulation, and words of encouragement. Specific parts of student task work I thought was interesting or showed effort were singled-out and commented on, including providing cautionary warnings when I felt students were not trying enough. My feedback also included replying to their lesson reflections, which created a kind of dialogue between thestudents and me. They would respond to my weekly reflection questions and I in turn would respond to their reflections. This can be seen in D5's notebook (see Figures 5.1, 6.0 and 6.1). My feeling is that this weekly dialogue demonstrated to the students a commitment on my part to their class work, and a teacher's commitment to his/her job as indicated to the students is crucial to student motivation (Dörnyei 2001). Reinforcing this conclusion, Garrison and Vaughan (2008: 25) report students equate successful learning environments with a strong teaching presence.

During the task repetition stage in module two I withdrew into an observer role during task repetitions, providing comments of encouragement during the lesson, but orally and arbitrarily. In module three these comments were systematic as they were embedded in the transcription process. Thus previously I felt having a clear pedagogical focus is essential for students to maximize the benefits of task repetition,

but would like to suggest that this goes beyond providing material and transparent learning phases and must also include an audible teacher voice embedded in the learning framework. When I planned this study I did not consider the impact my voice would play in encouraging students participation, but the systematic inclusion of it emerged through planning a transparent and principled pedagogical focus for the course. In summary I feel my teaching presence maintained student task engagement week to week and because the students stayed engaged in their task work they benefited from the task repetition through improved task performances.

7.2 Section two: research limitations and suggestions for improvement

The limited time period I had to conduct this study combined with a challenging teaching environment made some aspects of this research less than ideal and this section addresses the research problems encountered. This section starts with an analysis of the main the research design then moves to analysis of the CAF measures used and discussion of how theses measures could have been improved or alternative measures used to provide a finer picture of the data. The third part discusses the problems of using only one case study for the qualitative section of this research and in particular how it was difficult to address student performance variability. 7.3 concludes by discussing the various issues encountered in conducting this classroom-based research.

7.2.1 Quantitative limitations

The first limitation of this study was the number of participants; there were only two intact classes of 20 students each involved, who were recorded for only one academic semester. Data was gathered from four tasks sessions (with each session including four task performances) over a 15-week period. Data loss due to absences further

limited the data available for analysis; out of 160 potential data sets (each set containing all four task performances for one student), only 119 sets were collected. On average at each of the four task sessions I was comparing roughly 17 transcribers versus 15 reporters. This also meant there are issues concerning how evenly the data reflects the 40 students' task performances, as out of the 119 data sets, 12 students (30 percent of the total) make up around 40 percent of the total available data (48 data sets). 19 students did three sessions (at least one transcribing and reporting each). In contrast, 7 students' (one transcribing and one reporting)(18 percent of the total) data made up 12 percent (10 data sets) of the total. Two students were not included in the data because of absences or reading task performances that negated the use of a complete data set. This means students who attended every class are better represented in the data than students whose attendance was not as consistent. Outside of differences in attendance, I found no common characteristics between these two groups; students who attended every class varied in motivation and proficiency, as did students who were regularly absent. Nevertheless, these issues limit the generalizability of the quantitative results. On the positive side I collected 119 data sets (476 task performances) evenly spread across the four task topics for the task repetition part of this study, which in my opinion constitutes a sufficient amount of data for comparison of student performances across task sessions and repetitions.

7.3 Complexity, fluency, and accuracy measures (CAF)

This section discusses the different CAF constructs used in this research. In Chapter two I addressed the main issues of using CAF. In this section I suggest different CAF measures, or ways I could have tweaked the measures to produce a finer analysis of the data.

7.3.1 Examining the complexity measures used

With complexity I felt the measures: total words per task performance, number of words per AS-unit, and the complexity benchmark provided a fairly balanced set of global measures of task performance complexity. The first measure indicated quantity of speech while the latter two suggest quality of performance. One aspect of complexity that could have been added to this list is grammatical complexity, perhaps based on number of coordinators per performance, as students who appeared to be more engaged with the tasks relied heavily on clausal chaining. One limitation of number of words per AS-unit is that it does not differentiate between within clause additions and clausal coordination. Indirectly, the measure of largest single AS-unit per performance reflect some use of clausal integration or coordination in a performance but did not directly measure that aspect of student performances.

The other potential way I could have attempted to provide a more specific account of complexity would have been to examine lexical richness. This would have been perhaps beneficial for verifying how rich students' 'new meaning' was when they repeated a task.   In the study I concluded that the best performances were from students who talked about different things each time. While the qualitative case study supports this conclusion, the quantitative analysis presented in Chapter 3 does not provide a clear picture of how rich students' new vocabulary was in later repetitions of the tasks, a clear shortcoming of this study and one I hope to address in a follow-up analysis. In the qualitative section (see Chapter 6) I compared the diversity of verbs between D5's two latter repetitions and found an important difference, indicating lexical richness may reveal interesting quantitative results. Originally type-token ratio as a measure lexical richness was excluded from the research because of the level of

students and the fact that conversation can be characterized by low lexical density and lexical variety (Thornbury and Slade 2006: 44). The other problem with type-token ratio as a measure of vocabulary complexity is that it is influenced by text length, which may have been particularly problematic here because there was a considerable difference in total words spoken between task performances. For future rework on the data, one way I would approach lexical richness would be to count lexical verbs or establish a ratio of frequent to less frequent verbs. In my qualitative study I pointed out that one coping strategy of students, evidence of their struggling to speak, was depending on highly frequent verbs. Exploring frequent verb use versus infrequent verb use could offer a picture of whether student language became more or less complex across task repetitions.

The final measure of complexity it would be worth incorporating into future analysis with the data gathered for this research is number of turns per task performance, which could reflect interactional complexity (see Dörnyei 2002). A key finding from this study was that a speaker's task performance was dependent upon how well their partner asked follow-up questions. There is concern (see Seedhouse 1999) that pairs during task work will engage in very elliptical and minimal language production. One feature of such a task performance would be a lot of quick turn taking. In my study strong task performances had a lot of adjacency pairs where students seemed to achieve flow or intersubjectivity. In contrast task performances with few adjacency pairs generally produced a lot of silence or off task L1 talk. One way to validate these observations would be to count adjacency pairs containing follow-up questions with appropriate responses.

7.3.2 Examining the fluency measures used

For fluency the main problem encountered was non-significant results for pausing. Tavakoli (2011) suggests L2 learners, unlike native-speakers, pause mid-clause. However, measuring within AS-unit pausing appears inappropriate for the students in this study as it ignored the most substantial amount of pausing, which came when students started to respond to questions. Either there was a long silence before they spoke or a false start followed by a long filled or unfilled pause. In the case of the latter, I did not count pausing until students had clearly started to speak. Based on my data I would suggest students at a very basic level of English like those here pause more regularly at clausal boundaries, especially before they start speaking. This may be a feature of controlled processing, where students need time to think before they can speak. I concluded that for the level of student in my study, it might have been more effective to measure pausing or silence between AS-units rather than within AS-unit. The other fluency measure used was length of run, different from number of words per AS-unit (used for complexity). However, the means of both were comparable and their trajectories were the same. This is perhaps because of the students' English level; they may have been unable to produce longer utterances regularly throughout a task performance. Thus I would suggest future research with similar students use only words per AS-unit and move it from a complexity construct to a fluency construct, as it can serve as a "covert" (Bygate 2001: 34) fluency measure, a conclusion supported by my results as well.

7.3.3 Examining the accuracy measures used

With accuracy the main measure was number of error-free AS-units, and accurate performances appeared to mostly be the result of students accessing and producing

partially proceduralised language. In contrast, AS-units containing new language were filled with errors. One way to confirm this would have been to separate formulaic accuracy from accuracy achieved during risk taking, which may be more rule-based. However, identifying formulaic language with non-native speakers is difficult because it can be idiosyncratic (Foster 2001). In my study I was able to find partially proceduralised language by tracing it back to pre-task notes, reformulation work, and prior repetitions, an area of further research to pursue with the data from this study.

7.3.4 Summary of the CAF measures used

In summary the CAF measures used were quite amenable to the level of students used in this study. The main weakness as I have discussed was that at times my selection of measures were too global at and not specific enough to fully help answer the research questions being asked.

7.4 Case study limitations

One obvious limitations of the qualitative section of this research was the use of only one case study, although with the limited amount of space available including more cases would have compromised the depth of analysis. As the focus of the case study was examining in depth the transcription process and my participation as the teacher in that process, in the end only one case could be included in this part of the research. D5 was selected because her production was representative of the majority of students in this study. In addition, her task session on health and exercise resembled the central tendency of the quantitative data. Thus her task session highlighted what production features were behind the central tendency of student production in this study.

The problem with using only one case study to account for the central tendency of the quantitative measurements is that performance variability among students is backgrounded. This is unfortunate because performance variability characterized a substantial amount of data and its absence from the analysis limits the informative potential of this research, particularly because the central tendency of a study may not be true for any one individual in that study (Dörnyei 2009a). For example, with D5 only one of her task sessions fit the central tendency. Ellis and Larsen-Freeman (2006: 7) caution that to dismiss variability as "noise" may ignore important sources of developmental processes. However, with regards to this study, I would argue variability in task performances related more to interactional and motivational variables than differences in underlying production variables. This is because I chose my case study based on records I kept of the main production characteristics of the students. The process of choosing a case study first started with keeping records of the key features of student production and based on this I selected students whose performances reflected those general characteristics. Next this group was narrowed down to students whose attitudes towards the class were reflective of the majority, leaving about five potential students to use for the case study in the end. Future research using data from this study should provide a fuller picture of the language variation between students by investigating more individual students' task production.

7.5 Research design analysis

In this section I discuss the various constraints encountered during this classroom-based research. The first major obstacle involved adapting a quantitative research design to fit my teaching context. Initially, the statistical advisor for the study suggested using two separate classes each following one of the two different

treatments, but unfortunately the two classes available were not comparable as one group was significantly more proficient than the other. Overcoming this issue required creating two groups spanning both classes, forcing a more equal division of stronger and weaker students into the groups used for this research. Additionally, the statistical advisor recommended having students participate in only one treatment, but as the classroom teacher I felt it was impossible to have one group always do the more intensive transcriptions, particularly since professionally it was important for me to teach all of the students in the same way. The final design seems a best compromise and the statistical advisor was comfortable with the pseudo between-subjects design used here. Despite initial concerns, it was not difficult to have both groups in the same class and they were kept relatively separate, made easier by use of a classroom built to hold up to 60 students with only 20 students. In the end, while this study includes some compromises with relation to groups and treatments, I feel the design did not significantly influence the results. The results are clearly grounded in language learning and teaching issues**.**

7.6 Problems and realities of classroom-based research

During the study, I had to make a number of changes "on the fly" to the original study design to accommodate the realities of my teaching context. Rossiter (2001: 36) in her paper on classroom-based research points out that what are "problems" for researchers are in fact "realities" for teachers and so classroom research is messy. This study is no exception as it included continually adapting to the students' proficiency level and fluctuating motivation. First, this study intended to follow Lynch (2001) by measuring LRE's (language-related episodes) between pairs at the task and transcription stage. In the trialling stage, I had recorded pairs self-correcting and

reformulating their transcripts together. I quickly abandoned this data source because the students even in their L1 produced little metalanguage, perhaps because they were simply too low in L2 language proficiency. The next problem that emerged from the trials was task performances bogging down in negotiation of roles and turn taking which wasted task time. Some students with unfamiliar partners used role uncertainty as an excuse not to do the task. Faced with this problem, I created clearer task roles for the final study. This meant that instead of having one recording per pair for one task performance, there was now one recording of each student acting as the main speaker for each task performance, increasing the amount of data generated by the study and making it possible to analyse individual performances for each task performance. It also increased the amount of reformulation and feedback work to be done, which impacted the quality of my feedback to students and took time away from other work duties and responsibilities.

Another major issue was the prevalence of students reading their notes despite my verbally asking students not to do this during task performances. It became clear this was non-negotiable for them, particularly since for many students this was the first time they were being asked to perform a task without using notes or reading. Because this study employed cognitive-based measures, this issue needed to be resolved quickly and reading needed to be taken into account somehow in the data collection as it was compromising the quality of the data gathered. The compromise here involved a system where students underlined what parts of their task performance they had read, using different coloured markers for each task performance so they could be told apart. The problem of students reading continued throughout the study and had to be managed carefully each week as students would forget to underline what they had

read or would use the wrong colour of marker. The final substantial change to the task design involved adding a redo task performance (second task performance) to confirm students tried to self-structure during a repeat task performance. This is discussed further in the next section.

7.6.1 Balancing the research agenda with the teaching agenda

A number of the adjustments made in this study were intended to capture sufficient data for the research, as a major concern with this study with low English proficiency and low motivation students was whether the data would be of sufficient quality to use the analysis constructs suggested by the literature. At times I felt I was chasing ways to gather good-enough data without totally considering the pedagogical aspects of what I was doing. My students besides being my subjects were also often saboteurs of the study. There were the larger whole-class problems like the reading of notes, but there were countless examples of individual students who made it difficult for me to collect data. Thus one challenge of this research was balancing the sometimes conflicting research agenda with the teaching agenda. An important part of the teaching agenda involved maintaining social harmony even if it compromised the teaching goals of a particular lesson. Reading of notes during task work is an example of one such problem, as oral task work has little value if students simply read what they prepared; that is recitation, not conversation, yet many students regularly tried read instead of speak spontaneously. Yet if students had been instructed to not use their notes at all they likely would not have cooperated with the research.

The most controversial example I have of trying to balance the research and teaching agendas centred on having students self-structure their second through their fourth

177

task performances. In the trialling stages students did not self-structure task performances repeated after the first performance and it was unclear if they were ignoring the intended goals of the lesson or if they simply could not incorporate prior work into subsequent repetitions. To investigate this problem I introduced the second task performance as a redo task performance, even requiring students to work with the same partner to eliminate variables emerging from having a new partner. This additional repetition was motivated partly by pedagogical aims and partly by research aims. Regarding the research aims, self-structuring appeared to be a positive production strategy; however, when the students were not doing it, for the purposes of the research this needed to be confirmed. Pedagogically, halfway through the semester it became apparent students had problems with this redo task performance. They seemed to understand this as well and occasionally expressed displeasure at having to do a redo task performance with the same partner. Despite being aware of these problems, to keep the research design continuity, I continued to have students perform this repetition. While this felt like a poor teaching decision, it seemed an equally bad research decision to drop the redo task performance halfway through the study. Thus I grudgingly followed through with the redo task performance despite knowing students would view the decision poorly. The trade-off I made was to not reprimand students if they ignored the pedagogic focus of the second task performance; as their teacher, all I could do was encourage them to try.

In summary, the second section of this chapter acknowledges the limitations of the study and offers specific suggestions on how it could have been improved. It also provides preliminary ideas of how to further explore the findings of this research, including possible strategies for revisiting the raw data. Finally, this section described

the challenges of addressing the messy nature of classroom-based research, which required tweaking the CAF measures used, using a pseudo between-subjects design, employing ad hoc data gathering, and finally the issue of balancing the conflicting research and teaching agendas. All of these features of the study should provide a cautionary example of the difficulties of classroom-based research.

It may have been possible to avoid a number of the problems encountered through having a longer trial period, but I feel there are limits to how much trialling can eliminate problems in classroom-based research. In some cases it may even create problems if the trial group and research group are different. For example, the findings from my module two study informed the design of this study reported, so in this sense module two acted as an initial trial of this task repetition research. However, after this research had started it became clear, as demonstrated with my problems of having students self-structure their task repetitions, that the students in the two different studies are quite different. Similarly, the CAF measures used here are adapted from my module two study, but the measures length of run and within AS-unit pausing proved inappropriate for the data generated here. Thus if teachers do not have access to the same group of students for trialling then the trialling period may be of limited value. In addition, even trialling research with the same group of students will not eliminate all the problems that may arise, meaning that at some point, particularly when working under time constraints, the teacher/researcher just has to start their study and troubleshoot as they move forward, which is what happened in my situation. While the issues encountered in trialling are important, teacher/researchers have to be careful about not turn trialling into a cleansing of their classroom realities, particularly since classroom-based studies have a tendency to leave out the messy

details. The downside of this is that it denies other teacher/researchers interested in replicating that research or informing their pedagogy through it the very messiness that they know was present and that they need to know about in order to properly evaluate the applicability of findings to their own context.

7.7 Section three: implications for classroom practice

The final section of this chapter addresses the potential classroom implications of the findings of this study for similar teaching contexts. This section starts by briefly highlighting student responses to the course to provide the reader with a fuller account of the study from the students' perspectives. The final half of this section offers suggestions on how task repetition and transcription can be used in the classroom, including how task repetition could be incorporated into a larger EFL curriculum.

7.8 Student survey responses

There were two different surveys in this study; students each week responded in their notebooks to a set of lesson reflection questions, with questions and answers written in Japanese (see Appendix 2). The second survey was conducted at the end of the semester. There were two basic types of weekly reflection questions. With the first type students reflected on their participation and impressions of the lesson, such as being asked to reflect on what they did or did not do well during the lesson. The second set of questions focused on language awareness, such as asking them to explain what they focused on in their reformulation work or noticed about their language use. With the end of semester survey students were asked about their attitudes to English as well as how they felt about the course, including whether they would like to continue transcription in the next semester. In the last survey students also ranked task topics according to perceived difficulty.

7.8.1 Students' impressions of the course

In identifying what students liked about the class there were a number of common responses. Students regularly indicated they had fun talking with different classmates. Quite a few students commented they enjoyed the teacher-led reporting part. In relation to transcription many students stated that while they found transcribing tedious they did like self-correcting their language, mirroring Lynch (2001) and Stillwell, et al.'s (2010) conclusions. On the negative side, students mentioned they did not like listening to their voices on the recordings and many, after hearing their oral performances, responded with deflating self-assessments.

L7: *My grammar is all mixed up*.

D4: *Grammar is out of control.*

N1: *I can't use English properly.*

Other students were even more self-critical.

L6: *It makes me sad that I can't do English well.*

E8: *I don't like transcribing cause it disappoints me how I have a lack of English skill.*

Thus the act of listening to recordings of their task performances appeared to be a source of discouragement and demotivation, although it should be kept in mind that students only listened to their first performances, which were also their weakest. For teachers wanting to avoid similar problems, I would suggest they have students compare task repetitions. Having done this myself, I find it boosts student self-confidence (Moser, et al. 2012). The other suggestion is for teachers to record student

performances when students are expected to reach their interlanguage peaks. For example, in this research students did very well in their third task performances, so those may have been more appropriate to ask students to listen to and transcribe.

7.8.2 Metalanguage use

Regarding language awareness, students in the transcription group commonly wrote specific metalanguage comments, which seems to be the result of their having had access to their transcripts. Metalanguage comments were mostly from more proficient students, and always the transcribers. A sample of these comments is reproduced in Table 7.0.

Table 7.0 Transcription students' metalangauge comments

L4: *I need to use to or for I have to revise what word comes next after using an auxiliary verb.*

L2: I tend to skip using ing and I have problems with to and by.

E5: I had problems with the third person, plural, and verb tenses.

These types of comments were common, but not in overabundance. In module two I asked similar questions to students who had to recall their task performances. Like the reporting students in this study, they rarely responded adequately to the questions about language awareness, often leaving their response spaces blank. Thus in module two it appears students could not remember specific language issues they had in completing the task work, while in module three there was increased awareness of language issues among the transcription group students. In fact, this is my first experience of students at a low English proficiency level in an oral communication

class making specific metalanguage comments on their performances, and suggests transcription may be a useful tool for developing students' metalanguage awareness.

The students had little to say about reformulation work, and when they did comment, those comments tended to express negative feelings toward the activity. No student indicated enjoying reformulation or finding it useful. Instead students wrote they could not add details to their transcripts or repeat task performances with comments such as, *I couldn't write in details* (N8).

There were clear limits of how much new information students could add to their second task performances, and many stated they did not have the language resources to reformulate their transcripts or could not conceptualize how to expand on a topic, compounded by the fact that it was rare for students to ask me for help during this stage. In addition few students seemed to pay much attention to the extra input and feedback I provided to the whole class. As the semester continued students also realized that even if they wrote a lot during the reformulation stage, they had trouble employing it in a subsequent task repetition. Table 7.1 shares some of the concerns students expressed.

Table 7.1 Student comments on reformulation

L9: I had a hard time. The contents I wrote down in my notebook are different from what I said.

N4: I felt it was difficult for me to make conversation even though I wrote down my ideas in my notebook first.

L6: It's always hard to remember what I wrote.

Table 7.1 reveals students struggled to recall earlier work while simultaneously engaging in a real-time task performance. Related to the inability to recall language, almost every student at some point acknowledged they read or looked at their notes, the topic of Table 7.2

Table 7.2 Student comments about using their notes during task performances

D5: I tried not to look at my notes while I expanded on my ideas, but I wasn't
successful – I failed at it.

N3: It's hard to talk without looking at notes.

Either a student could not easily incorporate pre-task work or reformulation work into a task performance or her English proficiency was too low to talk without the aid of her notes.

7.8.3 Task repetition comments

For task repetition, regardless of whether it was transcription or reporting, almost all students at some point in the semester acknowledged they were able to improve in subsequent task repetitions. Students tended to rank their third task performance as *good* or *excellent*, as illustrated in Table 7.3.

Table 7.3 Student comments about the third task performance

E4: I was able to talk in more detail.

D2: I talked about what I hadn't before.

E6: I talked better than the last time.

E5: I could say English words and answer unexpected questions.

The fact that almost every student noted improvement on a subsequent task performance (especially the third performance) is a strong endorsement of task repetition as a pedagogic tool.

End of semester survey responses were positive towards the course. In particular, students clearly enjoyed repeating tasks with new partners. Regarding transcription, the consensus was it was useful but less would be preferable. On this point almost all students reflected a willingness to continue to do transcription. They particularly said they liked self-correcting. A number wrote they preferred this class to any other type of English class they had taken in the past. This is reinforced by my impression from watching the students throughout the semester that they did enjoy the unique features of this class.

7.9 Advice for classroom applications

7.9.1 Suggestions regarding transcription

This last part of this chapter discusses the classroom implications of this study and provides specific suggestions for using task repetition and transcription in similar contexts. Even though most students enjoyed transcription, and self-correcting in particular, the quantitative results for this study do not demonstrate any advantage for transcription over reporting in relation to subsequent task repetitions. This is disappointing because transcription was intended as a scaffolding tool to help students improve their subsequent task performances. At the end of this study I wondered if perhaps more closed tasks with 'task essentialness' accompanied with an intensive focus on form sequence might have resulted in more positive results for transcribing. However, Stillwell et al. (2010) in their transcription study with students doing a

185

closed task (poster presentation) facilitating the repeating of language, encountered the same working memory capacity issues as here despite the restricted nature of the task with its potential for recycling language; first performances were different from the second performances and students did not reuse language. Their students acknowledged they were unsuccessful incorporating changes made at the reformulation stage in a repeat task performance and some explained poorer performances on not reviewing their corrections (Stillwell et al. 2010). Their findings appears to reinforce the conclusion that because of working memory constraints students cannot retain earlier reformulation work in their working memory performing a real-time task. Like myself, Stillwell et al. (2010) may have missed the important point that reformulation is a concept originating in writing classrooms (see Johnson 1988), and may not be easily transferable to speaking classes, even through transcription. Adams (2003: 349) notes teachers have often "overlooked" the important difference between speaking and writing when it comes to reformulation in the classroom.

I would suggest that if teachers want to use transcription in similar contexts, it is more suited for stocktaking after a task performance. The transcription could then be followed by a set of practice activities or a form-focused task. The stocking taking should perhaps be for long-term language restructuring and unconnected to subsequent task repetitions. Thus transcription is more suitable at the end of a set of lesson phases rather than as a scaffold before a repeat task performance as transcription appears to have little impact on immediate repeat task performances. This may sound like a slight difference, but I am contending that transcription has little direct impact on an immediate subsequent task performance. This is an

important point for teachers to consider because of the amount of time and effort needed for student self-transcription. However, the suggestion to not use transcription to inform later task performances is not for all contexts, as teachers have reported successfully using transcription as a scaffolding tool; Lynch and Maclean's (2001) and Mennim's (2007) transcription studies seem to suggest it is particularly ideal for ESP or EAP courses where more proficient students are working with language in more restricted situational contexts. However, as the Stillwell et al. (2010) indicate, using transcription in these contexts can be challenging.

7.9.2 Suggestions for improving corrective feedback

The problem with providing corrective feedback in this research was the difficulty of addressing individual student interlanguage issues without face to face feedback sessions with the students. This is a significant difference between this study and other transcription studies, as other studies had a number of lesson phases with direct teacher to student feedback sessions (see Lynch 2007 and Mennim 2007). This direct contact appears to be crucial for facilitating student uptake, a point reinforced by Allwright et al. (1988), who suggest collaborative dialogue between the teacher and student is crucial for making reformulation successful in the classroom. They conclude this direct feedback and collaboration is more influential in getting the students to act on suggestions than the reformulation examples themselves. Thus in this research giving feedback through a notebook with no follow-up discussion between teacher and students seems insufficient and may have contributed to students apparent lack of interest in their finished transcripts. More direct teacher to individual student contact time may have encouraged more students to try and self-structure later task repetitions, an issue for further investigation. Thus it appears direct teacher to

student feedback is something teachers must try maximize. The obvious problem, and the one I faced, is finding the time to work with students individually. One solution could be, rather than reviewing students' transcripts on my own, instead using that same time to discuss task performances with individual students, with some of these individual feedback sessions occurring during class time as other students work on transcribing and reformulating. As I explained in Chapter three most students were very good at transcribing their spoken performances so teachers can be confident they are reviewing accurate transcripts.

7.9.3 Recommendations for improving task repetition

The most important finding for the task repetition component of this study was that students' production improved in subsequent repeat task performances relative to their first performances, demonstrating the importance of incorporating task repetition into language classrooms. I have four suggestions that teachers need to consider when using task repetition in similar contexts. First, when using open-ended tasks involving changing partners, do not expect students to recycle or reformulate very much language from their earlier task performances for two reasons. The first is because of working memory limitations, although proximity of task performances can remedy this to a certain extent. The second reason is that students when working with new partners tend to talk about different things.

The second suggestion for teachers considering using task repetition is that even though students may not rework prior performances into an improved subsequent performance, they do improve in terms of meaningfulness when repeating the same task, meaning students expand on previous topics and most talk about new topics, leading to a larger variety of language use in later repetitions (illustrated in the case

study). This improved meaningfulness, or lexico-grammatical access, is perhaps the result of eased processing pressure made possible by task repetition. When students are working at the limits of their interlanguage ability then there are lots of learning opportunities and processes. For the teacher one benefit of students working at the peak of their interlanguage is that it can facilitate a rich stocktaking phase after the final task, providing an abundance of language for the teacher and students to analyse and practice.

The third point for teachers considering using task repetition is they should be cautious about engaging in task repetition that resembles verbatim repetition. If meaningfulness is compromised then student task engagement seems to decrease. Thankfully, it appears that with a transparent pedagogical focus and an adequate teaching presence, meaningfulness can be maintained in open-ended tasks by simply having students change partners, which came through in the data for the third performance in this research. There were a number of factors that contributed to high levels of production or meaningfulness in the third performance, partially explained through task familiarity gained through task repetition and students being assigned to new partners. For task familiarity, minimizing time gaps between repetitions appears to influence performance positively, with the initial task appearing to act as a warm up task for the subsequent task. A final contributing factor to promoting productive task work is establishing clear task roles. In my study students initially struggled with managing their turn taking, which ultimately influenced their production. The students' lack of conversation strategies may have contributed to their poor turn taking management. Designing and assigning task roles that allow pair work configurations where one student can focus on output pushing without worrying about

the social maintenance aspects of their task performance appeared to help students complete their tasks more successfully and helped avoid a situation where the social aspects of the classroom hindered risk taking and extended turn taking.

My fourth pedagogical suggestion for teachers considering using task repetition is that an initial task performance, even with pre-task planning, is not going to be the same quality as a repeat performance, perhaps because many students cannot operate at the peak of their interlanguage in an initial task performance and need a second or third opportunity to attain their full current potential. Teachers using a strong version of TBL (Ellis 2003: 28) where tasks are central to the learning process may want to incorporate some type of task recycling into their curriculum.

7.10 Student production insights

This final section discusses the key features of student oral production in this study and what insights they may contribute to pedagogy. The first important finding here was the extent to which students used memorization in preparation for their task performances. First students who lacked sufficient procedural knowledge memorized their pre-task work. Second, students who wanted to incorporate their reformulated work into their task performances regularly memorized their work. Third, students that deliberately wanted to improve their accuracy did so by memorizing their best language from the pre-task or reformulation work. I would suggest that memorizing was more than just a student coping strategy and was also not the result of poor pedagogical planning. For these students it appears to be an important strategy they use to incorporate non-proceduralised knowledge into a speaking task.

Dörnyei (2009a) notes rote learning that is most commonly utilized in the PPP method has fallen out favour with the rise of TBLT with its more analytical and experiential dimensions of learning. Despite this loss of popularity among teachers my data suggests that students very much rely on memorizing language for task work. Ding (2007) found that the very successful language learners in his study all noted the importance of memorization and imitation to their learning success. After analyzing my students' task performances the importance of the role of memorization in their language learning is evident. My advice to teachers is to make space in their lessons for memorization, as it is not a symptom of students being lazy, particularly in the early stages of language learning. Even with more proficient students, memorization should be accepted as an important tool for students to rely on and teachers should encourage its use at points in their lessons, including during tasks.

7.10.1 Memorization and building up formulaicity

Students may rely on memorization when learning language because it is an important way to build language formulaicity. As lexicalization is a precursor to syntacticalization (Skehan 1998), at the early stages of lexicalization chunks are mimicked and memorized. As the students in this study lacked procedural knowledge or "behavioral routines" (Dekeyser 2007a: 3) of English, by memorizing language chunks and producing them during the task students were potentially making holistic meaning, form, and function mappings. Doing this holistically through task work increases the likelihood they will retain the language and reproduce it in subsequent repetitions, facilitating development of procedural knowledge. The other benefit is students, by repeating memorized chunks, are able to free attentional resources for attending to other features of their production, including their rule-based system. Thus

191

memorizing chunks and then repeating them in tasks appears to be an important process in the development of procedural knowledge.

This building up of stores of formulaic language, at least in the early stages of learning, is crucial for development of procedural knowledge. Interestingly, De Keyser (2007a: 10) notes automatization has until very recently been neglected in the field of applied linguistics. Gatbonton and Segalowtiz (2005: 327) contend communicative language teaching has not accounted for how to develop fluency in students, especially through interactive task work, noting teachers view it as irreconcilable to the "unpredictability" of communicative tasks. To ensure fluency development is not ignored Gatbonton and Segalowtiz (2005) suggest that task work should be genuinely communicative, functionally formulaic, and inherently repetitive, designing a relatively sophisticated set of phases to recycle language. Nunn's (2006) task-based units are another example of a way to recycle language more effectively than simply repeating a task. The lesson plan used here in this study appears to meet these three features of task work necessary for building automatization. I believe that my module three lesson plan, and revised lesson plan, meet the three features of task work that are deemed necessary for building automatization. In my lesson by using basic thematic tasks across the syllabus, this more than likely ensures the repetition of high frequency language. While I have yet to fully investigate it I would suggest that students when they repeated the same task in my study, and even when they produced new meaning that they were recycling a limited amount of high frequency language. Thus language repetition is possible in my syllabus on two levels, first through task repetition and second by using basic thematic tasks across the syllabus, with high frequency language likely recycled across themes.

7.11 Task repetition in an EFL curriculum

This final reflection section summarizes how this class could be integrated into a larger language learning curriculum. The classes in this study are part of what Swan (2005: 392) calls a '3hpw' (three hours per week) context and he argues task-based learning may be inadequate for such contexts because of problems with "input coverage," quoting Fotos (2002) as someone familiar with Japan who has rejected a task-based approach based on communicative goals because of a lack of input coverage. Fotos (2002: 139) further suggests that a communicative syllabus is inconsistent with the Japanese education system and its prioritizing of English exams for entrance into good universities or companies. While it is beyond the focus of this study, Fotos's (2002) argument is only partially accurate. For example, in my study almost every student in the final survey acknowledged the importance of knowing English for future career opportunities.

In response to Swan's (2005) questioning of task-based learning in an EFL context like Japan, I would suggest this course is appropriate for this context, as it is based on Ellis's (2002, 2003) communicative module in his modular syllabus, and is designed to build up formulaic language in students who lack procedural knowledge. The other half of his Ellis's (2002, 2003) modular syllabus is form-focused and there is little or no integration between his two modules. In the communicative module, students develop basic communicative ability by repeating thematic tasks that allow them to recycle high frequency language (Ellis 2002, 2003), which is practical considering in a 3hpw context there is not much time for the wide coverage as Swan (2005) advocates. In this class the goal was to help the students develop some procedural knowledge and expose them to how basic communicative skills can be acquired. For

students who have previously studied in classes based on grammar translation, the importance of exposing students to the process of how to acquire basic speaking ability should not be understated. Additionally, in the final survey students indicated they enjoyed the class and one reason for this may be that usually after third task performances almost every student indicated in their weekly survey they felt that their subsequent task performances improved in relation to their initial task performance, indicating that students (albeit indirectly) recognized the benefits of task repetition. Part of a larger curriculum that goes beyond 3hpw, I see task repetition as appropriate for fluency work. With Nation's (2007) four strands curriculum framework as an example, task repetition could be heavily employed in the fluency strand where the teacher is primarily concerned with automatization. With the focus on building procedural knowledge, teachers would not have to worry about the resulting limited grammatical and lexical improvements.

This final section of the chapter highlighted the importance of developing students' procedural knowledge, including the role task repetition should play at the classroom level and as part of a wider EFL curriculum. My hope is that these suggestions provide ideas for teachers in similar contexts who are using task repetition. The findings from both modules and my suggestions should demonstrate task repetition is difficult to effectively implement in the classroom and the suggestions should be seen as tentative and context-specific. My intention is to further develop and refine these initial recommendations as I revisit my data and undertake supplementary research.

7.12 Conclusion

To conclude, I feel there are four important levels to this study that should contribute to the understanding of task-based language learning and teaching literature. First, as an intact classroom-based study this module contributes to a growing list of studies that Samuda and Bygate (2008) have called for to validate task-based learning and teaching in the classroom. The benefit of this classroom-based study and other similar studies is that they are relevant to classroom practice and provide teachers with vivid accounts of key language learning concepts as they are applied in the classroom. This ultimately helps teachers evaluate the applicability of these concepts to their own contexts and provides practical teaching ideas. In this study the key concepts are task repetition, transcription, and reformulation. The second contribution this study makes to the current task-based literature is that it is just one of a few to date that provide a descriptive and analytical account of low beginner adults' language production during task work in an EFL context. This is important and reflects Norris and Ortega's (2009: 573) concern that features of production along a "developmental continuum" should not be seen as constants. Reflecting this, this research suggested different ways to view performance complexity, accuracy, and fluency for these students at this developmental level.

The third contribution this study makes to the task-based learning and teaching literature is that it demonstrates students in intact classrooms will take advantage of task repetition opportunities and these repetitions will improve their production, which should provide further support for teachers to give task repetition primacy in lesson planning and curriculum design. Its value to language learners goes beyond being an occasional option in the classroom. However implementing task repetition

effectively in the classroom is a complicated process with unintended consequences and only conditional solutions. The final contribution this study makes is probably the most important. A key factor in the relative success of this study was the sustained and effective teaching presence. Until very recently task-based learning has focused on the cognitive and social characteristics of tasks and task phases that lead to successful task-based learning. This study helps demonstrate that task-based research and teaching needs to devote equal attention to the teacher variable, as it appears that, at least in my teaching context, it is the teacher who determines the success of task-based learning in the classroom. In this sense it is fair to say that the teacher is more important than the task.

Appendix 1: tasks

Task one: health and exercise

- Is your lifestyle healthy or unhealthy? Give at least 4 reasons why.

1.
2.
3.
4.

- What do you do to stay healthy or what do you do that is unhealthy? Give 4 examples.

1.
2.
3.
4.

- How often do you exercise? What do you do? What sports or exercise did you used to do?

- Is your diet healthy or unhealthy? Give 5 reasons.
1.
2.
3.
4.
5.

- How often do you eat junk food? When do you eat it and what is your favorite junk food?

- Do you have a lot of stress? Please explain why you do or why you don't?

- How many hours of sleep do you get in a night? How about recently? Why?

- How can you improve your healthy and lifestyle? Give 4 reasons.
1.
2.

3.
4.

Task two: pets and animals

1. Do you have a pet or had a pet? What kind of pet? What is/was its name? What is its character like? What does it look like – hair- color- eyes?

2. What does it like to do and what are its habits?

3. Do you like dogs or cats better? Why? Give 3 reasons.

4. What kind of pet would like to have? Why? Give 3 reasons

5. What kinds of stuffed animals did you have when you were a child or even now? Write 4 things about one of your stuffed animals.

6. Have you had a scary experience with an animal? What happened?

7. If you could be any animal, what would you be? Give 3 reasons.

Task three: shopping and fashion

1. How often do you go shopping? Do you like shopping? Why? Why not?

2. Who do you usually go shopping with?

3. What is your favorite shopping area? Why? Give 3 reasons.

4. What are your two favorite shops? Give 3 reasons each.

5. Where do you buy your clothes? Why?

6. Do you like Uniqlo? Do you buy clothes from Uniqlo?

7. What shops don't you like? Why Give 3 reasons.

8. Are you a good bargain shopper? Why? Why not?

9. How much do you spend a month on shopping?

10. What's the most expensive thing you have bought? Why did you buy it?

11. Do you often buy things you don't need? Give examples and reasons why you bought? What is the worst thing you bought?

12. Is there something that you want to buy right now? What is it and why do you want to buy it?

Task 4: dating and marriage

1. Where is a good place to go on a date? Why?

2. Describe your perfect date. Where would you go? What you do? What would you eat? How would the date start and end?

3. What does your ideal date look like?

4. What 3 qualities must he have?

5. What would be your nightmare date?

6. Do you think it is okay to kiss on the first date? Why or why not?

7. Do you think it is okay for girls to ask boys out? Why or why not?

8. Do you think boys should pay for the date? Why or why not?

9. Do feel comfortable holding hands in public? Why or why not?

10. At what age do you want to get married? Why?

11. Would you do an arranged marriage?

Appendix 2: lesson reflection questions

(Note: English translations have been added for the reader)

First lesson: questions for transcribing students

１．今日の授業でのあなたの努力を評価してください。
Poor（あまりしなかった）, Okay(普通), Good (がんばった), Excellent (とてもがんばった)
の中から一つ選んでなぜそう思うか具体的に理由を書いてください。

How would you rate your effort today: okay, good, excellent? Choose one, and give a specific reason why you think so.

２．今日の授業のどの部分が好きでしたか、好きでなかったですか。具体的にその部分をあげて理由を書いてください。

What part of the lesson today did you like or not like? Give a specific part and reasons why.

３．今日の授業 はeasy （やさしかった）, ok (ちょうど良かった), difficult (難しかっただったですか)。一つ選んで具体的に理由を書いてください。

Was today's lesson easy, okay, difficult? Please choose one and give a reason why.

４．文法と語彙の何を直そうとしましたか。 具体的な例を書いてください.

What grammar and vocabulary did you try to fix. Give specific examples.

５．今日の英会話のレポートを読んで分析してあなたの英語について何に気がつきましたか。気がついた事を具体的に書いてください.

What did you notice about your English today from reading your conversation? Please be specific.

６．あなたの行った英会話を書き直した後で、どのようにあなたの英語を better and more interesting でふくらませたものにしようとしましたか。具体的な例を書いてください.

How did you try to make your own English better and more interesting after reading your English conversation performance?

First lesson: questions for reporting students

１．今日の授業でのあなたの努力を評価してください。
Poor（あまりしなかった）, Okay(普通), Good (がんばった), Excellent
(とてもがんばった)
の中から一つ選んでなぜそう思うか具体的に理由を書いてください。

How would you rate your effort today: okay, good, excellent? Choose one, and give a specific reason why you think so.

２．今日の授業のどの部分が好きでしたか、好きでなかったですか。具体的にその部分をあげて理由を書いてください。

What part of the lesson today did you like or not like? Give a specific part and reasons why.

３．今日の授業 はeasy （やさしかった）, ok (ちょうど良かった), difficult (難しかっただったですか)。一つ選んで具体的に理由を書いてください。

Was today's lesson easy, okay, difficult? Please choose one and give a reason why.

４．今日書いた英会話のレポートからあなたの英語について何に気がつきましたか。気がついた事を具体的に書いてください.

What did you notice about your English today from reading your conversation report? Please be specific.

５．今日文法と語彙について先生や教科書やパートナーから何を学びましたか。 具体的な例をあげてください。

What did you learn about grammar and vocabulary from the teacher or textbook or partner today? Please say from which and write down what you learned or noticed/

６．あなたの行った英会話を書き直した後で、どのようにあなたの英語を better and more interesting でふくらませたものにしようとしましたか。
具体的な例を書いてください.

How did you try to make your own English better and more interesting after reading your English conversation performance?

203

Second lesson: questions for both groups.

１．今日の授業でのあなたの努力を評価してください。
*Poor（あまりしなかった）, Okay(普通), Good (がんばった), Excellent
(とてもがんばった)*
の中から一つ選んでなぜそう思うか具体的に理由を書いてください。

How would you rate your effort today: okay, good, excellent? Choose one, and give a specific reason why you think so.

２．今日の授業のどの部分が好きでしたか、好きでなかったですか。具体的にその部分をあげて理由を書いてください。

What part of the lesson today did you like or not like? Give a specific part and reasons why.

３．今日の授業 はeasy （やさしかった）, ok (ちょうど良かった), difficult
(難しかっただったですか)。一つ選んで具体的に理由を書いてください。

Was today's lesson easy, okay, difficult? Please choose one and give a reason why.

４．パートナーと 会話をする前に、どのようにあなたの英語を better and more interesting でふくらませたものにしようとしましたか。
具体的な例を書いてください.

Before talking to your partner how did you try to make your performance more detailed? Please be specific and give examples.

５．今日のあなたの英語の良かった点と悪かった点は何でしたか. 具体的に例を挙げてください.

What were the strong points and weak points of your English today? Please be specific and give examples.

Questions asked orally to students at the end of the semester.

1. Do you like studying English?

2. Do you think English is important for your life?

3. What did you think about this class?

4. What this class useful or not useful for learning English Why? Why not?

5. Would you like to continue studying this way? Why? Why not?

6. Do you think your English improved or didn't improve? Why do you think so?

7. This semester we talked about health, pets, shopping, dating. Please rank them from the easiest to the hardest. And give a reason for each.

Appendix 3: D5's task performances for task session health and exercise

Performance 1: first time

Shared Silence 14.52

1P: How often do you exercise?

2D5:   /{I}[I play tennis]{5.70}{two}[two times a week]/**(7)(15.00 seconds)**

3P: Only tennis?

4D5:   /Yeah/

5P: Then what sports or exercise you used to do? Tennis or (in Japanese *anything else?)*

6D5:   /[I used to {used to?} play to basketball]{2.40} for {2.56}[junior high school and high school]/**(12)(21.70 seconds)**

7P: So six years

8D5:   /Yeah/

9P: Why do play tennis in college?

10D5: /{I I started I want to start}{6.03}{I} I want to started new sports (in Japanese *new*)/**(6)(23 seconds)**

11P: So which do you like basketball or tennis?

12D5: /{I}{5.55}[I like basketball and tennis but]{3.79}[now I play tennis twice a week so]{I}{3.07}][my best sports is tennis]/**(19)(34.19 seconds)**

Shared Silence 13.00

13P: Do you do any other sports?

14D5: /{I I I I I have played} I used to do swimming/**(5)(28.22 seconds)**

15P: Can you swim butterfly?

16D5: /Yeah/

17P: Can you swim well? Fast?

18D5: (In Japanese *One more time)*

19P: Can you play fast? (in Japanese *Fast?)*

20D5: /So so/

21P: Do you have a lot stress?

22D5: /{I} I don't have a lot of stress**/(8)(4.52)**

23P: Don't you have some problems?

24D5: /{If} If {I}[I have problems or stress but {I}]{4.31}{I}[I forget {8.23}{while} while sleep]**/(11)(32.5 seconds)**

Performance 2: rehearsal same partner

1D5: /My lifestyle is healthy/And a little unhealthy/Because I like junk food very much/Especially I like potato chips and {4.41} juice/{I I eat junk food}[I eat junk food {for}]{7.39}{more}[more than four times a week]/And I drink juice everyday/(4+4+7+6+11+5)

2P: What flavour of juice do you like?

3D5: /{I} I like fruit's juice/Mixed juice and fresh juice/(4+5)

4P: /100 percent/ (in Japanese)

5D5: /So {I I stopped} I stopped eat junk food but {I} I can't stop eat junk food/{But} but {I}{3.34} I think healthy life because {I have a} I have a breakfast anytime/I have a vegetables and fruits/(14+6+6)

6P: What vegetables do you eat for breakfast?

7D5: /{I} I eat breakfast is Japanese food/Miso soup and rice/So miso soup is in very many vegetables/{Onion} onion potato carrot/(6+4+8+3)

8P: What fruits do you eat for breakfast?

9D5:   /{I eat necessary} I eat for strawberry/And my mother's friend send to {3.11} orange {4.94} full box/(4+9)

10P: Mikan? (orange? in Japanese)

11D5:   /Hasaku/

12P: I like

13D5:   {So I} I eat hasaku {every} after dinner/(5)

14P: Do you exercise?

15D5:   /{I} I play tennis {for two week} for two times a week/And yesterday I play basketball in {Sho} Shoin basketball club/(8+9)

16P: Who do you play tennis with?

17D5:   /{I I play} I joined to Kinki University and Shoin University's tennis circle/So circle member these days/(10+5)

Performance 3: new partner

1P: Is your lifestyle healthy or unhealthy?

2D5:   /My lifestyle is maybe healthy but a little unhealthy because {I} I like junk food/And {I I have a junk food}[I love junk food very much {I} I eat junk food]{3.97}[more than {two} four times a week]/And I drink juice everyday/(14+17+5)

3P: What kind of junk food do you like?

4D5:   /{I} I like chocolate and potato chips/{It's} it is {li} little unhealthy but I can't stop junk food/[But {I} I think my lifestyle is maybe healthy because {I} I have a breakfast anytime and I have vegetables everyday]{and} and vegetables and fruits is my favorite food]/(6+10+27)

5P: What kind fruits do you like?

6D5:   /{Necessary I necessary necessary} especially I like orange/My mother's friends send to <u>hasaku</u> full a box/So {I} I eat <u>hasaku</u> every after dinner with family/(4+9+9)

7P: What is different between orange and hasaku?(In Japanese)

8D5:   /{Orange is}[orange is almost mikan but hasaku is]{5.02}[taste grapefruit and orange]/(11)

9P: Do you like to play sports?

10D5: /Yeah {I like} I like watching sports and I play sports/{I} I play sports every two times a week/{I} I join to tennis circle for Kinki University and Shoin University {10.00} combinations circle/(9+8+13)

11P: Where do you play?

12D5: /Tennis/{I play tennis court is I I} we play tennis court in near the Kinki University's tennis court/{It's} it's {in} indoors tennis court/So rainy's day we can play tennis/(11+5+7)

13P: What do you like sports?

14D5: /I like basketball/(3)

Performance 4: final

1P: Is your lifestyle healthy or unhealthy?

2D5:   /Maybe my {lifestyle} lifestyle is healthy because I almost don't dislike food/(12)

3P: /Great/

4D5:   /And {I} I have a lot of {sleepy} sleep time/(8)

5P: What do you do to stay healthy?

6D5: /{I I have I} I have a breakfast everyday and/As as (In Japanese: I eat breakfast)/(6)

7P: /Nani?/ (in Japanese: What?)

8D5: /As dekiru dake as/(In Japanese: 'What do I put in?')

9P: /Possible?/

10D5: As as nani/(In Japanese: What?)/

11P: /As/

12D5: /{As dekiru dake} as walk as possible (P: Okay okay) for station/(4)(1.20 minutes)

13P: Which do you eat rice or bread at morning?

14D5: /[Almost {I have} I have rice and miso shiru at morning][{but holiday is I have}[but I have bread in holiday]/(14)

15P: Do you do any exercise?

16D5: /{Now now now}{**17.96**} now {I I}[I don't play tennis because test time but]{5.59} [usually I play tennis so two times a week]/(18)

17P: Do you join to tennis circle?

18D5: /Yes I join tennis circle for Shoin and Kinki University tennis circle/(12)

19P: How often do you eat junk food?

20D5: /Yes/I like junk food/{I like I like}{**17.87**} I like potato chips and/(4+5)

References

Adams, R. 2003. L2 output, reformulation, and noticing: implications for IL development. *Language Teaching Research* 7: 347-376.

Ahmadian, M. and Tavakoli, M. 2011. 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: 35-60.

Allwright, R.L., Woodley, M. P., and Allwright, J. M. 1988. Investigating reformulation as practical strategy for the teaching of academic writing. *Applied Linguistics* 9: 237-258.

Baddeley, A. 2007. *Working Memory, Thought, and Action.* Oxford: Oxford University Press.

Batstone, R. 2002. Contexts of engagement: a discourse perspective on intake and pushed output. *System* 30: 1-14.

Birjandi, P. and Ahangari, S. 2008. Effects of task repetition on the fluency, complexity, and accuracy of Iranian EFL learners' oral discourse. *Asian EFL Journal* 10: 28-52.

Burns, A. 2010. *Doing Action Research in English Language Teaching: A Guide for Practitioners.* New York: Routledge.

Bygate, M. 1996. Effects of task repetition: appraising the development language of learners. In Willis, J. and Willis, D. (Eds.), *Challenge and Change in Language Teaching.* Oxford: Heinemann.

Bygate, M. 2001. Effects of task repetition on the structure and control of oral language. In Bygate, M., Skehan, P. and Swain, M. (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing.* Harlow: Longman.

Byagte, M. 2005. Oral second language abilities as expertise. In Johnson, K. (Ed.), *Expertise in Second Language Learning and Teaching.* Hampshire: Palgrave Macmillan.

Bygate, M. and Samuda, V. 2005. Integrative planning through the use of task-repetition. In Ellis, R. (Ed.), *Planning and Task Performance in a Second Language.* Amsterdam: John Benjamins.

Bygate, M. and Samuda, V. 2009. Creating pressure in task pedagogy: the joint roles of field, purpose, and engagement within the interaction approach. In Mackey, A. and Polio, C. (Eds.), *Multiple Perspectives on Interaction.* New York: Routledge.

Carter, R.A. and McCarthy, M.J. 2006. *Cambridge Grammar of English: Comprehensive Guide to Spoken and Written English Grammar and Usage.* Cambridge: Cambridge University Press.

Chafe, W. 1980. The deployment of consciousness in production of a narrative. In Chafe, W. (Ed.), *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production.* Norwood, NJ: Ablex.

Clennell, C. 1999. Promoting pragmatic awareness and spoken discourse skills with EAP classes. *ELT Journal* 53: 83-91.

Cook, G. 2000. *Language Play, Language Learning.* Oxford: Oxford University Press.

Corder, S.P. 1967. The significance of learners' errors. Reprinted in Corder, S.P. 1981. *Error Analysis and Interlanguage.* Oxford: Oxford University Press.

DeKeyser, R.M. 2007a. Introduction: situating the concept of practice. In DeKeyser, R.M. (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology.* Cambridge: Cambridge University Press.

DeKeyser, R.M. 2007b. Skill acquisition theory. In Van Patten, B. and Williams, J. *Theories in Second Language Acquisition*. New York: Routledge.

Ding, Y. 2007. Text memorization and imitation: the practices of successful Chinese learners of English. *System* 35: 271-280.

Dörnyei, Z. 2001. *Motivational Strategies in the Language Classroom.* Cambridge: Cambridge University Press.

Dörnyei, Z. 2002. The motivational basis of language learning tasks. In Robinson, P. (Ed.), *Individual Differences and Instructed Language Learning.* Amsterdam: John Benjamins.

Dörnyei, Z. 2007. *Research Methods in Applied Linguistics.* Oxford: Oxford University Press.

Dörnyei, Z. 2009. *The Psychology of Second Language Acquisition.* Oxford: Oxford University Press.

Dörnyei, Z. and Kormos, J. 2000. The role of individual social variables in oral task performance. *Language Teaching Research* 4: 275-300.

Dörnyei Z. and Teng, W. 2009b. Motivational processing in interactional tasks. In Mackey, A. and Polio, C. (Eds.), *Multiple Perspectives on Interaction*. New York: Routledge.

Duff, P. 2008. *Case Study Research in Applied Linguistics.* New York: Lawrence Erlbaum Associates.

Edwards, C. and Willis, J. (Eds.) 2005. *Teachers Exploring Tasks in English Language Teaching*. New York: Palgrave Macmillan.

Ellis, R. 2002. The place of grammar instruction in second/foreign language curriculum. In Hinkel, E. and Fotos, S. (Eds.), *New Perspectives on Grammar Teaching in Second Language Classrooms*. New Jersey: Lawrence Erlbaum Associates.

Ellis, R. 2003. *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

Ellis, R. 2009a. Implicit and explicit learning, knowledge and instruction. In Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J. and Reinders, H. (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing, and Teaching*. Bristol: Multilingual Matters.

Ellis, R. 2009b. The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics* 30: 474-509.

Ellis, R. and Barkhuizen, G. 2005. *Analysing Learner Language*. Oxford: Oxford University Press.

Ellis, N. and Larsen-Freeman, D. 2006. Language emergence: implications for applied linguistics – introduction to the special issue. *Applied Linguistics* 27: 558-589.

Essig, W. 2005. Storytelling: effects of planning, repetition, and context. In Edwards, C. and Willis, J. (Eds.), *Teachers Exploring Tasks in English Language Teaching*. New York: Palgrave Macmillan.

Estaire, S. and Zanon, J. 1994. *Planning Classwork: A Task Based Approach*. Oxford: Macmillan Education.

Field, A. 2009. *Discovering Statistics Using SPSS*. London: Sage Publications.

Foster, P. 2001. Rules and Routines: A consideration of their role in the task-based language production of native and non-native speakers. In Bygate, M., Skehan, P., and Swain, M. (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. Harlow: Longman.

Foster, P. and Skehan, P. 1999. The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research* 3: 215: 247.

Foster, P., Tonkyn, A., and Wigglesworth, G. 2000. Measuring spoken language: a unit for all reasons. *Applied Linguistics* 21: 354–75.

213

Fotos, S. 2002. Structured-based interactive tasks for the EFL grammar learner. In E. Hinkel and S. Fotos (Eds.), *New Perspectives on Grammar Teaching in Second Language Classrooms.* New Jersey: Lawrence Erlbaum Associates.

Freed, B. 2000. Is fluency, like beauty, in the eyes (and ears) of the beholder? In Riggenbach, H. (Ed.), *Perspectives on Fluency.* Ann Arbor: The University of Michigan Press.

Garman, M. 1990. *Psycholinguistics.* Cambridge: Cambridge University Press.

Garrison, D.R. and Vaughan, N. 2008. *Blended Learning in Higher Education Framework, Principles, and Guidelines.* San Francisco: Jossey-Bass.

Gass, S., Mackey, A., Fernandez, M., and Alvarez-Torres, M. 1999. The effects of task repetition on linguistic output. *Language Learning* 49: 549-80.

Gatbonton, E. and Segalowitz, N. 2005. Rethinking communicative language teaching: a focus on access to fluency. *The Canadian Modern Language Review* 61: 325-353.

Harmer, J. 2007. *The Practice of English Language Teaching.* Essex: Pearson Education Limited.

Hinton, P., Brownlow, C., McMurray, I., and Cozens, B. 2004. *SPSS Explained.* Sussex: Routledge.

Holliday, A. 2005. *The Struggle to Teach English as an International Language.* Oxford: Oxford University Press.

Hughes, R. 2002. *Teaching and Researching Speaking.* London: Longman.

Jenks, C. 2007. Floor management in task-based interaction: the interactional role of participatory structures. *System* 35: 609-622.

Jenks, C. 2009. Exchanging missing information in tasks: old and new interpretations. *The Modern Language Journal* 93: 185-194.

Johnson, K. 1988. Mistake correction. *ELT Journal* 42: 89-96.

Johnson, K. 1996. *Language Teaching and Skill Learning*. Oxford: Blackwell.

Kormos, J. and Denes, M. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32: 145-164.

Kumaravadivelu, B. 2007. Learner perception of learning tasks. In Van den Branden, K., Van Gorp, K., and Verhelst, M. (Eds.), *Tasks in Action: Task-Based Language Education from a Classroom-based Perspective.* Newcastle: Cambridge Scholars Publishing.

Lambert, C. and Engler, S. 2007. Information distribution and goal orientation in second language task design. In del Pilar Garcia Mayo, M. (Ed.) *Investigating Tasks in Formal Language Learning.* Clevedon: Multilingual Matters.

Lai, C. and Zhao, Y. 2006. Noticing and text-based chat. *Language Learning and Technology* 10: 102-120.

Larsen-Freeman, D. 2006. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics* 27: 590-619.

Larsen-Freeman, D. 2009. Adjusting expectations: the study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30: 579-589.

Larsen-Freeman, D. and Cameron, L. 2008. *Complex Systems and Applied Linguistics.* Oxford: Oxford University Press.

Levelt, W. 1989. *Speaking: From Intention to Articulation.* Cambridge: Cambridge University Press.

Loewen, S. 2007. The prior and subsequent use of forms targeted in incidental focus on form. In Fotos, S. and Nassaji, H. (Eds.), *Form-focused Instruction and Teacher Education Studies in Honour of Rod Ellis*. Oxford: Oxford University Press.

Lynch, T. 2001. Seeing what they meant: transcribing as a route to noticing. *ELT Journal* 55: 124-132.

Lynch T. 2007. Learning from the transcripts of an oral communication task. *ELT Journal* 61: 311-320.

Lynch, T. and Maclean, J. 2000. Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research* 4: 221–50.

Lynch, T. and Maclean, J. 2001. Effects of immediate task repetition on learners' performance. In Bygate, M., Skehan, P., and Swain, M. (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing*. Harlow: Longman.

Lyster, R. 2007. *Learning and Teaching Languages Through Content a Counterbalanced Approach.* Amsterdam: John Benjamins.

Mehnert, U. 1998. The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition* 20: 52-83.

Mennim, P. 2007. Long-term effects of noticing and oral output. *Language Teaching Research* 11: 265-280.

Michel, M., Kuiken, F., and Vedder, I. 2007. The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL* 45: 241-259.

Mochizuki, N. and Ortega, L. 2008. Balancing communication and grammar in beginning-level foreign language classrooms: a study of guided planning and relativization. *Language Teaching Research* 12: 11-37.

Moser, J. 2005. Using language-focuses learning journals on a task-based course. In Edwards, C. and Willis, J. (Eds.), *Teachers Exploring Tasks in English Language Teaching.* New York: Palgrave Macmillan.

Moser, J. 2008. An exploratory research study into classroom-based task repetition. Unpublished paper. University of Birmingham.

Moser, J., Harris, J., and Carle, J. 2012, forthcoming. Improving teacher talk through a task-based approach. *ELT Journal* 66.

Murphey, T. 2004. Participation, (Dis-)identification, and Japanese university entrance exams. *TESOL Quarterly* 38*:* 700-710.

Nation, I.S.P. 2007. The Four Strands. *Innovation in Language Learning and Teaching* 1: 2-13.

Nation, I.S.P. and Macalister, J. 2010. *Language Curriculum Design.* New York: Routledge.

Norris, J. and Oretga, L. 2009. Towards and organic approach to investigating CAF in instructed SLA: the case of complexity. *Applied Linguistics* 30: 555-578.

Nunn, R. 2006. Designing holistic units for task-based learning. *Asian EFL Journal* 8: 69-93.

Ohta, S. 2001. *Second Language Acquisition Processes in the Classroom.* New Jersey: Lawrence Erlbaum Associates.

Pallotti, G. 2009. CAF: defining and differentiating constructs. *Applied Linguistics* 30: 590-601.

Pinter, A. 2005. Task repetition with 10-year-old children. In Edwards, C. and Willis, J. (Eds.), *Teachers Exploring Tasks in English Language Teaching* New York: Palgrave Macmillan.

Prabhu, N.S. 1987. *Second Language Pedagogy.* Oxford: Oxford University Press.

Richards, K. 2003. *Qualitative Inquiry in TESOL.* Hampshire: Palgrave Macmillan.

Rossiter, M.J. 2001. The challenges of classroom-based SLA Research. *Applied Language Learning* 12: 31-44.

Saito-Abbott, Y. 2004. Designing an outcomes-based TBI Japanese language program. In Leaver, B. and Willis, J. (Eds.), *Task-based Instruction in Foreign Language Education Practices and Programs.* Washington: Georgetown University Press.

Samuda, V. and Bygate, M. 2008. *Tasks in Second Language Learning.* New York: Palgrave Macmillan.

Seedhouse, P. 1999. Task-based Interaction. *ELT Journal* 53: 149–56.

Segalowitz, S. 2010. *Cognitive Bases of Second Language Fluency.* New York: Routledge.

Skehan, P. 1998. *A Cognitive Approach to Language Learning.* Oxford: Oxford University Press.

Skehan, P. 2007. Task research and language teaching: reciprocal relationships. In Fotos, S. and Nassaji, H. (Eds.), *Form-focused Instruction and Teacher Education Studies in Honour of Rod Ellis*. Oxford: Oxford University Press.

Skehan, P. 2009. Modelling second language performance: integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics* 30: 510-532.

Skehan, P. and Foster, P. 2005. Strategic and on-line planning: the influence of surprise information and task time on second language performance. In Ellis, R. (Ed.), *Planning and Task Performance in a Second Language*. Amsterdam: John Benjamins.

Smith, B. 2004. Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition* 26: 365-398.

Stillwell. C., Curabba, B., Alexander, K., Kidd, A., Kim, E., Stone, P., and Wyle, C. 2010. Students transcribing tasks: noticing, fluency, accuracy, and complexity. *ELT Journal* 64: 445-455.

Swain, M. and Lapkin, S. 2000. Task-based second language: the use of the first language. *Language Teaching Research* 4: 251-74.

Swan, M. 2005. Legislation by hypothesis: the case of task-based instruction. *Applied Linguistics* 26: 376-401.

Swann, J. 2001. Recording and transcribing talk in educational settings. In Candin, C. and Mercer, N. (Eds.), *English Language Teaching in its Social Context*. London: Routledge.

Tavakoli, P. 2011. Pausing patterns: differences between L2 learners and native speakers. *ELT Journal* 65: 71-79.

Tavakoli, P. and Skehan, P. 2005. Strategic planning, task structure, and performance testing. In Ellis, R. (Ed.), *Planning and Task Performance in a Second Language.* Amsterdam: John Benjamin.

Thornbury, S. 2001. *Uncovering Grammar.* Oxford: Macmillan Education.

Thornbury, S. 2005. *How to Teach Speaking.* Essex: Longman.

Thornbury, S. and Slade, D. 2006. *Conversation: From Description to Pedagogy.* Cambridge: Cambridge University Press.

van Lier, L. 1988. *The Classroom and the Language Learner: Ethnography and Second Language Research.* London: Longman.

Weaver, C. 2007. Willingness to communicate: a mediating factor in the interaction between learners and tasks. In Van den Branden, K., Van Gorp, K., and Verhelst, M. (Eds.), *Tasks in Action: Task-Based Language Education from a Classroom-based Perspective.* Newcastle: Cambridge Scholars Publishing.

Widdowson, H.G. 1983. *Learning Purpose and Language Use.* Oxford: Oxford University Press.

Widdowson, H.G. 2003. *Defining Issues in English Language Teaching.* Oxford: Oxford University Press.

Williams, J. 2005. Form-focused instruction. In Hinkel, E. (Ed.), *Handbook of Research In Second Language Teaching and Learning.* New Jersey: Lawrence Erlbaum Associates.

Willis, J. 2004. Perspectives on task-based instruction: understanding our practises, acknowledging different practitioners. In Leaver, B. and Willis, J. (Eds.), *Task-based Instruction in Foreign Language Education Practices and Programs.* Washington: George Town University Press.

Willis, D. and Willis, J. 2007. *Doing Task-based Teaching.* Oxford: Oxford University Press.

Yalden, J. 1987. *The Principles of Course Design for Language Teaching.* Cambridge: Cambridge University Press.