



Libraries and Learning Services

University of Auckland Research Repository, ResearchSpace

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognize the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](#) and [Deposit Licence](#).

The Effect of Task Structure, Task Repetition, and Reformulation on Foreign Language Written Performance

by

Faezeh Mehrang

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Applied Language Studies and Linguistics.

The University of Auckland

2016

To my parents

&

my husband

Abstract

In the field of task-based language teaching, research on task design and task implementation variables has indicated positive effects of task structure as a task design variable and task repetition as a task implementation variable on oral performance (e.g. Tavakoli & Skehan, 2005; Bygate, 2001). Fewer studies, however, have investigated the effect of these variables on written task performance. To this end, the current study was conducted to explore how written performance is affected by task structure and task repetition. Oral performance studies investigating task repetition have reported that the performance of the same task benefits from task repetition, whereas the performance of a new task does not. Ellis (2009b) claims that positive effects of task repetition are carried over to a new task and learning takes place if feedback is provided between the repeated performances of tasks. Therefore, in this study written feedback (i.e. reformulation) was provided as an intervention variable to investigate how performance of the same task, a new task of the same type, and a new task of a different type are affected by task repetition both in presence and absence of feedback.

This is an experimental study involving four groups of EFL learners ($n = 106$) in two private language schools in Iran (i.e. structured group, structured + reformulation group, unstructured group, and control group). The study took place over a period of five weeks with all groups completing a pre-writing task in week 1 and a post-writing task in week 5. In weeks 2, 3, and 4, the structured group repeated structured tasks, the structured + reformulation group repeated structured tasks and was engaged in reformulation, and the unstructured group repeated unstructured tasks. Task structure was operationalised by adopting Hoey's (1983) problem-solution discourse structure. Written data were analysed in terms of both micro-measures (i.e. complexity, accuracy, and fluency) and macro-measures (i.e. textual organisation). Scores were compared across groups by means of parametric/non-parametric ANOVAs to discover the effects of the variables.

No significant improvements were found with regard to the performance of the same task. There was, however, strong evidence of the effectiveness of repeating structured tasks on the structural complexity and accuracy of the performance of a new task of the same type. Repeating the unstructured tasks facilitated the development of textual organisation of the performance of a new task of a different type and led to the production of well-organised texts in this group. Findings have implications for both theory and pedagogy. I propose that micro

aspects of written performance can best be promoted by providing learners opportunities for structured task repetitions, whereas unstructured task repetitions are beneficial in improving macro aspects of written language. I also suggest that writing education in Iran is reinforced by development of formal syllabi and training writing teachers.

Acknowledgements

I owe an enormous debt of gratitude to all the many people who have given me continued support over the course of my education, especially in the past few years.

To my main supervisor, Prof. Gary Barkhuizen for seeing my thesis over the finish line, responding to submissions with lightning speed and giving invaluable comments. To Dr. Shaofeng Li, my co-supervisor, for his fast reading of drafts, his comments and positive encouragement. To distinguished Prof. Rod Ellis, my first main supervisor, for launching me on this journey, his invaluable comments on the design of this study, and his assistance with writing up the results chapters.

To everyone at Auckland University who went out of their way to support me throughout my studies – the team at the Auckland University Student Association, at the School of Graduate Studies, and at the School of Cultures, Languages, and Linguistics, including Denise Lazelle, Prof. Laurence Simmons, Dr. Wendy-Llyn Zaza, and because he cannot be thanked enough, Prof. Gary Barkhuizen once again.

To the participants of both the pilot and main study for their continued participation and to my colleagues, Sahar Beheshti, Andre Breedt, and Madeleine Matanawa for helping with reformulating students' texts, rating data, and proofreading.

To PaCE at Massey University, especially Pamela Resnick, Liz Tully, and Andrea Flavel, for so much support and flexibility over the last two years.

To all my teachers from past to present, especially Prof. Masoud Rahimpour for his fatherly love and advice, as well as his support and encouragement during the process of applying for this PhD.

Above all to my friends and wonderful family, especially to my parents for emotional as well as initial financial support, and my amazing husband for keeping me sane and moving forward. Without you... there are no words.

Thank you.

Table of Contents

Abstract.....	iii
Acknowledgements	v
Table of Contents	vi
List of Tables	xi
List of Figures.....	xiv
Chapter 1 Introduction.....	1
1.1. Writing Education in Iran.....	1
1.2. English as a Foreign Language Education in Iran.....	3
1.3. Personal Rationale.....	5
1.4. Theoretical and Pedagogical Rationale	7
1.5. The Present Study	9
1.6. Structure of the Thesis	10
Chapter 2 Literature Review	11
2.1. Theoretical Background	11
2.1.1. Limited Attentional Resources Model	11
2.1.2. Kellogg’s Model of Writing	12
2.1.3. Writing and Speech: Similarities and Differences.....	15
2.2. Task-based Language Teaching Methodology	17
2.3. Task Design Features	18
2.3.1. Task Structure	19
2.3.2. Task Structure and Oral Performance.....	19
2.3.3. Task Structure and Written Performance	22
2.4. Task Implementation	23
2.4.1. Task Repetition	24
2.4.2. Task Repetition and Oral Performance	25
2.4.3. Task Repetition and Written Performance.....	31
2.5. Reformulation.....	35
2.5.1. Studies Investigating Reformulation	37
2.6. Teaching Discourse Structure	43
2.7. Research Questions	46

Chapter 3 Pilot Study	47
3.1 Research Questions	47
3.2. Setting.....	48
3.3. Overview of the Design	48
3.4. Participants	49
3.5. Instruments	49
3.5.1. Tasks	49
3.5.2. Reformulated Texts	52
3.5.3. Questionnaires	53
3.6. Procedures.....	55
3.6.1. Structured Group	55
3.6.2. Structured + Reformulation Group.....	55
3.6.3. Unstructured Group.....	56
3.6.4. Control Group.....	56
3.7. Coding and Scoring.....	56
3.7.1. Micro-measures	56
3.7.2. Macro-measures.....	56
3.8. Data Analysis	57
3.9. Summary of Main Findings.....	59
3.10. Problems Encountered in the Pilot Study	60
3.10.1. Low Proficiency Level of the Participants	60
3.10.2. High Mortality Rate	60
3.10.3. Participants' Low Level of Motivation.....	60
3.10.4. Operationalising Reformulation in Class	61
3.10.5. Effect of Memory on Reproduction Task Performance.....	61
3.10.6. Holistic Scale of Textual Organisation	61
3.10.7. Reformulating Student Texts	61
3.10.8. Questionnaire.....	62
3.11. Proposed Changes for the Main Study.....	62
3.11.1. Changes to the Design of the Study.....	63
3.11.2. Changes to the Research Questions	64
 Chapter 4 Methodology	 65
4.1. Overview of the Design	65
4.2. Participants	66

4.3. Setting.....	67
4.4. Materials for Data Collection.....	68
4.4.1. Problem-solution Discourse Structure.....	68
4.4.2. Tasks	70
4.4.3. Reformulated Texts	74
4.4.4. A piece of EFL Writing and its Reformulated Version	74
4.4.5. Questionnaires	74
4.5. Procedures for Data Collection.....	75
4.5.1. Structured Group	75
4.5.2. Structured + Reformulation Group.....	77
4.5.3. Unstructured Group.....	79
4.5.4. Control Group.....	79
4.6. Data Coding and Scoring.....	79
4.6.1. Micro-measures	80
4.6.2. Macro-measures.....	82
4.7. Statistical Analysis.....	93
 Chapter 5 Results and Discussion 1	95
5.1. Statistical Analysis.....	95
5.2. Results for Research Question One	96
5.2.1. Structural Complexity.....	96
5.2.2. Lexical Complexity	97
5.2.3. Grammatical Accuracy	98
5.2.4. Mechanical Accuracy	99
5.2.5. Fluency 1 (total number of words per text).....	100
5.2.6. Fluency 2 (number of words per T-unit)	101
5.2.7. Summary of the Results for Research Question One.....	102
5.3. Results for Research Question Two.....	102
5.3.1. Structural Complexity.....	103
5.3.2. Lexical Complexity	104
5.3.3. Grammatical Accuracy	104
5.3.4. Mechanical Accuracy	106
5.3.5. Fluency 1 (total number of words per text).....	107
5.3.6. Fluency 2 (total number of words per T-unit)	108
5.3.7. Summary of the Results for Research Question Two	108

5.4. Discussion of the Results of Research Questions One and Two.....	110
Chapter 6 Results and Discussion 2	122
6.1. Micro-measures	122
6.2. Macro-measures	122
6.3. Statistical Analyses.....	126
6.4. Within-group Differences	127
6.4.1. Structured Group	127
6.4.2. Structured + Reformulation Group.....	128
6.4.3. Unstructured Group.....	129
6.4.4. Control Group.....	130
6.4.5. Summary	132
6.5. Between-group Differences.....	133
6.5.1. Structural Complexity.....	133
6.5.2. Lexical Complexity	134
6.5.3. Grammatical Accuracy	135
6.5.4. Mechanical Accuracy	135
6.5.5. Fluency 1 (total number of words per T-unit)	136
6.5.6. Fluency 2 (total number of words per text).....	137
6.5.7. Idea Expression.....	138
6.5.8. Paragraph Division and Focus	139
6.5.9. Presence and Order of the Problem-solution Structure Elements.....	140
6.5.10. Development of the Problem-Solution Structure Elements.....	142
6.5.11. Summary	143
6.6. Discussion of the Results of Research Question Three	144
6.6.1. Overall Results.....	146
6.6.2. Within-group Differences	147
6.6.3. Between-group Differences	153
Chapter 7 Conclusion	155
7.1. Summary of Objectives and Design.....	155
7.2. Summary of the Key Findings.....	156
7.3. Research Implications.....	157
7.3.1. Theoretical Implications	157
7.3.2. Pedagogical Implications	159

7.4. Limitations of the Study	163
7.5. Suggestions for Future Research	165
References.....	167
Appendix A: Ethics Forms.....	179
Appendix B: Reproduction Tasks	183
Appendix C: Questionnaires (Farsi Version)	189
Appendix D: Pilot Study Results	191
Appendix E: Texts to Teach Reformulation.....	198
Appendix F: Example of a Reformulated Text	199
Appendix G: Main Study Group Differences.....	200
Appendix H: Main Study T1, T2, and T3 Descriptive Statistical Results	203
Appendix I: Mixed ANOVA Results for Research Question 3.....	206

List of Tables

Table 2.1 Task Characteristics and Influence upon Performance and Research Basis (Skehan, 2003)	18
Table 2.2 Implementation Variables During Each Phase of a Task-based Lesson (adapted from Ellis & Shintani, 2014).....	24
Table 3.1 Counterbalanced Sequence of the Free-production Tasks	49
Table 3.2 Design of the Pilot Study	50
Table 3.3 Writing Scale for Measuring Holistic Text Organisation	57
Table 3.4 Writing Scale for Measuring Problem-solution Textual Organisation	58
Table 3.5 Design of the Main Study	63
Table 4.1 Counter-balanced Sequence of the Pre- and Post-Writing Tasks	66
Table 4.2 General Writing Scale (Stage 3)	84
Table 4.3 Guidelines for Employing the General Writing Scale (Stage 3).....	85
Table 4.4 Problem-solution Structure Writing Scale (Stage 3).....	86
Table 4.5 Guidelines for Employing the Problem-solution Structure Writing Scale (Stage 3)	87
Table 4.6 General Writing Scale (Final)	89
Table 4.7 Guidelines for Employing the General Writing Scale (Final)	90
Table 4.8 Problem-solution Structure Writing Scale (Final)	91
Table 4.9 Guidelines for Employing the Problem-solution Structure Writing Scale (Final)..	92
Table 5.1 Descriptive Statistics Results for T1 – T2 Structural Complexity Gain/Loss Scores	96
Table 5.2 Descriptive Statistics Results for T1 – T2 Lexical Complexity Gain/Loss Scores.	97
Table 5.3 Descriptive Statistics Results for T1 – T2 Grammatical Accuracy Gain/Loss Scores	98
Table 5.4 Descriptive Statistics Results for T1 – T2 Mechanical Accuracy Gain/Loss Scores	99
Table 5.5 Descriptive Statistics Results for T1 – T2 Fluency1 Gain/Loss Scores.....	100
Table 5.6 Bonferroni Post Hoc Test Results for Fluency 1 Gain/Loss Scores	100
Table 5.7 Descriptive Statistics Results for T1 – T2 Fluency2 Gain/loss Scores.....	101
Table 5.8 Kruskal-Wallis H Post Hoc Test Results for Fluency 2 Gain/loss Scores.....	102

Table 5.9 Descriptive Statistics Results for T1 – T3 Structural Complexity Gain/Loss Scores	103
Table 5.10 Bonferroni Post Hoc Test Results for Structural Complexity Gain/Loss Scores	103
Table 5.11 Descriptive Statistics Results for T1 – T3 Lexical Complexity Gain/Loss Scores	104
Table 5.12 Descriptive Statistics Results for T1 – T3 Grammatical Accuracy Gain/Loss Scores	105
Table 5.13 Bonferroni Post Hoc Test Results for Grammatical Accuracy Gain/Loss Scores	105
Table 5.14 Descriptive Statistics Results for T1 – T3 Mechanical Accuracy Gain/Loss Scores	106
Table 5.15 Bonferroni Post Hoc Test Results for Mechanical Accuracy Gain/Loss Scores	107
Table 5.16 Descriptive Statistics Results for T1 – T3 Fluency 1 Gain/Loss Scores.....	107
Table 5.17 Descriptive Statistics Results for T1 – T3 Fluency 2 Gain/Loss Scores.....	108
Table 5.18 Summary of the Main Results of Research Questions One and Two	109
Table 6.1 General Writing Scale	123
Table 6.2 Guidelines for Employing the General Writing Scale	124
Table 6.3 Problem-solution Structure Writing Scale	125
Table 6.4 Guidelines for Employing the Problem-solution Structure Writing Scale.....	126
Table 6.5 Descriptive and Inferential Statistics for the Structured Group’s Pre-W – Post-W Performance	128
Table 6.6 Descriptive and Inferential Statistics for the Structured + Reformulation Group’s Pre-W – Post-W Performance.....	129
Table 6.7 Descriptive and Inferential Statistics for the Unstructured Group’s Pre-W – Post-W Performance	130
Table 6.8 Descriptive and Inferential Statistics for the Control Group’s Pre-W – Post-W Performance	131
Table 6.9 Summary of the Main Results for Pre- vs. Post-writing Performance.....	132
Table 6.10 Descriptive Statistics Results for Pre-W – Post-W Structural Complexity Gain/Loss Scores	133
Table 6.11 Descriptive Statistics Results for Pre-W – Post-W Lexical Complexity Gain/Loss Scores	134
Table 6.12 Descriptive Statistics Results for Pre-W – Post-W Grammatical Accuracy Gain/Loss Scores	135

Table 6.13 Descriptive Statistics Results for Pre-W – Post-W Mechanical Accuracy Gain/Loss Scores	136
Table 6.14 Descriptive Statistics Results for Pre-W – Post-W Fluency 2 Gain/Loss Scores	137
Table 6.15 Descriptive Statistics Results for Pre-W – Post-W Fluency 1 Gain/Loss Scores	137
Table 6.16 Bonferroni Post Hoc Test Results for Fluency 1 Gain/Loss Scores	138
Table 6.17 Descriptive Statistics Results for Pre-W – Post-W Idea Expression Gain/Loss Scores.....	138
Table 6.18 Descriptive Statistics Results for Pre-W – Post-W Paragraph Division and Focus Gain/Loss Scores	139
Table 6.19 Kruskal Wallis H Post Hoc Test Results for Paragraph Division and Focus.....	140
Table 6.20 Descriptive Statistics Results for Pre-W – Post-W Presence and Order of Elements Gain/Loss Scores.....	141
Table 6.21 Kruskal Wallis H Post Hoc Test Results for Presence and Order of Elements ..	141
Table 6.22 Descriptive Statistics Results for Pre-W – Post-W Development of Elements Gain/Loss Scores	142
Table 6.23 Summary of the Main Results for Pre-W – Post-W Gain/Loss	143
Table 6.24 Summary of the Main Results for Pre-W – Post-W Gain/Loss with	144

List of Figures

Figure 2.1 Kellog's Model of Writing (Kellog, 1996, p.59).....	13
Figure 4.1 Design of the Main Study	67
Figure 4.2 Problem-solution Discourse Structure (Adapted from Hoey, 2001)	70

Chapter 1 Introduction

Expressing ideas in a second language is no easy task, especially when it is in the form of writing. Second language writing is demanding as it requires acquiring not only the knowledge of the second language but also the skills to meaningfully communicate ideas through writing. This thesis is an attempt to discover ways to facilitate the process by investigating the effect of task repetition as a task implementation variable, task structure as a task design variable, and reformulation as an intervention variable on the performance of L2 written tasks in two language schools in Iran. In this chapter, I outline and explain the reasons why I chose this topic for my thesis. I first briefly explain the educational background in Iran in terms of teaching Farsi writing and teaching English. I then focus on my personal rationale both as an English learner and teacher, and finally, introduce the theoretical rationale of my topic.

1.1. Writing Education in Iran

In this section, I explain the type and quality of Farsi writing education the participants had received prior to participation in this study, to provide an insight into their minds and shed light on their writing background.

In recent years, public education in Iran has been undergoing transition (policy NO. 101/1/22734). However, since the participants of the present study were educated in the old system, I explain that system and how it was structured, especially with regard to teaching Farsi writing.

Previously, public education in Iran was divided into six levels including one year of pre-school, five years of primary school, three years of junior high school, three years of high school, one year of pre-university school, and finally, tertiary education. Tertiary education in Iran has been similar to tertiary education systems in many other countries. That is, it includes four years of study for Bachelor's Degrees, two years for Master's Degrees and three to four years for Doctoral Degrees (Farhady et al., 2010). Farsi writing has always been one of the many subjects in the curriculum at all levels. However, in the old system, it was not taught to meet students' needs in their professional life after they graduated. That is, students were unable to write well-structured essays or texts after they graduated from high school.

In primary school, writing carried less importance and there was no independent text book for writing. However, as said, it was included in the curriculum under the title of “composition”. A text book called ‘Farsi’ was used to teach the Farsi alphabet. The book included very short passages to teach each letter. That is, the alphabet was taught in context using short passages full of words containing the target letters. The focus was mainly on spelling, and regular assessments in the form of dictation were carried out to test students’ spelling skills. When composition was concerned, at lower levels, the focus was on constructing sentences. Students were given words and were asked to make sentences using those words. At higher levels, there was emphasis on writing compositions, but no teaching of any kind was offered to help students write. In fact, once a week, students were given a topic and were asked to write about it at home. They were then randomly called by the teacher to the front of the class to read their compositions to the other students and the teacher, who would mark it after the students had finished reading. Usually, students did not write their compositions themselves and asked their parents or an older sibling for help. Marking was completely subjective and students did not know why they got a certain mark. To assess writing at the end of the year, students were given two topics and were asked to choose one to write on. Assessment topics were chosen from the topics that had been covered in class during the year and usually students were told indirectly which topics were the most important, so they memorised the compositions they had written in class on those topics to be ready for the assessments.

In junior high school, high school and pre-university school, a textbook called ‘Farsi Language’ was used to teach writing and Farsi language. The book included four sections – grammar, linguistics, spelling, and writing skills. Among all sections, the main focus both in the book and in class was on grammar. That is, most of the lessons were devoted to teaching Farsi grammar, so the teachers spent much of the class time teaching grammar. Spelling and linguistics were of secondary importance and writing was the least important section with very few lessons focusing on writing skills. Moreover, the content of the lessons was mostly of an introductory type. That is, lessons were designed to introduce different genres without focusing on how to write any of them. In addition to these, teachers were not trained to teach writing, and in most cases teachers themselves could not write well-structured texts. Final assessments also mainly involved grammar, spelling, punctuation, sentence construction and use of literary techniques and devices.

At tertiary level, a two-credit paper called ‘General Farsi’ was a compulsory paper to be passed by students of all majors. However, the focus of this paper was ‘Farsi Literature’ and no teaching of writing skills was offered.

To conclude, compared to other subjects in the curriculum, Farsi writing was the least important in Iran and no instruction of any kind was offered in schools or universities to help improve students’ writing skills. However, in September 2012, there was a big change in the Iranian Education system. In the new system, writing is of more importance and there is an independent textbook for Farsi writing and composition at every level. Teaching writing through these textbooks is more systematic and there are lessons on how to organise a text into paragraphs and how to structure a proper paragraph. Of course, this system, as mentioned, has been put into practice very recently and whether it is implemented and practised in the classroom and whether the teachers are trained to teach the skills in the new books is yet to be investigated.

1.2. English as a Foreign Language Education in Iran

English language education has a long history in Iran. During the Pahlavi Dynasty (1925-1979) and due to rapid westernisation English became popular in Iran as a social need, a necessary requirement both in the job market and for securing scholarships to continue education overseas. However, after the Islamic revolution in 1979 and following changes in the infrastructure of the country, which involved the implementation of Islamic rules and the replacement of personnel, English was considered to be a foreign language. It was called an ‘alien’ language because it was the official language in the USA and the UK, which were perceived as Iran’s enemies (Farhady et al., 2010; Sadeghi & Richards, 2015). As a result, private English language institutes all closed down (Aliakbari, 2002; Farhady et al. 2010) and English language teaching was nationalised with the development of standard textbooks. Moreover, the learning and teaching of five other foreign languages, including German, French, Italian, Spanish, and Russian, were promoted to replace teaching English. However, as Farhady et al. (2010) state, since there was not a sufficient number of teachers and applicants for these languages, English remained the most dominant foreign language at high schools.

In the above-mentioned national system of English teaching, English was introduced in the second year of junior high school (Ghorbani, 2009). According to Sadeghi and Richards (2015), the syllabi and the content were fixed for all schools, and teachers were not authorised

to make any changes to the content or structure of the course. Course materials in junior high school mainly focused on alphabet recognition, pronunciation, and limited vocabulary instruction, whereas at high school, the focus was on reading comprehension, grammar, and vocabulary development; speaking skills were limited to grammar practice drills, and writing beyond decontextualised sentences and listening skills were ignored. In addition to the inappropriate locally developed textbooks, the system also followed a traditional teacher-centered approach and a grammar-translation method with teachers being unable to communicate in English themselves (Davari & Aghagholzadeh, 2015). This resulted in a failure of the system; after six years of formal English instruction, students normally were unable to communicate in English (Sadeghi & Richards, 2015).

At tertiary level, there was a compulsory 3-unit credit general English course for all university students, and students could also be required to take up to 4 units of ESP courses depending on their needs (Farhady et al., 2010). Unlike pre-university education, there were no prescribed syllabi for teaching English and tutors had to develop their own syllabi and materials. Some tutors chose books published outside of Iran, while others used books developed by SAMT (the organisation responsible for developing tertiary educational materials and textbooks). A typical English course at university focused on writing at the sentence level, reading comprehension, grammar, and vocabulary, ignoring listening and speaking skills and contextualised writing (Sadeghi & Richards, 2015). Teaching methodology at universities was translation-based as the main purpose was to help students read and understand the subject-based materials written in English (Farhady et al., 2010).

During the second decade of the revolution and due to the dissatisfaction with the national English language teaching system, the private English language institutes, which had been closed down after the revolution, reopened and new institutes were established (Farhady et al., 2010; Davari & Aghagholzadeh, 2015).

Since the third decade of the revolution, English has become even more popular as a result of globalisation and has been viewed as the language of communication, science, and technology. There has been a growing interest among young people to move to other countries to continue their education. English has become an essential requirement for employment in the private sector. Young people have become interested in communication with other cultures and travelling to other countries has become popular (Davari & Aghagholzadeh, 2015; Sadeghi & Richards, 2015). The need to be a proficient English speaker led to the establishment of private

English language institutes (Davari & Aghagholzadeh, 2015) where books published outside of Iran, such as *Interchange* and *Headway*, have been used. Private language schools have implemented communicative language teaching methods and emphasised the development of communicative skills. However, following the belief fostered by the way writing was taught in Farsi, the main focus in teaching English writing has also been on sentence structure and grammar with no emphasis on paragraph development and text structure.

Although private institutes have always been innovative and dynamic by trying to incorporate the latest methods of language teaching and the latest materials into their syllabi, the public education system had not undergone any changes with regard to its structure until very recently when there was a major change in the education system in all areas, including English. In the current system, materials are designed to develop English proficiency and the focus is on communicative abilities. However, since learners and sometimes the teachers are not able to communicate in English, classes are conducted in Farsi and the grammar-translation method is used in teaching. Consequently, as it was with the previous system, students fail to learn English in public schools.

1.3. Personal Rationale

In this section, I provide some information about my personal experiences in Iran, both as an English learner and an English teacher, and how they shaped the research I carried out for my thesis.

I started to learn English at a private language institute in Iran when I was eleven years old and continued to take English classes until I graduated from high school. At university, I chose to study English Language and Literature for my Bachelor's degree and got my first part-time job as an English language teacher at a private institute. This was the beginning of my career as an English teacher.

After I graduated from university, I knew I did not want to stop there and was keen to learn more. I therefore decided to study a Master's degree in English Language Teaching. I first found the shift from 'English Literature' to 'English Teaching' challenging, but then I started to develop an interest in 'English Teaching' since what I was learning was helping me facilitate my teaching. When I was writing my Master's research proposal, I chose to investigate the effects of 'task structure' and 'planning conditions' on the oral performance of English learners

in Iran. I chose this topic based on both my experience as an English teacher – I was trying to find ways to facilitate spoken language as it was the primary interest and the main goal of most of my students – and my reading – mainly motivated by Skehan and Foster's (1999) article on the effects of task structure and processing conditions on narrative retellings. As expected, the results indicated positive effects of both 'task structure' and 'planning conditions' on oral task performance.

Since I could also observe that students found writing in English extremely challenging (mainly because English writing skills were not taught in language schools), I felt responsible to help them improve their writing skills. However, I was not competent in writing myself because I had not received any instruction when English writing was concerned. My vocabulary and grammar knowledge was at an acceptable level, but I knew it was not sufficient to produce a well-structured piece of writing. I had to learn what ideas I was supposed to include in my writing and how to organise them into paragraphs and into a full text. Thus, I decided to first improve my own writing skills.

I started with reading articles and books and later practised what I was learning. It did not take long before I developed the necessary writing skills and an awareness of how a decent text is organised. Subsequently, I focused on designing materials to teach my students what I had learned. At the time, task-based language teaching was becoming popular in language schools in Iran, especially in the school that I was working in. As a result of this shift in language schools and because of my background in task-based research, I wondered whether the same variables that I tested in my Master's research – task structure and task planning – could also improve students' writing skills. I started reading so that I could find the answer to my question. However, I realised that there was a big gap in the task-based literature where writing was concerned. I found out that almost all research investigating the effects of task structure and task planning was conducted on oral performance, which left me with an unanswered question and led me to design and carry out the current research.

As a result of further reading, I decided to focus on the effects of 'task structure' and 'task repetition' on written performance. I was interested in 'task structure' because as mentioned above, the results of previous research (e.g. Skehan & Foster, 1999; Tavakoli & Skehan, 2005) show positive effects of structured tasks on oral performance and I was curious to find out if the same results would be borne out when writing was considered. I chose 'task repetition'

mainly because as a teacher, I believed in practice and wanted to investigate whether practising and repeating written tasks would result in a better written performance.

As I read more about task repetition, I realised that the research on the effects of task repetition on oral performance reported positive effects of repetition on the performance of the same task, but not on the performance of a new task. Ellis (2009b) claimed that the reason for task repetition failing to develop interlanguage was that no feedback was offered between the repeated performances. Consequently, I added a third variable to the design of my study; to investigate how ‘repetition only’ compared to ‘repetition + feedback’ would affect written performance. In the following section, I will provide the theoretical rationale of the study.

1.4. Theoretical and Pedagogical Rationale

Task-based Language Teaching (TBLT) was developed as a result of dissatisfaction with former language teaching methods, such as, audiolingualism and grammar translation, and is primarily concerned with developing communicative competence rather than focusing on language structure (Ellis & Shintani, 2014). In other words, TBLT emphasises the creation of meaning encouraging learners to use any language resources available to them to communicate the desired meaning and achieve task goals (Willis & Willis, 2001).

Like any other language teaching methodology, task-based syllabus design involves making decisions about what to teach; that is, task design. According to Skehan (2003), tasks should be designed according to specific characteristics (see Skehan, 2003, for the list of characteristics) as these can affect performance in different ways in terms of complexity, accuracy, and fluency. It is therefore evident that syllabus designers should first make decisions regarding which aspect of performance learners need to develop and improve, and then design tasks with characteristics that promote those aspects.

Task structure is one of the characteristics that Skehan (2003) lists that affects task performance by increasing its fluency and accuracy. Task structure is believed to affect performance in a positive way since, according to Skehan’s (1998b) limited attentional resources theory, it eases the processing burden on memory and sets attentional resources free to be allocated to both fluency and accuracy (Skehan & Foster, 1999; Tavakoli & Skehan, 2005).

In addition to task design, task implementation is also a factor which can affect how students perform a task. Task repetition as a task implementation variable gives learners the opportunity

to rehearse task performance at least once prior to the main performance of the task. When learners repeat a task for a second time, they can access the cognitive work undertaken during the first performance, and so have attention free to attend to all areas of performance including complexity, accuracy, and fluency (Bygate, 2001).

However, although task repetition eases the processing demands on working memory and sets attentional resources free to be allocated to improving performance areas (i.e. complexity, accuracy, and fluency), the research evidence suggests that the positive results are not carried over to the performance of a new task. Ellis (2009b) suggests that for positive effects of task repetition to be carried over to a new task (i.e. for language learning to take place) some type of feedback should also be offered between the repeated performances of the tasks.

To investigate Ellis's suggestion, in the current study, I offered feedback in the form of reformulation (rewriting students' texts while preserving the meaning but correcting all grammatical, lexical, and organisational errors) to some learners between their repeated performances to discover the effects of 'task repetition only' compared to 'task repetition + feedback' on the performance of both the same task and a new task.

It should be noted that the above theories and the findings of the research investigating the effects of task structure and task repetition on performance are exclusive to oral tasks and spoken language. Although according to Levelt's (1989) speech production model and Kellogg's (1996) model of writing, both speakers and writers go through similar processes to produce speech or texts, writing is different from speaking in that writing does not happen in real time. This allows more time for task execution and eases the processing constraints that apply to speech. It is thus possible that writers have more attention free to attend to form-meaning mappings than speakers do (Manchon, 2014), hence making it significant to study the effects of task design and task implementation variables on written tasks and performance, which have been neglected in the field of task-based language teaching.

The unsatisfactory teaching of English writing in Iran as well as the personal, theoretical, and pedagogical rationale explained above, led to the design of the present study which will briefly be explained in the next section.

1.5. The Present Study

To address the gap in the literature of task-based language teaching with regard to written tasks and performance, the present study investigates the effect of task structure as a task design feature, task repetition as a task implementation variable, and reformulation as an intervention variable on written performance. The aim is to answer the following research questions to discover how these variables affect written performance:

1. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of *the same task* in terms of micro-measures?
2. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of *a new task of the same type* in terms of micro-measures?
3. What is the effect of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of *a new task of a different type* in terms of both micro- and macro-measures?

To this end, 106 participants in the study were divided into three experimental groups – structured, structured + reformulation, and unstructured – and a control group. The experimental groups performed five written tasks over a course of five weeks, whereas the control group only performed two tasks with a five-week interval. That is, all groups first performed a pre-writing task which was a free-production task. The experimental groups then performed three note-expansion tasks with a week's interval between each performance. The 'structured' and the 'structured + reformulation' groups performed structured note-expansion tasks, whereas the 'unstructured' group performed unstructured note-expansion tasks. The 'structured + reformulation' group received reformulation as feedback between the first and the second performances of the note-expansion tasks. Finally, all groups performed a post-writing task which was a free-production task.

1.6. Structure of the Thesis

This thesis consists of seven chapters. Chapter 1 introduces the writing educational background in Iran and provides theoretical as well as my personal rationale for the study. It presents the aim of the study and outlines the organisation of the whole thesis. Chapter 2 explains the theoretical and pedagogical background of the study and reviews the empirical research on the efficacy of task structure, task repetition, and reformulation on task performance. Chapter 3 provides a brief report of the pilot study and outlines the problems identified and the adjustments made for the main study. Chapter 4 presents the methodology of the main study by describing the research settings, participants, instruments, and data collection, as well as data analysis, procedures. Chapter 5 reports the results of the first and the second research questions and discusses the main findings. Chapter 6 presents the results of the third research question and provides a discussion of the key findings. Chapter 7 summarises the study and the main findings and considers some theoretical and pedagogical implications for task-based writing instruction. The chapter concludes with an outline of the limitations of the study and suggestions for future research.

Chapter 2 Literature Review

The study reported in this thesis investigates the role task structure as a task design variable, task repetition as a task implementation variable, and reformulation as an intervention variable in the development of second language academic writing in terms of both language and textual structure. The aim is to examine how written performance is influenced by certain task characteristics and task implementation variables. To provide a rationale for the study, this chapter outlines the theoretical and pedagogical background to the study and presents a detailed review of the related literature. I start with explaining the underlying cognitive processes involved in performing a task and summarising the processes involved when a text is produced. I also outline the similarities and the differences between writing production and speech production. I then briefly review TBLT and define a ‘task’. Next, I focus on task design features, particularly task structure, and task implementation variables, especially task repetition. The focus is then on reformulation as a type of written feedback. The final section of the chapter reviews studies which investigated the effect of teaching discourse structure on textual organisation. It is worth mentioning that there are very few studies investigating the effect of either task structure or task repetition on *written* performance as studies in these areas focus mainly on oral performance. This chapter therefore includes a review of both oral performance and written performance studies.

2.1. Theoretical Background

2.1.1. Limited Attentional Resources Model

According to Schneider and Shiffrin (1977) and Shiffrin and Schneider (1977), human attention is limited and mastering new skills, like driving, is very attention-demanding at the beginning stages making it impossible for learners to allocate attention to other things simultaneously. However, through practice and automatisisation, performance of new skills requires less attention. Applying this model to language acquisition, Skehan (1998b) suggests that when learning a new language, learners have limited attentional resources and cannot attend to all three aspects of performance, namely fluency, accuracy, and complexity, at the same time. As a consequence, they prioritise one area over another. In other words, these three areas enter into competition with one another for attentional resources and focusing on one will reduce the attention which would otherwise be available to be allocated to other areas.

To further explicate how attention is afforded to different areas of performance, Skehan (1998b) distinguishes an exemplar-based and a rule-based system. The exemplar-based system consists of discrete lexical items as well as ready-made formulaic chunks which can easily and quickly be retrieved and result in fluent language performance. On the other hand, the rule-based system is made up of abstract representations of the underlying patterns of language. These patterns require more processing and so are utilised in more controlled but less fluent language, which Tarone (1983) calls “the careful style of language” as it is more accurate and complex. Skehan suggests that learners draw on the exemplar-based system when producing fluent language but the rule-based system when producing accurate and complex language. Complexity and accuracy are different as accurate language is produced as a result of learners trying to use the existing resources in a controlled way avoiding errors, while complex language is a result of learners taking risks to restructure. According to this view, there is a possibility that trade-offs occur between fluency, accuracy, and complexity (between accuracy and complexity in particular) and so, focusing on one of these areas will be at the expense of the other (Ellis, 2005).

In addition to Skehan’s limited attentional resources model, the present study draws on Kellogg’s (1996) model of writing to explain the processes writers go through when producing written texts. This model is summarised in the following section.

2.1.2. Kellogg’s Model of Writing

There are a number of current theories (e.g. Grabe & Kaplan, 1996; Kellogg, 1996; Zimmerman, 2000) to account for how written language is produced. However, the model proposed by Kellogg (1996) has been widely used and cited in the literature, and therefore will also be used as the basis of this study to allow the comparability of the findings with the other studies. Also this model provides a clear picture of detailed processes and sub-processes that are involved in text production.

2.1.2.1. Systems and Sub-processes

Based on his early research (Kellogg, 1987, 1988, 1990) and drawing on Baddeley’s model of working memory (1986), Kellogg (1996) divided text production process into three main systems: (a) formulation, which involves planning and translating, (b) execution, which involves programming and executing, and (c) monitoring, which involves reading and editing. The model is illustrated in Figure 2.1 below.

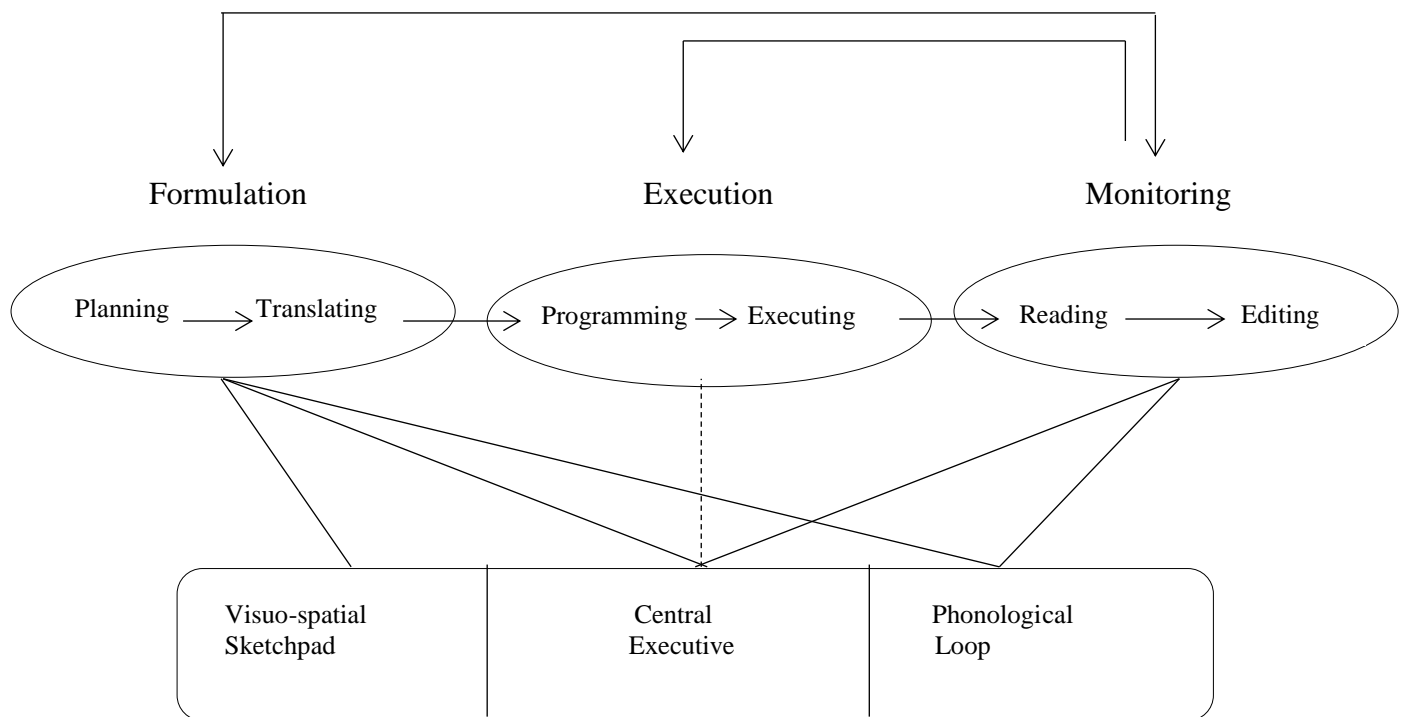


Figure 2.1 Kellogg's Model of Writing (Kellogg, 1996, p.59)

As the arrows in the figure show, the components of the model overlap and take place in parallel. That is, it should not be seen as a process through which a text is first fully formulated, then executed, and then monitored, but as a dynamic system which involves simultaneous activation of all processes and sub-processes. For instance, when a word or phrase is executed, a new piece of material can also be formulated or a previously executed piece of information be monitored.

Formulation involves two sub-processes of *planning* and *translation*. Planning involves goal setting, “thinking up ideas” to match the goals, and organizing ideas. The output of the planning stage can be either propositional representations or abstractions and feelings. Propositional representations can easily be forwarded to the translation stage, whereas abstractions and feelings need to undergo further planning as they “defy easy translation”. Translating involves activating semantic, syntactic, phonological, and orthographic sub-processes to select lexical units, build a syntactic frame, phonologically represent the lexical units in the frame, and convert phonemes into graphemes. As a result of this sub-process, an idea is converted into a written message. Planning is usually accompanied by partial translation in the form of Vygotsky's (1962) inner speech and Witte's (1987) pre-text. In other words, when planning,

the writer also has mental realisations of word meanings and their phonological representations with a “sketchy” structure which accounts for inner speech. These mental realisations can also be in the form of sentences which the writer tries out and edits before executing them (pre-text). Partial translation can also manifest itself in the form of visual symbols, diagrams, abbreviations, and topical outlines which carry ideas, but not adequate lexical and syntactic work is done to make the ideas intelligible for the reader.

Execution consists of two sub-components; *programming* and *executing*. Through programming, the writer converts the product of the translation process into production schema for the appropriate motor system in dictation, typing, or handwriting. Executing is the actual production of the written words, phrases, and sentences. Depending on the output mode, programming and executing components differ. That is, in dictation, the muscle movement system of speech is activated. In typing, arm, hand, and finger muscles are activated, and in handwriting, not only the size of the letters and their graphic form must be selected, but also arm, hand, and finger muscles must be activated.

Monitoring involves two sub-processes; *reading* and *editing*. When reading, the writer needs to activate several sub-processes to recognise words, understand sentences, and make connections between sentences and between the different parts of the text. When editing, writers make comparisons between their intentions and the actual product of each process. In other words, they evaluate their texts to find problems and fix them.

2.1.2.2. Demands on Working Memory

Kellogg argues that among the three main systems explained above, formulation and monitoring are considered as controlled systems and greatly demand the working memory. As Figure 2.1 illustrates, formulation places demands on visuo-spatial sketchpad, central executive, and phonological loop. Monitoring demands the phonological loop and the central executive. It is clear that formulation demands are larger than the monitoring demands. Execution demands only the central executive, and these demands are negligible when typing and hand-writing skills are automatised. However, Ellis (2005) argues that execution may occur automatically in the case of “adult, native-like automaticity in hand-writing or typing” but not in the case of L2 learners, especially those with a different L1 script. As a result, for L2 learners, all three systems are likely to tax the central executive, which Kellogg believes has a limited capacity. In other words, an L2 writer needs to make decisions on which process to prioritise when engaged in text production, especially when this involves time constraints.

Kellogg claims that formulation is critical and will be prioritised over execution and monitoring when trade-offs happen. This is to some extent supported by De Larios, Marin and Murphy's (2001) study, which examined the effect of L2 proficiency on on-line processing of written language. Results of their study indicate that writers in general focused on producing pre-texts and texts when writing (partial translation, translation, and execution processes in Kellogg's model) but tended to share the writing time with other processes (i.e. developing ideas and making changes to the written text) when they were not under pressure. Pressure in their study was operationalised through L2 proficiency. That is, low level of L2 proficiency was characterised as leading to greater demands on working memory while high L2 proficiency resulted in less pressure.

As mentioned earlier, task-based language teaching research has mainly investigated oral tasks and oral performance. There are only a few studies which have addressed written performance, and these studies mainly focused on the effects of strategic or online planning (e.g. Ellis & Yuan, 2004; Ellis & Yuan, 2005; Ong & Zhang, 2010; Johnsen et al., 2012; Abedifirouzjaie, 2014) or the effects of task complexity (e.g. Kuiken et al., 2005; Kuiken & Vedder, 2007a, 2007b, 2007c, 2011, 2012; Kormos, 2011; Ruiz-Funes, 2014; Revesz et al., 2016). Therefore, the present study is an attempt to examine written performance to discover whether and how written language is affected as a result of certain task design and task implementation variables. It is thus worth outlining the similarities and the differences between speech and writing in order to compare the effect on written performance.

2.1.3. Writing and Speech: Similarities and Differences

The similarity between writing and speech can be seen when the processes posited for written language production by Kellogg (1996) are examined next to the speech production model proposed by Levelt (1989). Like Kellogg, Levelt proposes a speech production model which consists of three main processes: conceptualisation, formulation, and articulation. Through conceptualisation, the speakers first determine the goal which they want to communicate and then develop the communicative goal into a number of sub-goals and map a speech act on to them to achieve their intentions. Finally, through the micro-planning stage, each of the sub-goals is realised by retrieving the necessary information. The product of this stage is a preverbal message which is ready to be converted into language. Formulation is concerned with the linguistic representations of the preverbal message and is realised by retrieving lemmas and lexemes from the speaker's mental lexicon. Since lemmas contain information regarding the

meaning and syntax of the lexical items, and lexemes carry morphological and phonological features, retrieving this information will result in building a surface structure as well as an articulatory plan of an utterance. These representations are called ‘internal speech’ by Levelt (1989) and are temporarily stored in the articulatory buffer. Finally, internal speech is forwarded to the last stage, articulation. During this stage, internal speech is retrieved and speech is produced and articulated. Levelt further asserts that there is a self-monitoring process where speakers monitor the output of every stage of the speech production process. It should be noted that these stages work in parallel; otherwise, there would be a lot of long pauses in speech when speakers go through these stages in real time. When comparing this model and Kellogg’s model (see 2.2), it becomes evident that both models put forward similar sets of processes for speech and written language production.

Although both speech and written language go through similar systems and processes, Manchon (2014) lists two major differences between them. First, unlike speech, apart from computer assisted online writing, most forms of writing do not occur in real time. That is, speech takes place in real-time and “is generally intolerant of significant pauses” (Ellis & Yuan, 2005, p. 174) whereas writers have the opportunity to “take time-out from on-line production” (Ellis & Yuan, 2005, p. 175). Manchon (2014) believes this removes the processing constraints that apply to speech by allowing more time for task execution. This is important, since when writing attention is released not only by certain attention releasing variables (task structure and task repetition in the case of this study) but also by availability of time, and as a result learners get the chance to better plan and formulate their messages and attend to form-meaning mappings (Manchon, 2014). Ellis and Yuan’s (2005) study provides support to Manchon’s claim. In their study, Ellis and Yuan compared the effect of careful vs. on-line planning on low proficient Chinese learners’ spoken and written language in terms of fluency, complexity (syntactic and lexical), and accuracy. Findings indicate that learners were less fluent, but more accurate, and more syntactically and lexically complex in their written production than in their oral production. Like Manchon, Ellis and Yuan explained these results by referring to the nature of writing. That is, writing does not occur in real time and results in a visual object thus making it easier to plan, formulate, execute, and monitor messages than speaking does. The second distinction Manchon (2014) argues is the availability of feedback in the case of writing. However, I will not consider this since feedback or any other type of intervention is also available in the case of oral language (e.g. Baleghizahed & Derakhshesh, 2012; Hawkes, 2011; Matsumura et al, 2008).

Ellis and Yuan (2005, p.175) list another distinction between speech and writing. That is, unlike speech, the result of writing is a “visual object (the written message), which is amenable to inspection in a way that the aural trace left by a spoken message is not”. Ellis and Yuan believe that this makes mistakes clear and therefore facilitates monitoring. It is clear from the above distinctions that writing is less demanding than speech and so, in the presence of attention releasing variables, writers should benefit more than speakers. The next section addresses task-based language teaching methodology to provide the pedagogical context of the study.

2.2. Task-based Language Teaching Methodology

According to Skehan (2003), during the 1980s there was a move both in second language pedagogy (Prabhu, 1987) and research (Long, 1989) towards activities that focused on communicating and conveying meaning rather than focusing on language structure. This move resulted in a major development in the field of SLA: the emergence of task-based language teaching (TBLT) with ‘task’ as the central unit “for both designing a language program and for planning individual lessons” (Ellis, 2009a, p.223). Task-based language teaching is based on the principle that language learning will be more successful if it is learned in contexts where “learner’s natural language learning capacity can be nurtured” (Ellis, 2009a, p. 222) instead of just teaching language systematically (Ellis, 2009a).

Tasks are the central units of TBLT and “their design can affect their use by teachers in the classroom, the actions of learners, and the performance and learning outcomes” (Bygate, Skehan & Swain, 2001, p.1). It is therefore important to define ‘task’ before any research can be conducted in this area. Although there are many definitions of ‘task’ available in the literature (see Ellis, 2003, pp. 3-4), giving a satisfactory definition to distinguish it from other types of instructional activities is somewhat problematic within task-based language teaching. Skehan (1998a) proposes four qualifying criteria for a task: (1) meaning is primary. (2) There is a goal which needs to be worked towards. (3) The activity is out-come evaluated. (4) There is a real world relationship (p. 268). However, these criteria are critiqued by Widdowson (2003) who argues that they are “loosely formulated” and fail to distinguish tasks from other kinds of language-teaching activities. In response to this criticism, Ellis (2009a), drawing on available definitions of a task, lists four criteria for a language-instructional activity to be qualified as a ‘task’; first, tasks should primarily engage learners in expressing and understanding both semantic and pragmatic meanings; second, tasks should create a need to convey information, express an opinion or infer meaning (i.e. a “gap”); third, task completion should involve

learners' own linguistic and non-linguistic resources, and finally, a task should have a clear outcome other than displaying correct language. That is, language should be used to achieve that outcome (Ellis, 2009a). Ellis and Shintani (2014) claim that these criteria will make it possible to distinguish a 'task' from other kinds of teaching activities as they ensure that when performing tasks, learners use language as a 'tool' to communicate rather than as an 'object' to learn.

2.3. Task Design Features

Like with any other language teaching methodology, task-based syllabus design involves making decisions about what to teach – task design. According to Skehan (2003), tasks should be designed considering specific characteristics which can affect performance in different ways in terms of complexity, accuracy, and fluency. These characteristics and the way they might affect performance are summarised in Table 2.1. It is evident that syllabus designers should first make decisions regarding which aspect of performance learners need to develop and then design tasks with characteristics which promote those aspects. One thing to be taken into account about Skehan's summary is that the research he is referring to is on oral performance.

Table 2.1 *Task Characteristics and Influence upon Performance and Research Basis (Skehan, 2003)*

Task Characteristics	Influence upon Performance and Research Basis
Structured tasks, i.e. clear timeline or macro-structure	Clearly greater fluency, tendency towards greater accuracy (Foster & Skehan, 1996; Skehan & Foster, 1997, 1999)
Familiar information	Greater fluency and greater accuracy (Foster & Skehan, 1996; Skehan & Foster, 1997)
Outcomes requiring justifications	Justifications lead to markedly greater complexity of language (Skehan & Foster, 1997)
Interactive vs. monologic tasks	Interactive tasks produce markedly more accuracy and complexity, monologic tasks more fluency (Foster & Skehan, 1996, 1999; Skehan & Foster, 1997, 1999)

As Table 2.1 shows, task structure as a task characteristic seems to considerably affect oral task performance and is therefore worthy of research. However, to be able to design structured tasks, task structure needs to be defined.

2.3.1. Task Structure

Skehan (2003) defined a structured task as a task with a clear time line or a macrostructure. The most comprehensive definition of task structure, however, was offered by Tavakoli and Skehan (2005) who claimed that a task was structured if it had the following features:

- A clear time line;
- A script;
- A story with a conventional beginning, middle and end;
- An appeal to what is familiar and organised in the speaker's mind; and finally,
- A problem-solution structure.

Tavakoli and Skehan (2005) further argued that it is not necessary for a structured task to have all these features at the same time. For example, a task can be structured if it only has a conventional beginning, middle, and end. However, there are different degrees of structure and a task with a conventional beginning, middle, and end as well as a problem-solution structure is more structured than a task with only a conventional beginning, middle, and end (Tavakoli & Skehan, 2005). One thing to note about the above definitions is that they only apply to narrative tasks, and so, task structure can be defined and operationalised in different ways if different task types are considered.

2.3.2. Task Structure and Oral Performance

Based on Levelt's (1989) speech production model and Skehan's (1998b) limited attentional resources model, it is hypothesised that tasks with a clear macro-structure, where one step in the task leads to the next, eases the processing burden of the task and sets attentional resources free to be allocated to different areas of performance (i.e. complexity, accuracy, and fluency) and results in more fluent, and more accurate performance. This hypothesis was proposed as a result of post-hoc analyses of earlier research (i.e. Foster & Skehan, 1996; Skehan & Foster, 1997) and has been empirically tested on oral performance since then. The following section provides a review of these studies.

2.3.2.1. Studies Investigating Task Structure

Skehan and Foster (1999) examined the effect of task structure and processing conditions on L2 task performance. To operationalise task structure, they used two Mr. Bean video prompts, 'Crazy Golf' and 'Mr. Bean goes to the restaurant'. The first episode involved Mr. Bean playing a round of golf where he starts the game making a very bad shot and hitting the ball outside the golf area. Having been warned not to touch the ball as a game rule, he ended up hitting the ball all over the town trying to get it back on the golf course. This task was an unstructured task because the series of events that occurred lacked interconnectedness and could not be predicted. In the restaurant task, however, Mr. Bean goes to a restaurant, gets the menu, and orders food. Since the events of this episode are predictable and sequenced, this task was regarded as a structured task. The 47 participants of the study were required to do these tasks under four different conditions: (a) watch and tell simultaneously (the most demanding), (b) storyline given, watch and tell simultaneously, (c) watch first, then watch and tell simultaneously, and (d) watch first, then tell (the least demanding). Data were analysed in terms of complexity, accuracy, and fluency. Results indicate that as predicted, fluency increased considerably as a result of performing the structured task, whereas complexity was left unaffected, and accuracy increased only when the structured task was performed under the fourth condition (watch first, then tell).

Subsequent to the above study, Tavakoli and Skehan (2005) investigated the effect of different degrees of task structure on the performance of 80 learners with different proficiency levels (elementary and intermediate) when they performed four tasks (no structure, schematic structure, weak causation structure, and problem-solution structure) under planned (five-minute planning time) vs. unplanned (no planning time) condition in a testing context. The first two tasks were considered as unstructured tasks, while the other two were treated as structured tasks. Task structure was operationalised through picture narrative tasks. The most structured task was a picture series with a clear sequential structure as well as a problem-solution structure, while the less structured task was a picture series with a sequential organisation and a weak causation structure. The unstructured tasks lacked a problem-solution and a causative structure. The less unstructured task had a sequential organisation, whereas the more unstructured task had no sequential organisation. Data were analysed in terms of complexity, accuracy, and fluency. Findings indicate that structured tasks resulted in more accuracy and fluency. Complexity results were complicated as performance on one of the structured tasks was more complex than the others. Regarding the planning condition, all performance areas

including complexity, accuracy, and fluency improved when participants did the tasks under the planned condition. To further investigate the effect of task structure on complexity, the researchers conducted a post-hoc analysis which clarified that it was the background information (foreground information which describes the main points and the central propositions of the story vs. background information which deals with the ideas and comments which are made to elaborate and explicate the foreground information [Tomlin, 1984]) present in one of the structured tasks, not the degree of structure of the task, which led to the increased complexity.

Post-hoc findings of Tavakoli and Skehan (2005) led to the design of a study by Tavakoli and Foster (2008) who examined the effect of task structure and storyline complexity on second language oral performance. Their study involved 100 participants (60 in Iran and 40 in England). The reason for this grouping was to see if the educational setting affected how learners performed. To operationalise task structure, the researchers employed four picture stories, two of which had a very tight structure (i.e., it was impossible to reorder the different events of the task without compromising the story), whereas the other two were loosely structured (the events could easily be reordered without changing the story). Results indicate that performance on the structured tasks was more accurate, and storyline complexity resulted in more complex performance. A comparison of the data from Tehran and London learners revealed that learners in London produced significantly more syntactically and lexically complex language.

In a similar study, Tavakoli (2009) investigated the effect of task structure and storyline complexity on oral task performance in a testing context. She defined task structure in terms of either a problem-solution structure (Hoey, 1983) or a schematic sequential organisation (Tavakoli & Skehan, 2005; Tavakoli & Foster, 2008). Her study involved 60 participants who performed six narrative tasks of different degrees of storyline complexity and structure in a testing context. Results provide further support to the findings of the previous studies. In other words, task structure resulted in more accurate and more fluent performance and tasks with both foreground and background information led to more syntactically complex language.

While all of the above studies focused on single performance of structured and unstructured tasks, Saeedi and Kazerouni (2014) investigated the effect of task structure when learners engaged in repeated performances of narrative structured tasks as opposed to unstructured tasks. In their study, task structure was operationalised following Tavakoli and Skehan (2005).

Participants, who were 60 students studying English at a university in Iran, were divided into four groups. The first group performed the unstructured task only once while the second group performed the same task twice with an interval of one week. The third group completed the structured task only once, whereas the fourth group completed the same task twice with an interval of one week. Data were analysed in terms of complexity, accuracy, and fluency. Group comparisons were made between the structured group and the structured repetition group as well as the unstructured group and the unstructured repetition group. Results indicate that repeating the unstructured task resulted in more complex and more fluent performance, whereas repeating the structured task promoted all three areas of complexity, accuracy, and fluency. No results were reported with regard to structured vs. unstructured task repetition.

In a more recent study, Ahmadian et al. (2015) examined the effect of task structure and online planning on the performance of 60 intermediate-level English learners. Narrative tasks based on animated videos were used in this study, and task structure was operationalised following Tavakoli and Skehan (2005). Participants were divided into four groups. The first group performed the structured task under careful online planning conditions. The second group performed the structured task under pressured online planning conditions. The third group completed the unstructured task while engaging in careful online planning. Finally, the fourth group carried out the unstructured task under pressured online planning conditions. The pressured online planning groups had only five minutes to perform the task. Data were analysed with regard to complexity, accuracy, and fluency. Findings indicate that performing the structured task under careful online planning conditions led to the most complex, accurate, and fluent performance, whereas performing the unstructured task under pressured online planning conditions resulted in the least complex, accurate, and fluent performance.

To summarise, in all of the above studies task structure was operationalised following Tavakoli and Skehan's (2005) criteria and employing narrative tasks. Findings indicate that by and large, tasks with a clear structure promote accuracy and fluency of oral performance. Complexity is not affected by task structure alone, but is improved when task structure is accompanied by another attention releasing task implementation condition (e.g. task repetition or careful online planning).

2.3.3. Task Structure and Written Performance

As mentioned earlier in this chapter, based on Levelt's (1989) speech production model and Kellogg's (1996) model of writing, underlying processes for the production of oral and written

language are the same. However, what makes writing different is that it does not happen in real time, and so it is hypothesised that different task design or task implementation variables might affect written performance in a different way from oral performance.

2.3.3.1. Studies Investigating Task Structure

To the best of my knowledge, there is only one study that investigated the effect of task structure on written performance. Rahimpour et al. (2011) investigated the written performance of 32 students studying English in a university in Iran. Task structure was operationalised following Tavakoli and Foster (2008). Two picture series were used; one with a problem solution structure in which the events could not be reordered and the other with no sequential structure allowing the events to be reordered. Each participant was required to perform two tasks (structured and unstructured) during two sessions. Data were analysed based on complexity, accuracy, and fluency (within group comparisons). Results indicate that task structure promoted the complexity and fluency of the written performance leaving accuracy unaffected. Findings of the above study suggest that task structure can affect written performance in a positive way, but more research is needed to shed light on the way in which written language can be influenced by task structure.

Another interesting point to note is that task structure in the area of both oral and written performance has mainly been investigated on the single performance of the task. In other words, only one study has addressed task structure effects on performance by engaging students in repeated performances of structured tasks. This creates another gap in this area. After tasks are designed, decisions need to be made on how they can be implemented in a language classroom. The following section focuses on task implementation, especially task repetition as the task implementation variable investigated in the current study.

2.4. Task Implementation

Ellis and Shintani (2014) claim that implementing a task-based syllabus involves considering the design of a lesson, the participatory structure of the lesson, and the roles the teacher and the students adopt.

The design of a task-based lesson, as Skehan (1996b) and Ellis and Shintani (2014) argue, has three stages: pre-task stage, during-task stage, and post-task stage. There are a number of options which can be implemented during each stage. Ellis and Shintani (2014) propose a list of these options in each phase of a task-based lesson (see Table 2.2). These options each affect

how students perform a task. For example, pre-task planning makes it easier for learners to perform the task, while time pressure, which is a possible during-task variable, makes performing a task difficult for learners.

Table 2.2 *Implementation Variables During Each Phase of a Task-based Lesson (adapted from Ellis & Shintani, 2014)*

Phase	Options
Pre-task	<ol style="list-style-type: none"> 1. Modelling performance of the task 2. Pre-teaching language 3. Schema-developing 4. Strategic planning
Main-task	<ol style="list-style-type: none"> 1. Time pressure 2. Contextual support 3. Explicit instruction 4. Surprise element
Post-task	<ol style="list-style-type: none"> 1. Repeat performance 2. Report 3. Language work

Pre-task planning is a pre-task phase implementation variable which can be operationalised in two ways: strategic planning or task repetition. Strategic planning provides learners with the opportunity to prepare for task performance by allowing them time to generate the ideas they will need to perform the task and to plan ways to express those ideas. Task repetition gives learners the opportunity to rehearse the task at least once before the main performance (Ellis, 2005).

Since in this study ‘task repetition’ is employed as the task implementation variable, the following section provides a detailed account of what task repetition is and how it can be implemented.

2.4.1. Task Repetition

Task repetition is a task implementation variable which has generated a great deal of interest in the area of TBLT and oral language production since 1996 when Bygate first attempted to

explore task repetition effect on the performance of oral narrative tasks. As mentioned above, task repetition provides a chance for learners to practise task performance at least once before the main performance. That is, the second performance of the task is the main performance while the first performance serves as rehearsal. Task repetition has been investigated in three different types: exact task repetition, content task repetition, and procedural task repetition (Manchon, 2014). Exact task repetition refers to performing the same task (same content, same procedures) several times. Content repetition refers to performing several tasks which have the same content but require different procedures to achieve the communicative goal. Procedural repetition refers to performing several tasks with different content but the same procedure to achieve the communicative goal.

2.4.2. Task Repetition and Oral Performance

When oral performance is considered, task repetition is believed to have beneficial effects on performance. That is, since oral communication takes place in real time, giving learners the opportunity to repeat a task would ease the processing burden posed by the task because the speaker is able to access the cognitive work undertaken previously to conceptualise information and this sets attentional resources free for new cognitive processes. The repeat performance therefore would be more complex, more accurate, and more fluent (Bygate, 2001).

To validate the above assumption, a large body of research in the area of TBLT has focused on investigating the effect of different types of task repetition on oral performance. According to Manchon (2014), these studies fall into three groups: (a) studies that investigate the effect of task repetition and task-type repetition on the characteristics and features of oral production (e.g. Bygate, 1996, 2001; Gass et al 1999; Kim & Tracy-Ventura, 2013; Lynch & Mclean, 2000; Patanasorn, 2010; Saeedi & Kazerouni, 2014), (b) studies that examine the effect of different types of task repetition on attention to language/focus on form in language related episodes (e.g. Kim, 2013), (c) studies that investigate the effect of different types of task repetition on the use of discourse and interactional features (e.g. Plough & Gass, 1993; Mackey et al., 2007). However, these three groups of studies do not exhaust the other ways task repetition effects can be investigated. That is, there are studies that have either looked at students' and teachers' perceptions of the usefulness of task repetition (Kim, 2013) or investigated the combined effect of task repetition and some kind of intervention between the repeated performances (Baleghizadeh & Derakhshesh, 2012; Hawkes, 2011; Sheppard, 2006).

In the following section, some of the above studies which are within the scope of this study are reviewed briefly.

2.4.2.1. Studies Investigating Task Repetition

Bygate (1996) designed a pilot study to explore how task repetition affects second language oral performance. In his study, Bygate required a single participant to perform the same task (an episode of a Tom and Jerry cartoon) twice with an interval of three days under the same conditions. The only difference between the two occasions was that at the second encounter, the learner was familiar with the task. Bygate posited a number of possibilities regarding the effect of task repetition. First, there would be no differences between the two performances because either the learner starts afresh in the second encounter with the task, and the performance is simply a reflection of his overall proficiency, or a single exposure to the task is not sufficient to affect subsequent performance. Second, there would be “some random” differences between the performances; that is, some aspects of language would be more successful at time one and others at time two. And finally, since according to Levelt’s (1989) model of speech production (see 2.1.3), spoken language involves some planning work before it is produced, the third possibility was that repeating a task for a second time would involve less planning work and set attention free to be allocated to formulation rather than content and would therefore lead to better language production in terms of complexity, accuracy, and fluency. The findings of the study strongly supported the last hypothesis. The learner mostly focused on retrieving the necessary information to communicate at time one while monitoring this language and trying to better formulate it at time two, which resulted in more complex, more accurate, and more fluent performance.

Since the study above only examined the performance of a single participant, Bygate (2001) examined the effect of task repetition and task-type practice on a number of L2 learners’ oral performance. His study involved two types of tasks (narrative and interview tasks) and required the participants to engage in repeated performances of the tasks over ten weeks. The main purpose was to explore whether task repetition or practising a task-type in a language classroom resulted in better performance of a new task of the same type and facilitated form-meaning mappings under pressured conditions where L2 learners needed to identify the meanings they wanted to convey and to access appropriate forms to attach to those meanings. Results of the study indicate that practising a certain type of task over a period of time influenced performance of the same task by promoting complexity, accuracy, and fluency but had no effect on the

performance of a new task of the same type. That is, when participants performed the same task after ten weeks, their performance was more complex, accurate, and fluent, but no changes were found when they performed a new task of the same type.

Similar results were reported in Gass et al. (1999) who examined the effect of exact and procedural task repetition on the oral task performance of Spanish learners. Their study involved three groups of learners: two experimental and a control group. One of the experimental groups performed the same task three times with an interval of two to three days and completed a new task of the same type a week after their last performance (exact task repetition). This was the same for the second experimental group except that they did not repeat the same task. Instead, they completed a new task of the same type on each occasion (procedural task repetition). The control group, however, only performed the first and the fourth tasks which were the same for all groups. Results indicate that there was only a limited influence of exact task repetition on overall proficiency, target-like use of the verbs *ser* and *estar*, and lexical sophistication, but this effect was not carried over to the new task. In other words, task repetition did not affect the performance of a new task of the same type.

While the two studies reported above investigated the effect of task repetition on the performance of both the same task and a new task, the following two studies only focused on the performance of a new task. Kim and Tracy-Ventura (2013) investigated the effect of exact task repetition and procedural task repetition on performance. Their study followed a pre-test – post-test design and involved two experimental groups. In the exact repetition group, participants performed the same information-exchange task three times before post-test, while the procedural repetition group performed three different information-exchange tasks and were then post-tested. Results indicate that the procedural repetition group produced more syntactically complex performance at the post-test. Both groups improved with regard to task-induced linguistic features but showed no change regarding fluency.

Another study which investigated the effects of different types of repetition on performance of a new task was Patanasorn (2010). The study examined the effects of exact, content, and procedural task repetition on accuracy and fluency and so involved three experimental groups. Each group was pre-tested, received treatment, and was given an immediate and a delayed post-test. The exact task repetition group performed the same task three times. The content repetition group performed three tasks with the same content but different procedures, and finally, the procedural repetition group performed three tasks with different content but the same

procedure. Results revealed that procedural repetition resulted in improved accuracy – use of past simple tense – in the performance of a new task, whereas content repetition improved fluency. Exact repetition did not have notable effects on the performance of a new task.

In general, results of the studies reviewed above are indicative of positive effects of task repetition on the performance of the same task. However, when performance of a new task is considered, there are controversies in the findings of different studies. Bygate (2001) and Gass et al. (1999) found no transfer of the positive effects of task repetition to the performance of a new task, whereas the other studies (Kim & Tracy-Ventura, 2013; Patanasorn, 2010) reported gains in the performance of a new task. The difference between the results of these two groups of studies can be attributed to the type of the tasks they used – monologic in Bygate (2001) and Gass et al. (1999), but dialogic in Kim and Tracy-Ventura (2013) and Patanasorn (2010). That is, Bygate (2001) and Gass et al. (1999) used narrative tasks, whereas Kim and Tracy-Ventura (2013) used communicative information-exchange tasks, and Patanasorn (2010) used decision-making information-exchange tasks. The question then is what needs to be done if positive effects of task (monologic task) repetition are to be transferred to a new task or in other words, if language development is desired? Ellis (2009b) suggests that offering some kind of intervention between the repeated performances might assist students to allocate attention to the performance of the task and result in improvements in the performance of the same task or a new task. This is in a way proved by the results of the studies above which employed dialogic tasks in their studies (Kim & Tracy-Ventura, 2013; Patanasorn, 2010). That is, dialogic tasks require learners to collaboratively perform tasks with peers and since collaboration can provide scaffolding during task repetition, it can be considered a kind of intervention. The three studies that investigated the effect of task repetition involving intervention will be reviewed in the following section.

2.4.2.2. Studies Investigating Task Repetition and Intervention

Baleghizahed and Derakhshesh (2012) investigated whether giving learners a second opportunity to perform the same task, after their errors have been corrected, would result in any gains in their second performance with regard to error correction. Their study involved 4 intermediate level female Iranian students who were required to read a graded reader chapter by chapter and orally retell the plot in class. When presenting, the students' voices were audio recorded, and the recordings were given to the students to transcribe. The students then corrected the transcribed version of their own presentations and gave them to the second

researcher who further corrected the errors and in some cases, provided some explanations as to why a certain structure was erroneous (e.g. since you are narrating a story, you should use past simple) and gave the transcripts back to the learners to study before they presented them the second time. To analyse data, the number of errors students had in their first performance, the number of errors they corrected in their transcripts, and the number of errors they corrected on the second performance were counted. To locate errors, the focus was on verb usage, direct translation from students' L1, noun modifiers and prepositions, and some less frequently occurring errors categorised as 'others'. Findings indicate that the four students made 105 errors in their first performance, of which 48 were corrected in the transcripts and 51 were avoided at the second performance. It can thus be concluded that the reactive focus on form led to improvements in the second performance of the same task.

In a similar study, Hawkes (2011) examined the effect of focus on form activities on the second performance of the same task in a Japanese junior high school. The task sequence implemented in this study was pre-task, main task, form focus, and repeat performance. The pre-task stage involved brainstorming activities to introduce the topic as well as theme-based listening where two advanced speakers performed the same tasks participants were about to perform. The pre-task stage was followed by the main task where the 12 participants were required to perform three tasks in pairs (opinion exchange, describe and draw, timed conversation). The form-focus stage involved consciousness-raising activities and explicit presentation of language items to help students notice the useful language which can help them perform the tasks more successfully. Finally, students were asked to repeat the same tasks with the same partners. Data were analysed qualitatively in terms of the use of the target language and the accuracy (form and pronunciation) of that language. Findings of the study revealed that as a result of the form-focus stage, students were able to incorporate the target language practised into their repeat performance and increase the number of corrections (form correction in all three tasks; pronunciation correction in the 'describe and draw' task only).

In a semi-longitudinal case study, Matsumura et al. (2008) investigated the effect of procedural task repetition on oral performance of a new task in terms of structural complexity, accuracy, and fluency. In their study, tasks were used in normal language classes where students engaged in Willis' (1996, 2004) sequence of pre-task, task cycle and post task language focus. The study involved two Japanese participants who were required to perform nine different tasks of the same type – one performed nine narrative tasks and the other completed nine decision-making tasks – in nine lessons. The participants were first pre-tested by performing a narrative task and

giving a speech on a given topic. They then received nine lessons in three weeks and were finally post-tested by performing a narrative task (different from pre-test) and giving a speech on a topic which was different from the pre-test speech topic. As mentioned, the lessons took place in the usual language classes the participants attended. During the lessons, the participants were guided on the use of helpful lexical items as a pre-task activity. At the task cycle stage, however, each participant received a different treatment as they were from different classes and performed different types of tasks – narrative and decision making. In the class with narrative tasks, after students narrated a story to their partners, they were given time to discuss the storyline and the expressions which could help them in performing the task. They were then required to narrate the story to each other again. The lessons concluded with a brief focus on the task-related grammar and vocabulary. In the class with decision-making tasks, the students were first asked to individually make a decision on a given topic and were then required to work in pairs with their partners to reach a joint conclusion in ten minutes. Then, in pairs, they prepared a presentation and presented it to a new partner. The whole process was followed by a task-related grammar and vocabulary review. Partners changed every session, so the participants worked with different partners on each occasion. Overall, both participants performed more complex, more accurate, and more fluent language at post-test, but the way they gained in performance was different (i.e. the participant who performed the narrative tasks gained more in accuracy and fluency, whereas the one who performed the decision-making tasks produced more complex and more fluent language) which can be attributed to the nature of the tasks they performed (narrative *vs.* decision making).

It is difficult to reach a robust conclusion about the effects of task repetition plus a kind of intervention from the results of the above studies as there are some differences between them. That is, two of the studies (Baleghizadeh & Derakhshesh, 2012; Hawkes, 2011) investigated the effect of task repetition plus intervention on the performance of the same task, while Matsumura et al. (2008) examined the performance of new tasks. The type of tasks they used and the intervention they employed between the repeated performances were also different. Moreover, they used different measures and different data analysis procedures. In addition to the differences among these studies, they are also limited by the fact that they did not include a control group (task repetition only group), and had a limited number of participants. This gap, therefore calls for more research in this area.

2.4.3. Task Repetition and Written Performance

As discussed earlier, although the underlying cognitive processes in producing written language (Kellogg, 1996) is almost similar to those of speech production (Levelt, 1989), writing is different from speech in a number of ways (see 2.1.3), and in theory, written performance might be affected by task repetition in different ways from oral performance. However, not much research has been conducted to study these effects. To the best of my knowledge, there are only five studies that have investigated task repetition effects on written performance. These studies are reviewed in the following section.

2.4.3.1. Studies Investigating Task Repetition

In a longitudinal study based on dynamic systems theory, Larsen-Freeman (2006) studied five high intermediate female Chinese-speaking learners' written performance over a course of six months. The participants were asked to write a narrative based on a past episode that they wanted to share. They were not allowed to consult dictionaries and were encouraged not to worry about whether or not their texts were in "perfect English". Learners wrote the same narrative four times (once every six weeks) without receiving any feedback between the performances. Data were analysed in terms of grammatical complexity, lexical complexity, accuracy, and fluency. Although no statistical analyses were carried out in this study, group averages indicate improvements on all four measures over six months. In other words, participants made improvements on each measure and wrote more grammatically and lexically complex, more accurate, and more fluent texts after every repetition of the task.

Similarly, Indrarathne (2013) reported positive effects of task repetition on written performance in a case study of a 26-year-old Greek female at a mid-proficiency level. The study involved repetition of two tasks over four days. That is, the participant first performed a picture narrative task on day one and repeated the same task on day two. On day three, she performed a new picture narrative task and repeated it on day four. The participant had a maximum of forty-five minutes to perform each task and did not receive any feedback or instruction between the performances. All four performances were analysed in terms of syntactic complexity, lexical sophistication, accuracy, and fluency. Although the findings of the study cannot be generalised as they are of a single participant, results provide some support that task repetition could improve written performance with regard to all three measures, with complexity and accuracy gains being more striking than the fluency gains.

In another longitudinal study (thirty weeks), based on dynamic systems theory, Nitta and Baba (2014) investigated the effect of task repetition on the writing performance of the same task and a new task in terms of syntactic complexity, lexical complexity, and fluency. They conducted the study in real classroom settings with 46 Japanese university students majoring in English. There were two classes in their study: class A with a focus on writing skills and class B with a focus on all four skills of speaking, listening, reading and writing. They used ten-minute writing tasks where students were required to write on a topic for ten minutes without stopping. Topics were chosen with care to make it easy for Japanese university students on the basis of their personal experiences. Each week participants were given a list with three different topics and were asked to choose one which seemed easier for them to write on and then to write a text on it for ten minutes. The topics students were given to choose from changed every two weeks. In other words, their study involved both exact task repetition and procedural task repetition. Student writing was checked every week and returned to them with some feedback the week after. Feedback did not include any linguistic correction and was to encourage positive attitudes towards writing. They analysed data based on syntactic measures (i.e. average sentence length [ASL]), STRUT measures (i.e. sentence syntax similarity, all sentences across paragraphs), lexical measures (i.e. Measure of Textual Lexical Diversity [MTLD] and word frequency values from the CELEX corpus), and fluency measures (i.e. text length). Unlike findings of oral task repetition, results of their study indicate limited effects on the performance of the same task but show notable effects on the performance of a new task in terms of grammatical and lexical aspects. Their study further supports Bygate's (2001) claim that for positive effects of task repetition to be transferred to a new task, more repetitions over an extended period of time is required.

Overall, results of the studies reviewed in this section indicate positive effects of task repetition on the written performance of the same task as well as a new task. However, these findings should be interpreted with care as the first two studies (Larsen-Freeman, 2006; Indrarathne, 2013) did not carry out any statistical analyses and had a limited number of participants. None of the above studies provided feedback or any other type of intervention between the repeated performances and mainly focused on investigating the role of task repetition. The two studies that incorporated some kind of intervention between the repeated performances will be reviewed in the following section.

2.4.3.2. Studies Investigating Task Repetition and Intervention

Jung (2013) reported a pilot study which investigated the effect of task repetition, corrective written feedback, and the interaction of the two on the academic written production. Six L1 Korean-speaking learners of English were randomly divided into four groups: (a) repetition with feedback, (b) repetition with no feedback, (c) no repetition with feedback, and (d) no repetition with no feedback. Two writing tasks from the iBT Test of English as a Foreign Language (TOEFL) were adapted to be used for data collection: Task A and Task B. The study took place over one week. Groups 1 and 2 performed task B on day one and repeated the same task on day seven, whereas groups 3 and 4 performed task A on day one and Task B on day seven. All groups had thirty minutes to perform the tasks. On day three, groups 1 and 3 received unfocused direct corrective written feedback on their first performance with a brief metalinguistic explanation on all lexical, morphological, and syntactic errors and were given fifteen minutes to review the feedback. They then revised their texts without having access to the feedback. No feedback was given with regard to content and organisation. Groups 2 and 4, on the other hand, were given fifteen minutes to have a look at their first performance, find and correct their mistakes independently. First and second performances in all groups were analysed in terms of complexity, accuracy, and fluency. Since the groups were too small to carry out statistical analyses, data were studied in terms of descriptive results. Findings indicate some positive effect of task repetition (i.e. both ‘repetition’ groups improved in terms of accuracy, fluency, *and* complexity, but the ‘no repetition’ groups improved only with regard to accuracy and fluency), while no clear evidence of feedback effect was reported.

In another study, Pakbaz and Rezai (2015) investigated the effect of task repetition and consciousness-raising activities on written task production. Their study involved 60 lower-intermediate English learners (two intact classes): repetition only (control group) and repetition with consciousness-raising (experimental group). The participants’ proficiency level was determined employing the Oxford Quick Placement Test (OQPT). One week after the placement test, both groups were pre-tested using the writing section of the Test of English as a Foreign Language (TOEFL). Later, participants in the experimental group engaged in consciousness-raising activities. Since the focus of the study was ‘article use’ and ‘comparative/superlative forms’, the researchers first developed a two-paragraph text which contained target structures and asked the participants to read the text silently. Secondly, the instructor read a few sentences which included the target structures and encouraged participants to work out the meanings of the structures in the sentences. Thirdly, the whole class did a

number of grammar judgement activities; that is, some grammatically correct and some incorrect sentences were written on the board, and participants were asked to judge whether they were correct or not. In the following stage, the participants were encouraged to come up with a rule for the mentioned structures, and finally, they were asked to make two sentences using the new structures. The control group, on the other hand, only had the opportunity to look at their pre-test writing and try to locate their mistakes and correct them. A week later, both groups were post-tested. The delayed post-test took place four weeks after the initial post-test. The task used for the post-test and the delayed post-test was the same task employed for the pre-test. Data were analysed in terms of complexity and accuracy. Findings indicate that the experimental group increased with regard to accuracy, while the control group improved in terms of complexity. Both effects were long lasting; that is, the results of the delayed post-test also show more accurate performance for the experimental group and more complex performance for the control group.

In a unique study conducted recently, Van de Guchte et al. (2016) examined the effect of task repetition along with corrective feedback on student performance. Their study is unique in that on the one hand, they examined the effects of oral task repetition and feedback on both students' written accuracy as well as oral accuracy and fluency and on the other hand, they investigated whether task repetition effect depends on which grammar structure is being focused on; they studied two German structures (i.e. dative case after a preposition and comparatives). Forty-eight Dutch learners of German as a foreign language were divided into two groups: repetition (R) and no-repetition (NR). The study took place over eight weeks and involved an experimental pre-test – post-test – delayed post-test design for both structures. Groups were first pre-tested on the two structures. They then performed an oral focused task which was designed to elicit German dative structure and received corrective feedback on that target structure as they were performing the task. Two weeks later, the R group repeated a similar task with the same linguistic demands and grammar structures. The NR group, however, did a filler task which was not related to the target structure being investigated. After one week, both groups were post-tested, and the delayed post-test took place two weeks after the first post-test. The same procedures were followed for German comparative structures. Pre- and post-tests examined both students' implicit and explicit knowledge through learners verbalising the target structure rules in metalinguistic knowledge tests, using rules in fill-in-gap exercises, and using both structures in two meaning-based oral tasks. These tests allowed the researchers to test students' declarative knowledge of the target structures, their written accuracy, their oral

accuracy, and their oral fluency. The results of their study show that regardless of grammar structures, the R group outperformed the NR group in terms of metalinguistic knowledge and written accuracy, but no significant differences between the groups were found with regard to oral accuracy and fluency. These findings indicate that students in the R group benefited from repeating similar tasks with the same cognitive demand and target structures to develop understanding of the target structures but needed to engage in more intense repetitions to be able to automatise that knowledge.

In summary, the review of the above studies does not conclusively show clear effects of task repetition along with some type of intervention on written performance; that is, Jung (2013) found no clear effect of repetition and corrective feedback as opposed to repetition only on written performance, whereas Pakbaz and Rezai (2015) reported positive effects on complexity as a result of task repetition only but on accuracy as a result of task repetition and consciousness-raising activities. Van de Guchte et al. found that oral task repetition and corrective feedback help learners develop their declarative knowledge of certain structures and their written accuracy, but not their procedural knowledge. There is therefore a big gap in this area, and research is needed to further investigate the effect of task repetition and task repetition along with intervention on written performance, especially because the above studies are limited in that they either did not include a control group or had a limited number of participants or both. In addition, Van de Guchte et al.'s study did not investigate written accuracy in task performance but through writing exercises.

Since reformulation as a type of written feedback has been used in the present study as an intervention between the repeated performances, the following section explains reformulation and presents a review of the studies in this area.

2.5. Reformulation

Reformulation is defined by Cohen (1983) as “having a native writer of the target language rewrite the learners’ essay, preserving all the learner’s ideas, making it sound as native-like as possible” (p.4). In other words, a native or a native-like speaker of the target language rewrites the students’ texts “using the content the student has provided but recasting it so that the rewritten draft approximated as closely as possible to a putative target language model” (Thornbury, 1997, p.327). According to these definitions, reformulation involves addressing both surface errors and organisational flaws (Swain & Lapkin, 2002).

Reformulation is claimed to have positive effects on improving L2 writing skills based on two hypotheses, namely, the ‘output hypothesis’ (Swain, 1985) and the ‘noticing hypothesis’ (Schmidt, 1990). According to Schmidt, if a form in the target language is to be acquired, it needs to be noticed by the learners. Schmidt and Frota (1986) argued that noticing can take place in two different ways. First, learners pay attention to the language forms in the input that they are exposed to and second, they make “cognitive comparisons” (Ellis, 1995) between their interlanguage and the target language to notice the gap between their own output and the target language system. Swain’s (1985) output hypothesis claims that in addition to comprehensible input, comprehensible output is also necessary for language learning. Swain sees target language production as a trigger that pushes the learners to develop awareness of the inadequacies in their output and the gaps that they need to fill to be able to convey their intended meaning. This, in turn, will promote noticing in subsequent input and affect what becomes intake and what may be learned. To put this another way, Swain (1995) proposed that output is not just an outcome of language learning process but a means for learners to notice linguistic forms, test hypotheses, and develop metalinguistic knowledge.

Thus, learners get to notice the deficiencies in their interlanguage when they are engaged in target language production, especially written production (Qi & Lapkin, 2001). Therefore, providing them with a reformulated version of their text will give them the opportunity to make comparisons between their own text and the reformulated text and actively look for, discover, and notice the gap between the target language and their own interlanguage. In other words, reformulation will act as a catalyst for noticing the gap and thereby promote language development (Adams, 2003).

One could argue that traditional ways of written feedback, say error-correction, would serve the same objectives. However, there are several reasons why reformulation is potentially superior to other types of written feedback. First and foremost, reformulation “provides a good balance between focus on form and focus on meaning since it exploits both meaning-driven and form-focused potential” (Qi & Lapkin, 2001; p. 282) or as Sachs and Polio (2007) assert, reformulation is especially helpful in assisting form-function mappings because learners already know the meaning of their original output and have attentional resources free to focus on form. Second, unlike traditional corrections, which provide negative evidence (Kassen, 1988), reformulation provides both negative evidence and positive evidence about how to express the writer’s own ideas in a target-like way (Adams, 2003) without which “L2 learners are left alone to figure out the solutions themselves” (Qi & Lapkin, 2001, p. 282). Third, as

mentioned earlier in this section, in comparison to typical error correction, reformulation goes beyond the surface level of the text (Thornbury, 1997) and leads to improvements in cohesion (Cohen, 1983). Fourth, reformulation provides target language forms in the same context that the learners themselves provided earlier when performing a written task (Qi & Lapkin, 2001). Fifth, reformulation engages learners in actively searching and discovering the mismatches between their own output and the native-speaker version (Adams, 2003; Sachs & Polio, 2007). As Tocalli-Beller and Swain (2005) argued, reformulation promotes cognitive conflict because it confronts students with information that contradicts their ideas, beliefs, and theories. This conflict, in turn, may result in learning, which is the main goal of providing feedback.

It should, however, be noted that despite the advantages of reformulation mentioned above, there are some concerns about its implementation. “Text appropriation”, as Qi and Lapkin (2001) call it, may be an area of concern. That is, care needs to be taken by the reformulators to “ensure maximal respect for the content of the original text” (Qi & Lapkin, 2001, p. 282). Another problematic aspect of reformulation may be the fact that it is very difficult and time-consuming, especially in EFL settings (Lazaro Ibarrola, 2009; Yang & Zhang, 2010).

The following section is a review of a number of studies that investigated the effect of reformulation as feedback on written performance.

2.5.1. Studies Investigating Reformulation

With a focus on proficiency differences, Qi and Lapkin (2001) conducted a case study in which two participants with different levels of proficiency engaged in performing a three-stage writing task (writing along with think aloud, comparing written output and its reformulated version while thinking aloud, revising the original draft). It was hypothesised that the first writing task would promote awareness of the forms that learners lack in their interlanguage system and would, in turn, draw students’ attention to look for and notice those forms in the reformulated version. However, they argued that the “quality” of this noticing might be different for different proficiency levels and might affect their final output in different ways. Quality of noticing in their explanation was either “perfunctory” (noticing without giving reasons) or “substantive” (noticing as well as giving reasons). The results of the study show that both learners improved their final writing in terms of lexis, form, and discourse, but the degree of improvement differed. The findings indicate that reformulated texts provoked noticing of the changes, but the quality of this noticing was directly related to the improvement. In other words, changes in the final products were mostly the result of substantive noticing, not

perfunctory noticing. Furthermore, the participant with the higher level of proficiency tended to provide reasons for the noticed changes, while the lower-proficient participant only noticed the reformulated forms without giving any reasons due to her limited knowledge of the target language rules and structures.

In another study, Lapkin, Swain, and Smith (2002) investigated the effect of reformulation on the learning of French pronominal verbs in a French Immersion program. Their study focused on proficiency differences as well as task type and involved four pairs of students who were from different French proficiency levels. Two out of four pairs performed jigsaw tasks, while the other two completed dictogloss tasks. The design of the study involved five stages: writing (pre-test), noticing, stimulated recall, post-test, and interviews. In the writing stage, each pair had as much time as they needed to collaboratively write a text. In the noticing stage, they were given their pre-test texts as well as the reformulated texts and were asked to notice the differences between the two texts. In the stimulated recall stage, each pair watched the movie of their noticing stage and was asked what they were thinking about when they compared the two texts. At the post-test stage, the pre-test texts were given back to the participants who were asked to individually make changes to their texts. The interview stage was conducted for each student individually, and they were asked to comment on all stages of the task. Analyses of language related episodes as well as post-test texts indicate that all students regardless of task type or proficiency level made progress in the use of French pronominal verbs at the post-test stage.

While the above studies focused on proficiency differences and task type to investigate the effectiveness of reformulation, Swain and Lapkin (2002) used reformulation as a stimulus for collaborative dialogue. Their study examined two French immersion learners' use of French pronominal verbs. The two participants first wrote a text collaboratively, and later they were given the opportunity to collaboratively compare the text with its reformulated version. Then they were required to participate in a stimulated recall session where they commented on what they were thinking when making comparisons. Finally, they were asked to rewrite their original stories independently. Data were analysed in terms of language related episodes (LREs) – parts of the dialogue where learners discussed their language use, talked about the language they were producing, and corrected their own or another learner's production when engaged in collaborative writing (Swain & Lapkin, 1995). The results indicate that reformulation had a positive effect on the final draft of the text, which Swain and Lapkin argued was because the

comparison stage was done in pairs, and the participants had the opportunity to discuss the reasons for the reformulations.

However, Swain and Lapkin's (2002) study was criticised by Adams (2003), who argued that the findings were not just due to the reformulation, as two other factors, namely task repetition and stimulated recall, were also at work and might have influenced the results. Therefore, Adams designed a study to separate the differential impact of these three variables: reformulation, task repetition, and stimulated recall. Her study followed a pre-test, treatment, and post-test design. Fifty-six Spanish L2 learners were divided into three treatment groups: noticing, noticing + stimulated recall, and task repetition (control). The control group did the same task twice without receiving any treatment in between. The noticing group had the chance to make comparisons between their original texts and the reformulated texts which was hoped to stimulate noticing, and finally, the noticing + stimulated recall group compared their original drafts with the native-speaker version and also attended a stimulated recall session to investigate what they were thinking while noticing the changes between their texts and reformulated texts. Following Swain and Lapkin (2002), the participants did the pre-test in pairs but were post-tested individually. Data were analysed in terms of the number of the reformulated errors that had been corrected in the post-test writing. Results show that all three groups made progress in their final drafts in terms of accuracy, but the degree of improvement was different for the different groups. That is, participation in the noticing + stimulated recall treatment group promoted the incorporation of target-like forms more than participation in the noticing group, and participants in the noticing group outperformed participants in the task repetition group. These findings gave support to Adams' (2003) initial argument that Swain and Lapkin's (2002) results could not be attributed to reformulation alone.

Sachs and Polio (2007) conducted a comparative study to examine the effectiveness of reformulation and traditional error correction on global accuracy. Their research involved a single experimental group who engaged in a three-day sequence of composition, comparison, and revision, three times over a period of three weeks. The first and the last stages of the sequences were the same; however, the comparison stage involved error correction, reformulation, and reformulation + think aloud respectively. The difference between reformulation and error correction was that in the case of reformulation, the texts were rewritten, whereas in the case of error correction, the errors were crossed out and the correct forms were written above them. This study was different from the other studies on reformulation because participants went through all three stages independently without

engaging in peer discussions. Sachs and Polio reported that error correction led to more accurate sentences than reformulation, and reformulation resulted in greater accuracy than reformulation + think aloud. Careful analysis of the exit interviews revealed that error correction was easier for participants to locate their errors compared to reformulation where participants had to try and find the changes themselves, and as a result, they had more free time to memorise the location of the errors and their correct forms and incorporate them later into their final drafts. Thus, they concluded that little understanding took place in the case of error correction.

To compensate for the limitations of the above study, Sachs and Polio (2007) designed a second study which involved three experimental groups with each group receiving a different type of feedback. Moreover, they extended the time period between the comparison and the revision stages. They also included a control group in addition to the three experimental groups. It was expected that these changes would eliminate the effect of memorisation on participants' final texts. Findings show that all three experimental groups were more accurate than the control group on the post-test with the error correction group being superior to the other two groups (reformulation and reformulation + think aloud). An obvious conclusion is that error correction was more effective than reformulation, but Sachs and Polio stated that their results had to be interpreted with care because their study did not examine the long-term effects of these two types of feedback in a new context. Additionally, the advantage for error correction reflected the fact that they analysed language production only in terms of accuracy. However, reformulation can be considered a better method than traditional types of feedback because it does not only deal with surface structure but also addresses cohesive writing. In other words, error correction may increase accuracy of written output, but accuracy is not the only aspect that is important for improving L2 learners' texts.

In another comparative study, Lazaro Ibarrola (2009) analysed the effect of reformulation compared to self-corrections on two low-proficient participants' written performance. Their study was valuable in that in contrast to all previous reformulation studies, it was carried out in a classroom setting making it easier to draw pedagogical implications. The study took place in four sessions with an interval of one week between the sessions. First, the two participants (Bea and Daniel) were asked to perform a writing task individually. In the second session, they were asked to work in pairs to compare Bea's original text with its reformulated version, which was followed by them explaining the errors they had noticed to the teacher. In the next session, they were given Daniel's original text with no corrections and were asked to work together to

find possible errors, explain, and correct them. During the last session, they were given their first drafts and were required to rewrite them independently. Analyses revealed that reformulation was more successful than self-correction because no errors remained unnoticed in the case of reformulation, while self-correction did not encourage noticing of all their errors. The two techniques were similar in that with both, learners correctly incorporated most of the noticed errors in their final draft. The only unfavourable result of this study was that the learners did not seem to notice errors beyond the sentence level (i.e. they paid no attention to structural and textual aspects), which could be attributed to the methodology used in the learners' writing classes.

Lazaro Ibarrola (2013) replicated the above study under the same conditions, using the same tasks. The only difference, however, was the number of participants; that is, this time, the study involved a larger scale (16 participants). As expected, the same results as Lazaro Ibarrola (2009) was reported in this study, too.

In another classroom study (Yang & Zhang, 2010), five groups of participants were first given a picture prompt to write a story in pairs. The second stage involved two sub-stages. First, the five groups worked in pairs to compare their original texts with their reformulated versions. Then they were given a model text (a text which was written by a native-speaker in response to the written task of the study) and were told to compare it with their original texts. Finally, two weeks after the second stage, the participants were post-tested independently by rewriting their original texts and revising them. Yang and Zhang reported that reformulation along with model text comparisons improved the quality of the participants' texts in the post-test. However, the improvements occurred mainly in the areas of lexis and form leaving macro-level aspects of writing, such as cohesion, mostly unaffected. The final stage questionnaires and interviews show that learners paid more attention to micro-level aspects, failing to notice textual reformulations because they believed "good" writing involved good vocabulary and correct forms. In addition, the results of the study demonstrate that model texts were also helpful in improving writing skills because unlike reformulation, they did not need to be faithful to the original text and so could offer a good sample of native writing. Furthermore, as mentioned earlier, they are more applicable than reformulation, especially in EFL settings, where it is very difficult for teachers to reformulate every single student's text.

It is worth mentioning that all of the above studies vary in a number of potentially significant ways, which should be investigated to arrive at robust conclusions about the effect of

reformulation on written performance. One source of variation is the way reformulation is implemented. That is, in some studies (e.g. Qi & Lapkin, 2001; Sachs & Polio, 2007) participants went through all stages (writing, noticing, rewriting) independently, whereas in others (e.g. Lapkin et al., 2002; Swain & Lapkin, 2002) learners had the opportunity to collaboratively write their texts and engage in noticing. Time allocations for noticing (comparing original texts with reformulated versions) were also different among the studies ranging from ten minutes in Swain and Lapkin (2002) to fifty-five minutes in Lazaro Ibarrola (2013). Another variation is in the post-test stage of these studies. That is, most studies (e.g. Swain and Lapkin, 2002) gave students' original texts back to them at the post-test stage and asked the students to revise them, while two studies (Adams, 2003; Sachs & Polio, 2007) required learners to write a whole new text.

Apart from the differences mentioned, some of these studies (e.g. Qi & Lapkin 2001; Swain & Lapkin, 2002) are also limited in that they did not include a control group and were conducted on a small scale. These limitations therefore make it difficult to generalise the results of these studies and call for more research to discover the effects of reformulation on different aspects of written performance.

Overall, reformulation seems to be an effective method of written feedback and has positive effects on lexis and form (e.g. Adams, 2003; Lazaro Ibarrola, 2009, 2013; Qi & Lapkin, 2001) but limited effects on discourse-level and textual aspects (positive effect on discourse level has been reported only in Qi & Lapkin, 2001). It also seems to be more effective for learners of intermediate levels and above (Cohen, 1983).

The current study was designed not only to investigate the effects of task structure, task repetition, and reformulation on the micro aspects of written performance, it also attempts to discover how macro aspects of student texts, namely, text organisation and structure, are affected as a result of performing tasks which are designed to implicitly teach textual structure. Hence, it is worth reviewing the studies which investigated the effect of either implicit or explicit teaching of discourse structure on student writing. To the best of my knowledge, there are only very few studies that have examined this, and all of these studies are within the scope of genre-based teaching of writing. The following section provides a review of these studies.

2.6. Teaching Discourse Structure

Henry and Roseberry (1998) designed a study to determine the effect of genre-based instruction and materials on students' writing and their ability to produce effective tokens of the genre. To this end, they divided 34 participants of the study into two groups: genre and nongenre. The study followed a pre-test – post-test design. Both pre-test and post-test involved students choosing some relevant information from a bank of both relevant and irrelevant information provided to them and using that information to write a tourist information text of 150-200 words. The instruction took six hours over a three-week period. Six authentic tourist information texts were used to develop materials for instruction. In the genre group, students read the texts, identified their moves (genre segments), and determined which moves were obligatory, which were optional and in which order they should appear. The instruction took place in the form of class discussions, and the teacher's role was to facilitate the discussions rather than provide the right answers. Participants in the genre group were made aware of the grammar for each move through consciousness-raising activities. In the nongenre group, the six authentic texts were used to develop a set of traditional materials, such as cloze passages, sentence joining exercises, and error correction exercises. Class discussions of which tenses, adjectives, collocations and so on should or could be used in the different exercises and in students' writing were facilitated to teach grammar. At the end of the instruction, both groups had two opportunities to use the genre: (1) They rewrote a text in which two obligatory moves were missing and which had some grammatical errors. (2) They chose a tourist destination and wrote a text to promote it. Pre- and post-test data were analysed based on motivation index (to what extent the readers were motivated to visit the place described), move score (how well students followed the allowable move structure), and texture index (information on conjunction, conjunctive reach, specificity, connectivity, topic, and topic shift). Findings indicate that the genre group improved significantly from pre- to post-test with regard to motivation score and texture index, and these improvements were significantly different from the nongenre group which made no gains from pre- to post-test. Although no significant results were yielded with regard to move score in the genre group, the level of significance indicates a tendency towards producing texts which followed the move structure.

Bacha (2010) described an explicit instructional approach in teaching argumentative essays in an advanced EAP course for L1 Arab students. The instructional approach involved five steps: (a) building the context, (b) modelling and deconstructing texts, (c) constructing texts jointly, (d) constructing texts independently, and (e) linking related texts. Instruction took place during

a period of four weeks. During the first step, students first wrote a timed argumentative essay on a given topic. Then text types were introduced to the students and the differences among them were discussed. Author's purpose, the intended audience, as well as the way each can affect organisation of the arguments, were also analysed. During the second step, some argumentative essays from the course textbook as well as some student sample essays were analysed with regard to the organisational differences and their effectiveness. During step three, students produced texts jointly. During step four, students produced texts independently, which was followed by class discussions and group work on effectiveness of the organisation. Subsequently, students wrote three timed argumentative essays under test conditions. The topic of the last essay was the same as the one they wrote before the instruction started. During the last step, students were exposed to the similarities and the differences among different text types. Four participants' first and last essays were analysed qualitatively for their organisation. Findings indicate that all four participants made improvements in the manifestation of the argumentative structure in their essays.

In another study which is within the area of both genre-based and task-based teaching of writing, Yasuda (2011) designed a fifteen-week genre-based writing course incorporating email writing tasks to investigate how foreign language writers develop genre awareness, linguistic knowledge, and writing competence. The study involved two intact classes ($N = 70$) of lower-intermediate level English learners who met once a week for ninety minutes. The first session of the course focused on email fundamentals, followed by writing emails to introduce yourself. Sessions two to thirteen each focused on writing a specific type of email, such as writing emails of apology. Session fourteen was a wrap-up session and session fifteen was the final exam. Each lesson followed Norris's (2009) task sequence: task input, pedagogic tasks, target tasks, and task follow up. During the task input phase, the participants were shown two email samples which were written to the same person and for the same purpose but by different people and in different styles. Students were then engaged in pair discussions and later class discussions to analyse the emails based on three metafunction variables: ideational, interpersonal, and textual. The main focus of this phase was to explicitly teach genre-based expressions and help students develop language knowledge as well as genre knowledge. During the pedagogic task phase, students were provided with a real world context and were asked to write two emails to two different people trying to achieve the same functional goal. The student emails were then shared with the whole class to be analysed in terms of language choice. During the target tasks phase, students were asked to write an email in response to a specific

context, reader, and purpose. These tasks were performed outside the class as assignments. The emails were returned to the students along with the instructor's comments and feedback. During the task follow-up phase, students reflected on their performance, their strengths, and their weaknesses according to their instructor's feedback. Data included student surveys, interviews, as well as the two emails students produced at the start and the end of the course. Student emails were analysed in terms of both global and specific measures. Globally, emails were rated based on three analytic criteria: task fulfilment and appropriacy, cohesion and organisation, and grammatical control. Specific measures included measures of fluency (number of words per email), lexical diversity (D index), and lexical sophistication (frequency count of formulaic genre-based expressions). Analyses of qualitative data indicate that learners came to the course with either no or little email writing knowledge, but they perceived that they had made improvements by the end of the course. Quantitative data analyses indicate that learners improved significantly in terms of all global measures, fluency, and lexical sophistication, whereas they made no gains with regard to lexical diversity. Results of this study suggest that explicit teaching of language aspects and discourse structure within a task-based course can help learners improve their texts with regard to both language and structure.

Unlike the above study which focused on email writing, Khodabنده (2014) investigated the effect of genre-based instruction on EFL learners' argumentative writing. The 79 participants of the study were assigned to three groups: explicit, implicit, and self-study. The study took place over two weeks with the explicit and the implicit groups attending eight sessions and the self-study group attending six sessions. All groups were first required to write two argumentative essays (one in L1 – Farsi – and one in English) for pre-testing purposes. Then the participants in the explicit group attended two instructional sessions where they were given two model essays. The researcher then read the model essays to the learners and explained the moves, their functions, and the language used in them. After that, the participants engaged in four practice sessions. During each session, the participants were given a topic to write on collaboratively. The last part of the sessions involved participants choosing a topic and writing individually. The implicit group engaged in the same eight sessions, but they did not receive any instruction when they read the model texts. There were only class discussions and the instructor asked questions to which learners responded orally. The self-study group was given the model texts as homework to read at home, and they only engaged in four practice sessions. Findings indicate that within the groups, the self-study group did not show any improvements

from pre- to post-test. Among the groups, the explicit group outperformed the other two groups in their use of argumentative essay moves (counterarguments and rebuttals).

In summary, findings of the above studies suggest that teaching discourse structure is beneficial for foreign language learners in developing better understanding of how to structure their texts. However, it should be noted that results of the studies by Bacha (2010) and Yasuda (2011) should be interpreted cautiously as they did not include a control group in the design of their studies.

2.7. Research Questions

This chapter provided a detailed review of the robust research in the area of task structure, task repetition, and reformulation as a basis for the development and design of the current study. The following areas which require further research were identified:

- Effects of task structure on written performance,
- Effects of task structure on performance when learners engage in repeated performances of structured tasks,
- Effects of task repetition on written performance,
- Effects of task repetition + intervention on written performance, and
- Effects of task repetition on a new task.

The present study was therefore designed in an attempt to address the above gaps in the existent literature and to answer the following research questions:

1. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of *the same task* in terms of micro-measures?
2. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of *a new task of the same type* in terms of micro-measures?
3. What is the effect of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of *a new task of a different type* in terms of both micro- and macro-level measures?

Chapter 3 Pilot Study

This chapter reports on the methodology and results of the pilot study which was conducted from April to May 2012. The purpose of this small-scale study was to assess the feasibility of the research design, pilot the instruments, including the tasks, the measures and the writing scales for the main study, and identify the potential problems in the methodology. I start by presenting the research questions and then focus on the description of the setting, the design of the study, the participants and the instructional and testing instruments. This is followed by explanation of the instructional and testing procedures, including coding and scoring data, and data analysis methods. Then results are briefly reported. The chapter concludes with an outline of the problems identified in the pilot study and the adjustments that were made for the main study.

3.1 Research Questions

The pilot study aimed to answer the following questions:

1. What effect does repetition of a written structured task have on L2 learners' performance of the same task in terms of micro-measures?
2. What effect does repetition of a written structured task have on L2 learners' performance of a new task of the same type in terms of micro-measures?
3. What effect does repetition of a written structured task + reformulation have on L2 learners' performance of the same task in terms of micro-measures?
4. What effect does repetition of a written structured task + reformulation have on L2 learners' performance of a new task of the same type in terms of micro-measures?
5. What effect does repetition of a written unstructured task have on L2 learners' performance of the same task in terms of micro-measures?
6. What effect does repetition of a written unstructured task have on L2 learners' performance of the new task of the same type in terms of micro-measures?
7. What effect does instruction consisting of task repetition of a structured task have on learners' performance of a new task of a different type in terms of both micro-measures and macro-measures?
8. What effect does repetition of a structured task + reformulation have on learners' performance of a new task of a different type in terms of both micro-measures and macro-measures?

9. What effect does repetition of an unstructured task have on learners' performance of a new task of a different type in terms of both micro-measures and macro-measures?

3.2. Setting

The pilot study was conducted at a small private university in Tabriz, Iran. At the time of data collection, the university was small and newly-established and only offered Bachelor programs in a few disciplines including English Translation. Like most other universities in Iran, students were selected based on their performance on the national entrance exam (Konkour in Farsi). The students who ranked middle to low on the entry exam could choose to study at this university.

The pilot study started in April 2012 when the university re-opened after the Iranian New Year's holidays and it spanned six weeks. Prior to the study, Ethics Approval was obtained from The University of Auckland Human Participants Ethics Committee (Ref. 7911). Consent forms were signed by the participating students, their tutors, and the Head of the English Department (see Appendix A for the Ethics forms).

3.3. Overview of the Design

The study employed a quasi-experimental, pre-writing, repetition, post-writing design with a task repetition study woven into this design (see Table 3.2). Data collection took place over a course of six weeks involving four intact classes (i.e. three experimental groups and a control group). In week 1, all four groups attended a pre-writing session. From week 2 to week 5, the experimental groups engaged in task repetition. Two groups performed the structured tasks (structured group and structured + reformulation group), whereas the other completed the unstructured tasks (unstructured group). In week 3, the structured + reformulation group also had the opportunity to compare their week 2 performance with its reformulated version. In week 4, the experimental groups were required to complete a questionnaire to elicit whether task repetition and reformulation helped them improve their performance and how. In week 6, all four groups attended a post-writing sessions. Two writing tasks (i.e. *Unemployment* task and *Generation Gap* task) were employed at pre- and post-writing sessions and to avoid the practice effect, a counter-balanced design was adopted as shown in Table 3.1.

Table 3.1 *Counterbalanced Sequence of the Free-production Tasks*

	First half of the participants in each group	Second half of the participants in each group
Sequence 1 (pre-writing)	Unemployment task	Generation gap task
Sequence 2 (post-writing)	Generation gap task	Unemployment task

Table 3.2 summarises the design of the study.

3.4. Participants

Participants in the study were forty-seven (mostly female) full-time students studying English Translation in an Iranian private university. They were aged between eighteen and twenty-four years old. Most of them spoke Azeri Turkish as their first language and Farsi as their second language. They had learned English as a second language for four years at high school and two years at junior high school. At the time of data collection, they had sixteen hours of classes every week. Courses were mostly English-related, such as English translation, English writing, and English grammar, but classes were mostly conducted in Farsi. They had no opportunity to use English outside the classroom. Their level of English proficiency was estimated as lower-intermediate.

Participants were in four intact classes which were used to form the four groups of the study. At the start of data collection, there were about thirty students in each class, but due to the high mortality rate, at the end, there were eight students in the structured group, thirteen students in the structured + reformulation group, fifteen students in the unstructured group, and eleven students in the control group.

3.5. Instruments

The study used tasks, reformulated texts, writing scales, and questionnaires as instruments.

3.5.1. Tasks

Two types of tasks were used: free-production tasks for pre- and post-writing and reproduction tasks for task repetition.

Table 3.2 *Design of the Pilot Study*

Group	Pre-writing	Reproduction Task 1 (Time 1)	Reformulation	Reproduction Task 1 (repeat) (Time 2)	Reproduction Task 1 (repeat) (Time 3)	Questionnaire	Reproduction Task 2 (new task of the same type) (Time 4)	Post-writing
Str. N=8	YES	YES	NO	YES	YES	YES	YES	YES
Str. + ref. N=13	YES	YES	YES	YES	YES	YES	YES	YES
Unstr. N=15	YES	YES	NO	YES	YES	YES	YES	YES
Control N=11	YES	NO	NO	NO	NO	NO	NO	YES

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

3.5.1.1. Free-production Tasks

Two free-production writing tasks were used for pre- and post-writing purposes to test participants' writing proficiency both before and after task repetition. These tasks were employed as the aim was to find out to what extent task repetition would facilitate text production as the ultimate goal of any writing instruction. Since the focus of the study was to implicitly teach problem-solution textual structure, the free-production tasks introduced two social problems (i.e. *Unemployment* and *Generation Gap*) which were culturally familiar to the participants. However, the task rubrics did not signal the use of the problem-solution structure in the texts since the objective was to discover whether learners would develop the knowledge of the structure after they engage in performing repeat tasks (see 3.5.1.2). Both tasks required the participants to write at least 250 words.

These free-production tasks were designed based on four defining criteria of pedagogic tasks. That is, they primarily engage learners in expressing meaning; they create a “gap” and require learners to convey information; learners should draw on their own linguistic and non-linguistic resources to complete the task, and finally, they have a clear outcome which is ‘writing an article’ to be published in the local newspaper (Ellis, 2009a).

Unemployment task:

Unemployment is one of the major problems concerning nearly all university graduates in our society. Write an article on this topic to be published in the local newspaper. Write at least 250 words.

Generation Gap task:

The generation gap is one of the major concerns of young adults and parents in our society. Write an article on this topic to be published in the local newspaper. Write at least 250 words.

3.5.1.2. Reproduction Tasks

Two reproduction tasks with a problem-solution textual structure were used as repeat tasks: *Spam* task and *Sea Level* task. According to Hoey (1983), texts with a problem-solution discourse structure follow a situation, problem, solution, and evaluation pattern. By ‘situation’, Hoey means the context and background information; ‘problem’ is “an aspect of situation which requires a response” (Hoey, 1983, p.49); the action to solve the problem is the ‘solution’, and ‘evaluation’ is evaluation of the solution to the problem (see 4.5.1 for full accounts of this pattern). They were texts which participants had to read and reproduce. The texts were

extracted from English text books and were revised by the researcher so that their level of difficulty matched the participants' level of proficiency. To ensure both texts were at the same level of difficulty, their readability was calculated on-line using readability formulas (web-site: <http://www.readabilityformulas.com>). Seven popular readability formulas including the Flesch Reading Ease formula, Flesch-Kincaid Grade Level, Fog Scale (Gunning FOG Formula), SMOG Index, Coleman-Liau Index, Automated Readability Index, and Linsear Write Formula were used to measure the readability of the texts. The *Spam* text was scored 64.6, 8.3, 10.5, 7.6, 9, 8.3, and 8.9 and the *Sea Level* text was scored 64.6, 8.1, 10.3, 7.7, 9, 8.3, and 8.5 on each of the above formulas, respectively. Since the scores were either the same or very close to each other on each formula for both texts, it was confirmed that the texts were at the same level of difficulty. The *Spam* task was used as reproduction task 1 which was repeated 2 times in weeks 2, 3, and 4, whereas the *Sea Level* task was used as reproduction task 2 (new task of the same type) which was performed in week 5.

To operationalise task structure, changes were made to the macrostructure of the problem-solution texts, in that they were organised in such a way as to be unfamiliar to the reader. That is, the unstructured tasks started with a paragraph which evaluated undefined responses to an undefined problem. They then moved on to propose responses to a problem which was not yet defined. The third paragraph defined the problem, and finally the last paragraph focused on the situation which provided the context for the problem.

Like the free-production tasks, the reproduction tasks also meet the criteria for a pedagogic task. That is, they primarily focus on understanding and expressing meaning; they create a “gap” and require learners to convey information; learners need to draw on their own linguistic resources to perform the task, and they have a clear outcome which is ‘writing an article’ for a scientific newspaper (Ellis, 2009a) (see Appendix B for the reproduction tasks).

3.5.2. Reformulated Texts

To operationalise reformulation, a reformulated version of the first performance of reproduction task 1 was used. That is, each participant's first performance of reproduction task 1 in the structured + reformulation group was reformulated by a native-like English teacher who lived in Germany at the time of data collection. The researcher typed the texts and emailed them to the reformulator who reformulated the texts within a week and sent them back to the researcher. The reformulator had been instructed to keep the meaning and content of the texts and only correct mechanical, syntactical, and lexical errors. She was also asked to structure the

texts with regard to paragraph organisation and order following Hoey's (1983) problem-solution textual structure. In other words, the reformulator did not stay loyal to sentence order and replaced the sentences within or between the paragraphs where necessary in a certain text to help improve the structure of the text. The texts and their reformulations were printed out and used in class in week 3 in the structured + reformulation group. That is, participants in this group first compared their original performance to its reformulated version and then performed the same task for the second time.

3.5.3. Questionnaires

Participants in the experimental groups were asked to answer a questionnaire in week 4 immediately after the last repeat performance of reproduction task 1. Two questionnaires were employed: one for the structured and the unstructured groups (repetition only) and the other for the structured + reformulation group (repetition + reformulation). The purpose was to find out how and in what ways repetition and reformulation helped participants to improve their texts, what changes they made to their texts while repeating the task, and what they learned. These questionnaires were translated into Farsi to make it easier for the participants to understand the questions and to avoid any misunderstandings (see Appendix C for the translated versions). The participants were asked to answer them in Farsi, too.

3.5.3.1. Questionnaire for the Structured and the Unstructured Groups

As mentioned, the questionnaire for the structured and the unstructured groups, as shown below, focused only on task repetition and was designed to elicit participants' accounts of whether they benefited from repetition and if yes, how and what changes they made to their texts when they repeated the task for the first and the second time.

Questionnaire

Please read the following questions and answer them.

1. How useful was it for you to repeat the same task for a second time?
(a) very useful (b) somewhat useful (c) a little useful (d) not useful
2. How did repeating the tasks help you improve your writing?
3. Do you think you learned any new language (vocabulary and grammar) or anything about the organisation of writing in English as a result of repeating the tasks? If yes, please say what you learned.

4. What changes did you make when you repeated the tasks?

(a) To the content?

(b) To the language (vocabulary, grammar, spelling)?

(c) To the organisation?

3.5.3.2. Questionnaire for the Structured + Reformulation Group

As shown in the following, the questionnaire for the structured + reformulation group elicited participants' ideas about the effectiveness of both repetition and reformulation.

Questionnaire

Please read the following questions and answer them.

1. How useful was it for you to repeat the same task for a second time?

(a) very useful

(b) somewhat useful

(c) a little useful

(d) not useful

2. How did repeating the tasks help you improve your writing?

3. Do you think you learned any new language (vocabulary and grammar) or anything about the organisation of writing in English as a result of repeating the tasks? If yes, please say what you learned.

4. What changes did you make when you repeated the tasks?

(a) To the content?

(b) To the language (vocabulary, grammar, spelling)?

(c) To the organisation?

5. Did you find reformulation useful?

6. How did reformulation help you improve your writing?

7. Do you think you learned any new language (vocabulary and grammar) or anything about the organisation of writing in English as a result of reformulation? If yes, please say what you learned.

8. What changes did you make after you compared your performance to the reformulated version of it?

(a) To the content?

(b) To the language (vocabulary, grammar, spelling)?

(c) To the organisation?

3.6. Procedures

The following procedures were employed for different groups in the study:

3.6.1. Structured Group

First, the participants in this group were randomly assigned to two sub-groups to perform the pre-writing free production tasks. That is, half of the participants were asked to perform the *Unemployment* task and the other half performed the *Generation Gap* task to avoid a potential practice effect. Each participant was given a piece of paper with the task topic on top and was required to write at least 250 words on that topic. They were given as much time as they needed to write. A week later, participants were asked to complete structured reproduction task 1 for the first time (T1). They were given the texts and 15 minutes to read and take notes. They were not allowed to consult dictionaries. After 15 minutes, both the texts and the notes were collected by the researcher, and the participants were provided with a piece of paper to reproduce the text they had read. They were asked to write at least 250 words. They had as much time as they needed to perform the task and were reminded to proofread their performance. The purpose was to avoid any “slips” and to make sure that the texts represented students’ best performance (Qi & Lapkin, 2001; Yang & Zhang, 2010). The same task was repeated following the same procedures two more times (T2 & T3) with an interval of a week. Immediately after the participants’ third encounter with the same task, they were required to answer a questionnaire (see 3.5.3). Another week after that, the participants completed structured reproduction task 2 (T4), following the same procedures as for structured reproduction task 1. Finally, the group performed the post-writing tasks. Post-writing tasks were the same free production tasks as the pre-writing, and the same procedures as pre-writing were employed for task completion. This time however the participants who performed the *Unemployment* task on the pre-writing, completed the *Generation Gap* task and vice-versa.

3.6.2. Structured + Reformulation Group

The same procedures mentioned above were employed in this group except that this group had the opportunity to compare their first performance of structured reproduction task 1 with its reformulated version a week after they performed the task for the first time. That is, the participants’ first performance of structured reproduction task 1 was reformulated by a native-like English teacher. The reformulator was instructed to maintain the participants’ original meaning, ideas and level of production, but make any necessary lexical, syntactic and discourse

level changes, so that the text reflected native speaker usage. A week after the first performance of structured reproduction task 1, the participants were invited to attend a reformulation session with the researcher. During this session, the participants were given their original texts and the reformulated texts to make comparisons between the two and discover the differences. They were given 30 minutes to do this. After that, the original texts and the reformulated texts were collected, and the participants repeated the same task for the second time following the same procedures that were employed the first time.

3.6.3. Unstructured Group

The same procedures employed in the first group, the structured group, were followed in this group except that this group performed the unstructured tasks instead of the structured tasks.

3.6.4. Control Group

Participants in this group only performed the pre- and the post-writing tasks. The same procedures as for the experimental groups were followed to complete these tasks.

3.7. Coding and Scoring

Data were coded to measure their structural complexity (SC), lexical density (LD), accuracy, fluency (i.e. micro-measures), and textual organisation (i.e. macro-measures).

3.7.1. Micro-measures

To calculate structural complexity, the total number of subordinate clauses was divided by the total number of T-units. Lexical complexity was measured in terms of the percentage of the number of the lexical words per text by total number of words per text (i.e. lexical density). Accuracy was achieved by the ratio of the error-free T-units to total number of T-units. An error-free T-unit was a T-unit with no grammatical and mechanical (i.e. spelling, punctuation, and capitalisation) errors. Fluency was measured by means of the number of words per T-unit.

3.7.2. Macro-measures

Textual organisation was measured by employing two writing scales: holistic scale and problem-solution structure scale.

3.7.2.1. Holistic Writing Scale

The holistic scale as shown in Table 3.3, was used to measure the holistic organisation of the texts. The raters were required to read the participants' texts and decide to what extent, ranging between 0 (the lowest) and 5 (the highest), they were organised.

Table 3.3 *Writing Scale for Measuring Holistic Text Organisation*

The student's text was well-organised and easy to follow.					
Agree					Disagree
5	4	3	2	1	0

3.7.2.2. Problem-solution Structure Writing Scale

The problem-solution structure scale as shown in Table 3.4, measured to what extent the problem-solution discourse structure was manifested in the participants' texts. The scale focused on two criteria: the presence of the problem-solution structure elements (i.e. situation, problem, solution, and evaluation) and the order of the present elements. The highest mark (i.e. 4) was assigned to the texts that included all four elements of the model in the right order, whereas the lowest mark (i.e. 1) was given to the texts that only included one element of the model.

3.8. Data Analysis

Data included pre-writing and post-writing task performances in all four groups, the first and the third performances of reproduction task 1, and the performance of reproduction task 2. That is, the participants in the experimental groups had five pieces of writing while the participants in the control group had only two pieces to be coded.

Data were coded by the researcher, and to ensure inter-rater reliability, 20% of the data were also coded by a second rater. The inter-rater reliability coefficients were above .70 for the measures of lexical density (correlation coefficient = .81) and problem-solution structure rating (correlation coefficient = .74). However, the reliability coefficients for the measures of structural complexity, accuracy, fluency, and holistic rating of the textual organisation did not reach a satisfactory level. These measures therefore were discussed with the second rater until satisfactory levels of inter-rater correlations (lexical density: .81, accuracy = .88, fluency = .98,

and holistic rating: .90) were achieved. Then the scores were fed into the SPSS software Version 21 for further analysis.

First, descriptive statistics were computed for all data sets and all measures mentioned above. Then to answer research questions one, three, and five, in each experimental group, the participants' first and third performances of reproduction task 1 were compared to each other employing either paired-samples t-tests or Wilcoxon Signed Rank tests for each measure.

To answer research questions two, four, and six, in each experimental group, participants' first performance of reproduction task 1 was compared to their performance of reproduction task 2 (new task of the same type) employing either paired-samples t-tests or Wilcoxon Signed Rank tests for each measure.

The last three research questions focused on learners' post-writing performance (new task of a different type). To answer these questions, comparisons were made among the four groups' post-writing performance employing either one-way ANOVA tests or Kruskal-Wallis tests for each measure.

Table 3.4 *Writing Scale for Measuring Problem-solution Textual Organisation*

Problem-solution Structure Manifestation	Score
Performance includes all four elements of a problem-solution structure and they are sequenced in order (situation, problem, solution, evaluation).	4
Performance includes all four elements of a problem-solution structure, but they are not sequenced in order.	3.5
Performance includes just three elements and they are sequenced in order.	3
Performance includes just three elements, but they are not sequenced in order.	2.5
Performance includes only two elements and they are sequenced in order.	2
Performance includes only two elements, but they are not sequenced in order.	1.5
Performance includes just one of the four elements.	1

3.9. Summary of Main Findings

Detailed results of the pilot study are not presented in this section as the purpose was to investigate whether the design and the materials used were satisfactory (see Appendix D for the detailed results). The summary of the main findings however is presented below.

Research question 1:

No significant differences were found between the first and the third performances of reproduction task 1 in the structured group.

Research question 2:

There was a significant increase in fluency of the performance of a new task of the same type (reproduction task 2) compared to the first performance of the reproduction task 1 in the structured group.

Research question 3:

No significant differences were found between the first and the third performances of reproduction task 1 in the structured + reformulation group.

Research question 4:

There was a significant increase in fluency, but a significant decrease in lexical density of the performance of a new task of the same type (reproduction task 2) compared to the first performance of the reproduction task 1 in the structured + reformulation group.

Research question 5:

There was a significant increase in the problem-solution structure manifestation of the third performance of reproduction task 1 compared to its first performance in the unstructured group.

Research question 6:

No significant differences were found between the first performance of the reproduction task 1 and the performance of the reproduction task 2 (new task of the same type) in the unstructured group.

Research questions 7, 8, and 9:

Results of the tests carried out among the four groups indicate no significant differences suggesting that the groups did not differ from each other with regard to the post-writing performance.

3.10. Problems Encountered in the Pilot Study

The pilot study helped identify the problems with the participants, design, and tasks of the study. These problems are addressed below.

3.10.1. Low Proficiency Level of the Participants

The proficiency level of the participants at the university where data were collected was very low (low-intermediate) which was not appropriate for the purposes of this study. As this study focuses on academic writing, participants with an intermediate or upper-intermediate level of English proficiency had to be recruited.

3.10.2. High Mortality Rate

Since the design was semi-longitudinal and data collection took place over five weeks and required every participant to perform all 6 tasks, there was a high mortality rate. I started data collection with about 30 participants in each group, but at the end, there were 8 participants in the structured group, 13 in the structured + reformulation group, 15 in the unstructured group, and 11 in the control group. This in turn resulted in two problems:

1. The counter-balanced design used for pre- and post-writing was very difficult to accomplish. The participants who performed the *Unemployment* task at the pre-writing, were supposed to perform the *Generation Gap* task at post-writing and vice-versa. However, as a result of mortality, in some groups, there was not an even distribution of participants at the post-writing to perform either task. In other words, I had lost the participants who performed, say the *Unemployment* task, at the pre-writing which made it impossible to achieve the counter-balanced design.
2. I lost almost all male participants in the groups.

3.10.3. Participants' Low Level of Motivation

There were only a few participants who were eager to participate in the study and perform the tasks. The others were not motivated and felt bored after the first two performances, and so they refused to continue to participate in the study and if they did, their performance was not reliable.

3.10.4. Operationalising Reformulation in Class

Unfortunately, due to logistic issues and since the lecturers did not give the researcher enough time, reformulation was not operationalised in class as proposed in the initial design of the study. It was proposed that reformulation would be operationalised through the following three steps: (1) a set of guidelines will be produced to explain to students what they need to do when they get back their texts and the reformulated texts. (2) The researcher will go over the guidelines with the whole class providing examples. (3) Students will individually compare their own texts and the reformulated texts. In sum, there was not enough time for the researcher to show the participants how they were supposed to make comparisons between their original texts and the reformulated texts.

3.10.5. Effect of Memory on Reproduction Task Performance

Another major problem concerned the effect of memory on task performance. That is, Repeating the reproduction task 1 three times helped some participants with a good memory, memorise some parts of the texts, and so they just copied the same sentences in their texts instead of producing their own sentences, particularly on the third performance of reproduction task 1. Their performance therefore was not a real reflection of their language ability. In other words, reproduction tasks failed to meet one of Ellis's (2009a) criteria for a pedagogical task (i.e. participants copied the structures from the texts rather than using their own language resources to perform the task).

3.10.6. Holistic Scale of Textual Organisation

The writing scale which was used for holistic rating of textual organisation (see Table 3.3) was too subjective, and therefore it was very difficult to decide on a score based on it.

3.10.7. Reformulating Student Texts

The reformulator, who was a native-like English teacher living in Germany, had complaints about not being able to understand the students' texts, especially when they had produced grammatically incorrect sentences. Hence, in some cases, she just made guesses, and reformulated sentences which resulted in sentences with a different content from the original sentences.

3.10.8. Questionnaire

Analysis of participants' answers to the questionnaires indicates that some of them had not understood what was meant by the organisation of writing. Very few students answered the questions related to the textual organisation, whereas the others either did not answer or pointed to sentence structure instead.

3.11. Proposed Changes for the Main Study

In order to avoid the problems that arose during the pilot study, a few changes needed to be made to the main study. However, before any changes could be proposed, it should be noted that the problems mentioned above are divided into two groups: logistic problems and problems with the design of the study.

It was assumed that the logistic problems (i.e. participants' low proficiency level, participants' low level of motivation, high mortality rate, and operationalising reformulation) stemmed from the nature and conventions of the setting in which the data were collected. In other words, since the university at which the pilot study was conducted was a low-ranking university, the participants were not proficient speakers of English. Moreover, because the researcher was not teaching the classes which participated in the study, students were not very motivated to participate in the study. To solve these problems, data were collected at two private language schools for the main study. In these language schools, it was possible to recruit students at a variety of proficiency levels. Moreover, data were not collected during normal class hours. That is, English writing classes were advertised within both language schools for both intermediate and upper-intermediate level students. The students who were keen were asked to sign up for the classes. The days and times were then announced to the participants based on their availability and room availability within the schools. It was predicted that recruiting participants in this way and running independent classes for the purposes of data collection would solve the first four problems, as students at the right level of proficiency who were also keen and motivated would join the study. It would also allow the researcher enough time to fully operationalise the design of the study.

The other four problems mentioned above involve the design of the study, and thus changes were made to eliminate those problems. Firstly, to overcome the problem of memory effect on task performance, in the main study note-expansion tasks (see 4.5.2.1) were used instead of reproduction tasks. Secondly, to deal with the subjectivity of the holistic rating scale, a more

detailed and objective data-driven scale was developed for the main study. The difficulty of reformulating student texts was expected to be removed automatically if students at the right level of English proficiency were recruited for the main study. That is, it was expected that intermediate and upper-intermediate level students' English would be comprehensible. Finally, to clarify the questions regarding textual structure in the questionnaires, the researcher orally and briefly explained to the participants what text structure was and how they were supposed to answer those questions.

In addition to the changes proposed, changes were also made to the design and the research questions of the study. These changes are summarised below.

3.11.1. Changes to the Design of the Study

Another change that was made to the main study concerns the design. That is, there would be only one repetition of the same task instead of two. It was anticipated that this would remove the boredom that was reported by the participants in the questionnaires and which caused them to refuse to continue to take part in the study. Therefore, the design would be as illustrated in Table 3.5.

Table 3.5 *Design of the Main Study*

Group	Pre-writing	Note-expansion Task 1 (Time 1)	reformulation	Note-expansion Task 1 (repeat) (Time 2)	Questionnaire	Note-expansion Task 2 (new task of the same type) (Time 3)	Post-writing
Str.	YES	YES	NO	YES	YES	YES	YES
Str. + ref.	YES	YES	YES	YES	YES	YES	YES
Unstr.	YES	YES	NO	YES	YES	YES	YES
Control	YES	NO	NO	NO	NO	NO	YES

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

3.11.2. Changes to the Research Questions

Changes were made to the research questions both for reasons of simplicity and to allow more effective ways of analysing the data. Research questions for the main study are as follows.

1. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of *the same task* in terms of micro-measures?
2. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of *a new task of the same type* in terms of micro-measures?
3. What is the effect of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of *a new task of a different type* in terms of both micro- and macro-measures?

Chapter 4 Methodology

This chapter describes the research methodology of the main study. I start with providing an overview of the design. This is followed by a detailed account of the participants and the settings where data were collected. Next, I outline the materials used for data collection and explain the data collection procedures. The chapter concludes with a description of the testing procedures including data coding, scoring, and methods of analysis.

4.1. Overview of the Design

This study aimed to investigate the effects of task structure as a task design variable, task repetition as a task implementation variable, and reformulation as an intervention variable on the written performance of the same task, a new task of the same type, and a new task of a different type (free-writing) in terms of both micro and macro measures.

To investigate the effectiveness of the variables of the study on free-writing (the ultimate goal of any writing instruction), the main study adopted an experimental design: pre-writing, repetition, post-writing. Also woven into this design was a task repetition study to observe the effect of task structure, task repetition, and reformulation on the performance of the same task and a new task of the same type (see Figure 4.1). The entire process of data collection took place over a period of five weeks involving three experimental groups and a control group. In week 1, all four groups attended a pre-writing session. From week 2 to week 4, the experimental groups engaged in task repetition. Two experimental groups performed structured tasks, and the other group completed unstructured tasks. One of the two groups that performed structured tasks was involved in reformulation in week 3. The experimental groups also completed a questionnaire in week 3 to ascertain whether and how task repetition and reformulation helped them improve their performance. In week 5, all four groups attended a post-writing session. Two writing tasks were employed during pre- and post-writing sessions (i.e. *Unemployment* task and *Generation Gap* task) adopting a counter-balanced design (see Table 4.1.) to avoid the practice effect. Figure 4.1 presents the design of the main study.

Table 4.1 *Counter-balanced Sequence of the Pre- and Post-Writing Tasks*

	First half of the participants in each group	Second half of the participants in each group
Sequence 1 (pre-writing)	Unemployment task	Generation gap task
Sequence 2 (post-writing)	Generation gap task	Unemployment task

4.2. Participants

Participants in the study were learners of English as a foreign language at two language schools in Tabriz, Iran. They were 196 (90 of whom quit at different points during the course of the study) English learners at intermediate or upper-intermediate proficiency levels and were aged between 14 and 50. Their native language was Azarbaijani Turkish and they learnt Farsi as a second language because it is the official language in the country. They all had passed at least 2 years of English classes ('Interchange' and 'American files' books) either at the language schools where data were collected or another language school prior to participation in the study. They had also passed some English courses at junior high school, high school, or university. Two participants were also majoring in 'English Language Teaching' at Bachelor's level when the study was carried out. At the time when data were being collected, they were all studying English as a foreign language and had 3 hours and 30 minutes of English classes every week. Participants were recruited from two schools to make sure there were enough participants in each group and mortality would not affect the study. To recruit the participants, intermediate and upper-intermediate level learners in both schools were informed of free writing classes for which they could sign up. In one school, announcements were sent to learners through text messages, and in the other the researcher visited the usual classes in the school and informed students. Fifty-nine participants were recruited from one of the schools and 137 from the other. At the end of data collection process, there were 27 participants in the structured group (7 males, 20 females), 28 in the structured + reformulation group (7 males, 21 females), 27 in the unstructured group (4 males, 23 females), and 24 in the control group (6 males, 18 females).

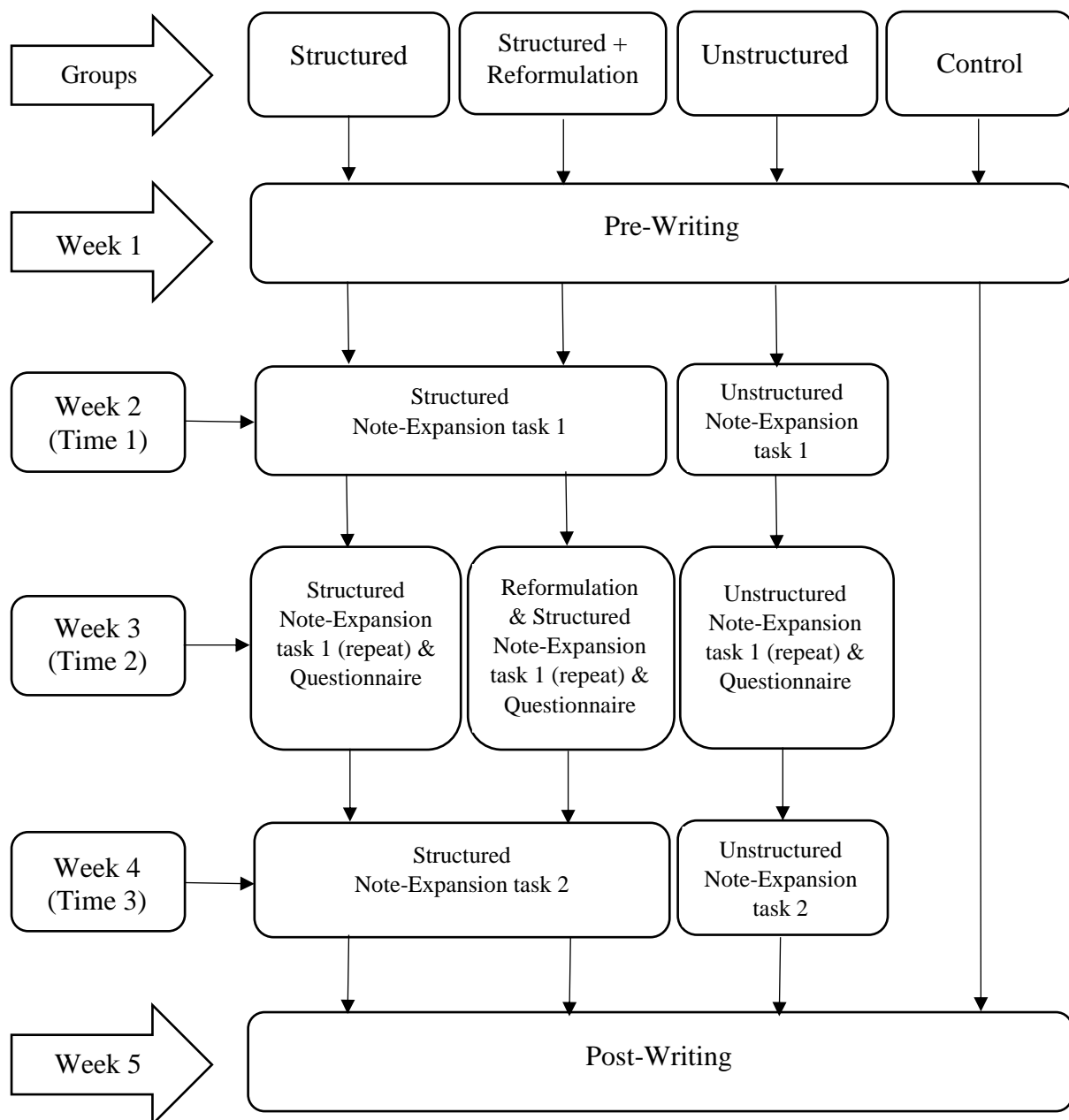


Figure 4.1 *Design of the Main Study*

4.3. Setting

The two language schools in which data were collected are among the leading language schools in Tabriz, Iran and have different branches all over the city. One school was founded in 1993 and the other in 2004. The schools offer classes for both children and adults at various levels ranging from elementary to very advanced (TOEFL or IELTS). Classes are all conducted in English and students are not allowed to speak Farsi. At the time of data collection, *Interchange* textbooks were taught in one of the schools and the other had very recently changed from that

to *American Files*. The schools also used some supplementary materials focusing on vocabulary and grammar. Both schools were managed by their founders and every branch had a separate supervisor who supervised the teachers, their classes, and their teaching through the CCTV cameras the classrooms were equipped with. To recruit English teachers, the schools administer an English proficiency test and the applicants who obtain results that are indicative of their near native proficiency are selected to be interviewed. After the interview, the successful applicants are required to take a short teacher training course to ensure that their teaching style matches the teaching and learning criteria of the schools. At the end of this course, every candidate is assessed by teaching a section of the books that are taught in those schools and any candidate who can demonstrate an acceptable teaching style is appointed.

4.4. Materials for Data Collection

4.4.1. Problem-solution Discourse Structure

Hoey's (1983) problem-solution discourse structure was employed to design the repeat tasks in this study – reproduction tasks in the pilot study and note-expansion tasks in the main study. According to Hoey (1983), texts with a problem-solution pattern follow a Situation – Problem – Response – Result/Evaluation pattern. In this pattern, situation provides some background information and a context for the subsequent sentences (Hoey, 1983, 2001); problem is defined as an “aspect of situation which requires a response” (Hoey, 1983, p. 49); response (in Hoey's terms) is what is done about the problem, and finally, result or evaluation is the consequence and evaluation of the response to the problem. The reason why the term “response” is preferred over “solution” is that for a “response” to be a “solution” to the problem it should be evaluated positively. However, not every response is necessarily evaluated as positive. Authentically, there are sometimes responses that do not work and result in negative outcomes. In these cases, the negative outcome either triggers the original problem or leads to a new problem. In both cases, the problem needs a new response until it is resolved completely. Therefore, a text is not complete unless it ends up in a positive result or a positive evaluation of a response. This leads to the conclusion that since a text with a problem-solution structure must always end in a positive response to be considered complete, it is preferable to use the word “solution” instead of “response”. However, there are texts where it is impossible to retrieve a negative result, and therefore, the negative result “functions exactly like a positive evaluation for the purposes of pattern completion” (Hoey, 2001, p. 132). To clarify this, Hoey (2001) provides an example of a traditional British folk song, ‘The Old Woman Who Swallowed a Fly’, which is a story of an

old woman who swallows a fly and then swallows a spider to eat it. Swallowing a spider is considered to be a response to the problem of swallowing a fly. However, this response not only does not solve the problem, but makes the problem even worse. The old woman then continues to swallow a bigger animal each time to catch the previous one and ends up swallowing a horse and dies. This is a good example of a text where the final negative result is beyond retrieval and acts as a positive response to complete the text.

As mentioned, the situation element of the model sets the background and the context for the subsequent elements. In the following example by Hoey (2001), the first sentence is considered as the “situation” element of the problem-solution pattern which introduces the context and the situation in which the problem exists (lack of ability to write).

(1) I was once a teacher of English Language. (2) One day some students came to me unable to write their names. (3) I taught them text analysis. (4) Now they all write novels (Hoey, 2001, p. 123).

However, it should be noted that it is not this sentence, or in other words, it is not the “situation” element that “triggers any expectations of a pattern to be followed” (Hoey, 2001, p. 123). That is, “situation” in itself is not the trigger of the problem-solution pattern and does not generate any expectation by the reader/listener to read or hear the subsequent sentences. It is only when the problem is signalled (sentence 2) using the word “unable” that the reader or the listener expects a response and the result of the response to the problem (sentences 3 and 4) because without them the text would sound incomplete. Thus, among the elements of the pattern, “situation” is the element which is optional and is not necessary to be included in the text (Hoey, 1983, 2001).

The final point to be considered regarding the problem-solution discourse pattern is the distinction between the “result” and the “evaluation”. According to Hoey (1983), a “result” is a statement of fact which is elicited by the question “what was the result?” while an “evaluation” is a statement of opinion and is elicited by the question “how successful was this?” Consequently, there are three different possibilities according to which a text is organised; first, a positive result and evaluation are provided together in a single part; second, a positive evaluation is provided and then a positive result is stated which serves as a basis for that evaluation, and third, a positive result is first provided and then evaluated. To clarify this, consider the above example. The last sentence of this example (sentence 4) illustrates the first of the three possibilities. That is, it both states the positive end result of the response, which is

teaching text analysis, and evaluates it indirectly. Figure 4.2 summarises the pattern, its elements, and the dynamics among the elements.

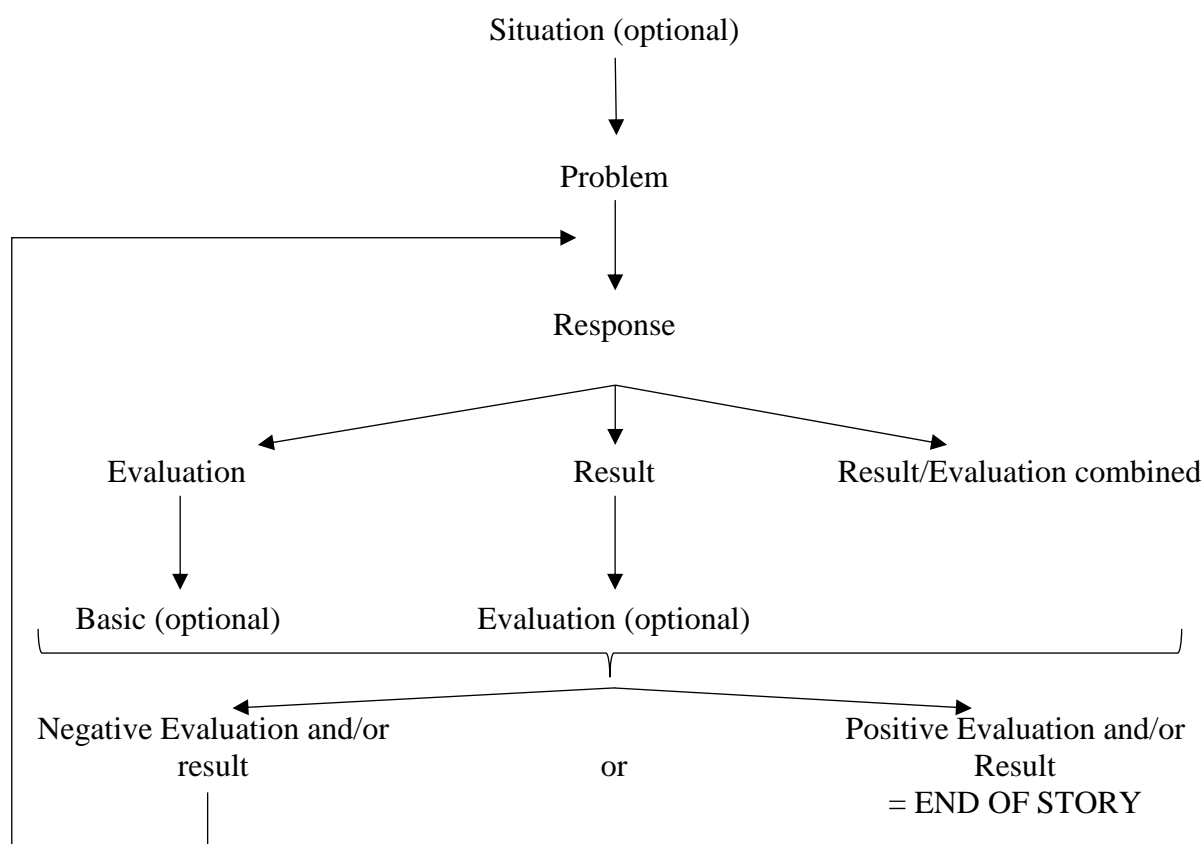


Figure 4.2 *Problem-solution Discourse Structure (Adapted from Hoey, 2001)*

4.4.2. Tasks

Two types of tasks were used in the main study: free-production tasks at pre- and post-writing sessions and note-expansion tasks for task repetition.

4.4.2.1. Free-production Tasks

The same free-production writing tasks employed in the pilot study were used at pre- and post-writing sessions to test participants' writing proficiency both before and after task repetition. The aim was to investigate whether and how task repetition would facilitate text production as the ultimate goal of any writing instruction. As mentioned in chapter 3, since the focus of the study was to implicitly teach problem-solution textual structure, the free-production tasks introduced two culturally familiar social problems (i.e. *Unemployment* and *Generation Gap*) without signalling the use of the problem-solution structure in the texts. The objective was to discover whether learners would develop the knowledge of the structure after they engage in

performing the repeat tasks (see 4.4.2.1). Both tasks required the participants to write at least 250 words.

Unemployment task:

Unemployment is one of the major problems concerning nearly all university graduates in our society. Write an article on this topic to be published in the local newspaper. Write at least 250 words.

Generation Gap task:

The generation gap is one of the major concerns of young adults and parents in our society. Write an article on this topic to be published in the local newspaper. Write at least 250 words.

4.4.2.1. Note-expansion Tasks

Two note-expansion tasks were employed during the task repetition sessions to elicit participants' written performance. Each task had a topic (*Drug Addiction* and *Divorce*) which was a social problem in Iran when data were collected and therefore culturally familiar to the participants. Relevant ideas to each topic were developed in the form of key words and notes. The notes then were organised based on Hoey's (1983) problem-solution discourse structure. That is, tasks started with a paragraph which provided notes related to the situation (context) in Iran (e.g. youth, curiosity, lack of awareness, drugs). Then a problem in that situation was introduced and the notes encouraged participants to focus on the effects of that problem (e.g. drug addiction, effects on society and families). In paragraph three, notes proposed a negatively evaluated response (e.g. one solution, police force, drug dealers [inefficient – financial burden, stop drug users not dealers]), and in paragraph four, a positively evaluated solution was put forward (e.g. another solution, public education [efficient – awareness, drugs, dangers]). Thus, the participants were given the topic, the ideas, and the organisation of the text and were required to expand the notes into full paragraphs and then into a full text. They were required to write at least 250 words.

To operationalise task structure, changes were made to the macrostructure of the tasks in that they were organised in such a way as to be unfamiliar to the reader. That is, the unstructured tasks started with a paragraph which proposed a positively evaluated solution to an undefined problem. They then moved on to propose a negatively evaluated response to the same problem. The third paragraph provided notes which mentioned the problem and encouraged participants

to write about the effects of the problem, and finally, the last paragraph included notes focusing on the situation which provided the context for the problem.

The note-expansion tasks were designed to satisfy the criteria for a pedagogic task. That is, they primarily focused on understanding and expressing meaning; they created a “gap” by requiring participants to convey information; they required learners to draw on their own linguistic and non-linguistic resources to complete the tasks, and finally, they had a clear outcome which was ‘writing an article’ to the local newspaper (Ellis, 2009a).

After one of the tasks (*Drug Addiction*) was designed, it was piloted on 10 learners who were studying English at one of the language schools where data were collected. The purpose was to investigate whether learners would be able to understand and perform this type of task and to establish whether the tasks would elicit texts long enough to be used in the study. Since the results of the trial were satisfying (i.e. learners could easily understand and perform the tasks and they produced texts which were long enough to be analysed), it was confirmed that the tasks would suit the purposes of the study. The *Drug Addiction* task was used as the first repeat task and the *Divorce* task was employed as the second repeat task. These tasks are presented below.

Drug Addiction Task (Structured)

Use the notes below to write an article on 'Drug addiction' to be published in the local newspaper. You will need to expand the notes into full sentences by adding your own words. Write at least 250 words.

First paragraph:

- Youth – curiosity – lack of awareness – drugs

Second paragraph:

- Drug addiction – effects on society and families

Third paragraph:

- One solution – police force – drug dealers
- Inefficient –
 - (a) Financial burden
 - (b) Stop drug users not dealers

Fourth paragraph:

- Another solution – public education
- Efficient –
 - (a) Awareness – drugs – dangers

Drug Addiction Task (Unstructured)

Use the notes below to write an article on 'Drug addiction' to be published in the local newspaper. You will need to expand the notes into full sentences by adding your own words. Write at least 250 words.

First paragraph:

- One solution – public education
- Efficient –
 - (a) Awareness – drugs – dangers

Second paragraph:

- One solution – police force – drug dealers
- Inefficient –
 - (c) Financial burden
 - (d) Stop drug users not dealers

Third paragraph:

- Drug addiction – effects on society and families

Fourth paragraph:

- Youth – curiosity – lack of awareness – drugs

Divorce Task (Structured)

Use the notes below to write an article on 'Divorce' to be published in the local newspaper. You will need to expand the notes into full sentences by adding your own words. Write at least 250 words.

First paragraph:

- Couples – personal differences – cultural differences

Second paragraph:

- Divorce – effects on society, couples, and children

Third paragraph:

- One solution – professional advice before marriage
- Inefficient –
 - (a) General advice for specific cases
 - (b) Unexpected problems after marriage

Fourth paragraph:

- Another solution – mutual discussions
- Efficient –
 - (a) Unimportant issues do not become serious
 - (b) Shared commitment

Divorce Task (Unstructured)

Use the notes below to write an article on Divorce' to be published in the local newspaper. You will need to expand the notes into full sentences by adding your own words. Write at least 250 words.

First paragraph:

- One solution – mutual discussions
- Efficient –
 - (c) Unimportant issues do not become serious
 - (d) Shared commitment

Second paragraph:

- Another solution – professional advice before marriage
- Inefficient –
 - (c) General advice for specific cases
 - (d) Unexpected problems after marriage

Third paragraph:

- Divorce – effects on society, couples, and children

Fourth paragraph:

- Couples – personal differences – cultural differences

4.4.3. Reformulated Texts

In the structured + reformulation group, participants' first performance of the *Drug Addiction* task (i.e. week 2 performance) was reformulated by the researcher and employed in week 3 to provide feedback to the participants in that group (see 4.5.2 for detailed procedures).

4.4.4. A piece of EFL Writing and its Reformulated Version

A piece of EFL writing and its reformulated version (reformulated by the researcher in the same way as the participants' performance was reformulated) were employed in week 3 to teach the participants in the structured + reformulation group as to how they are expected to compare their own performance with its reformulated version (see Appendix E).

4.4.5. Questionnaires

As explained in the pilot study chapter (chapter 3), two questionnaires were employed in week 4 in the experimental groups (one in the structured and the unstructured groups and one in the structured + reformulation group) to elicit participants' accounts of whether and how they benefited from task repetition and reformulation and what changes they made to their

performance of the same task. The same questionnaires used in the pilot study were employed in the main study and they were translated into Farsi to make it easier for the participants to understand the questions and to avoid any misunderstandings (see 3.5.3 for the questionnaires and Appendix C for their translations). The participants were asked to complete the questionnaires in Farsi. Responses to the questionnaires were used to better interpret and discuss the results.

4.5. Procedures for Data Collection

After participants were recruited, they were randomly assigned into four groups (i.e. structured, structured + reformulation, unstructured, and control). Then the following procedures were followed in each group to collect data.

4.5.1. Structured Group

Participants in this group took part in five writing sessions during a course of five weeks. In weeks 1 and 5 they attended pre- and post-writing sessions and performed the free-production tasks. In weeks 2, 3, and 4 (Times 1, 2, and 3) they engaged in task repetition and completed the note-expansion tasks.

Pre-Writing

As in the pilot study, at pre-writing, participants were randomly assigned to two sub-groups: *Generation Gap* group and *Unemployment* group. In other words, half of the participants were required to perform the *Generation Gap* task, and the other half completed the *Unemployment* task. To perform the tasks, each participant was provided with a piece of paper with the task topic on top and was asked to write at least 250 words on the topic. Before they started writing, the researcher briefly explained the topics to them and translated them into Farsi as some of the students did not fully understand the meaning of *Generation Gap*. It was also clarified that they were free to focus on any aspect of the topics. They were not allowed to ask any questions regarding what they should or should not write. The objective was to make sure that participants were not guided directly or indirectly to follow a specific textual structure. They had as much time as they needed to write and were reminded to proofread their production. The purpose was to make sure that the texts represented students' best performance (Qi & Lapkin, 2001; Yang & Zhang, 2010). Drafting was also allowed as there were no time restrictions and participants could follow their own writing styles to complete the tasks. They were also required to write down the times when they started writing and when they stopped. They were

also asked to use pens for writing and cross their mistakes out avoiding correction pens or any kind of erasers. After they finished writing, they submitted their texts to the researcher and were allowed to leave.

Time 1

A week after the pre-writing session, participants met again with the researcher to perform the structured note-expansion task 1 (i.e. *Drug Addiction*). Since the participants had never encountered this type of task, after they were given the task, the researcher explained to them in detail what note-expansion tasks were and how they had to be completed. Participants were told that they needed to use the notes and expand them into paragraphs and write at least 250 words. It was explained that the actual key words and phrases did not necessarily need to be included, but the ideas and the content they carry should be conveyed. Then they were given a few minutes to go over the notes and ask any questions regarding the words and their meanings. Then they started writing. Again, they were given as much time as they needed to complete the task and were allowed to follow their own writing style and draft their texts before finalising them. They were also asked to write down the times when they started writing and when they stopped. They were asked not to use pencils or correction pens or erasers and rather to cross their mistakes out. They submitted their texts to the researcher after they finished writing and were allowed to leave.

Time 2 (The Same Task)

In week 3, participants met again with the researcher to repeat the structured note-expansion task 1 (i.e. *Drug Addiction*). After they were given the task, they were asked to perform the same task for a second time writing at least 250 words. This resulted in confusion as some participants thought they needed to try and remember the exact sentences they had written at Time 1, whereas others assumed they had to write a text avoiding any repetitions of the language or content they had used in their first performance. Therefore, the researcher clarified that the participants did not need to write the same sentences or use the same vocabulary as in their first performance and that they were allowed to use any content or language that they could easily remember. They were given as much time as they needed to perform the task and were allowed to follow their own style of writing and draft their texts. They were also asked to write down the times when they started writing and when they stopped. They were asked not to use pencils, correction pens, or erasers, rather to cross their mistakes out. They submitted their texts to the researcher after they finished writing, but were not allowed to leave. After all

participants submitted their texts, the researcher distributed a questionnaire (see 3.5.3.1) to the participants. The questionnaire required them to answer some questions regarding whether and how task repetition was helpful. To avoid any misunderstandings and to clarify that full and complete answers had to be given to the questions, the researcher explained every question in detail giving examples of how each question, especially the questions concerning textual structure, needed to be answered. After that, the participants were given time to answer the questions and were allowed to leave after they submitted the questionnaire.

Time 3 (New Task of the Same Type)

In week 4, participants met with the researcher to perform the structured note-expansion task 2 (i.e. *Divorce* task). First, the researcher distributed the tasks among the participants. Since the participants were familiar with this type of task, no explanation was given as to how the task should be performed. However, the researcher went over the task briefly explaining the notes and translating the words into Farsi if participants had problems understanding them. Then the participants were given as much time as they needed to perform the task following their own writing style. They were required to write at least 250 words. Like the previous weeks, they were asked to write down the times when they started writing and when they stopped. They were also asked not to use pencils, correction pens, or erasers and rather to cross out their mistakes. They submitted their texts to the researcher after they finished writing and were allowed to leave.

Post-Writing (New Task of a Different Type)

In week 5, participants attended the post-writing session and performed a free-production task. The same procedures as were used at the pre-writing session were followed during this session. The only difference was that to ensure a counter-balanced design, the participants who had completed the *Generation Gap* task at pre-writing performed the *Unemployment* task at post-writing and vice-versa.

4.5.2. Structured + Reformulation Group

The procedures to collect data in this group were exactly the same as the structured group at pre-writing, Time 1, Time 3, and post-writing sessions but different at Time 2.

At Time 2 (week 3), participants in this group had the opportunity to compare their Time 1 performance (i.e. first performance of the *Drug Addiction* task) with its reformulated version; that is, the researcher reformulated participants' Time 1 performance preserving the content,

the original meaning, and the level of proficiency of the texts but correcting all errors regarding grammar, vocabulary, spelling, punctuation, style, and discourse (i.e. connection of ideas or sequence of events). The researcher also changed the order of the sentences when it was necessary to establish a better connection between ideas, therefore enhancing cohesion. Moreover, sentences were occasionally moved between paragraphs to make sure that supporting ideas belonged to the paragraph with the correct topic sentence. This way the organisation of the text was maintained. Changes were made to the texts so that they reflected native-like use of English in written texts. In doing so, the researcher, who is not a native English speaker, consulted different dictionaries and widely searched the internet when she was not sure about certain grammatical or lexical structures (see Appendix F for an example).

When participants met with the researcher at Time 2 (week 3), they were first told that they would be given the opportunity to compare their Time 1 performance with its reformulated version. Then they were given a model text (written by an English learner) and its reformulated version (reformulated by the researcher). The purpose was to teach the participants how to compare their performance with its reformulated version. To do this, the researcher asked the participants to have a look at both texts and started to read one sentence from the original text and the same sentence from the reformulated text. After that, she encouraged the participants to find out the differences between the two sentences (i.e. the original sentence and the reformulated sentence). After the participants pointed out the differences, the researcher asked them to determine the reasons for those changes and differences. The same procedure was repeated for a few more sentences until the researcher was satisfied that the participants understood how to make comparisons between the two texts. The participants' attention was also directed to the sentences that had been replaced among the paragraphs in the reformulated text. This ensured that the students would notice this kind of change when comparing their own texts and the reformulated texts. After that, the researcher collected both texts and distributed the participants' own texts from Time 1 and their reformulated version. The participants were given 15 minutes to make comparisons between the two texts. They were encouraged to discover the differences between the two texts and to think about the reasons for the differences. They had to do this individually and were not allowed to ask questions or seek help. After 15 minutes, the researcher collected both versions of the texts (original and reformulated) and gave each participant a task sheet (i.e. *Drug Addiction* task) to complete. After participants finished writing, they completed the questionnaire for this group (see 3.5.3.2). The same

procedures as in the structured group were followed for both performing the task and completing the questionnaire.

4.5.3. Unstructured Group

Data collection procedures in this group were exactly the same as the structured group during all five sessions (i.e. pre-writing, Time 1, Time 2, Time 3, Post-writing) except that the participants in this group performed the unstructured note-expansion tasks (see 4.4.2.1).

4.5.4. Control Group

Participants in this group only took part in two sessions: pre-writing session in week 1 and post-writing session in week 5. The procedures to collect data were the same as the pre- and post-writing procedures followed in the other groups.

4.6. Data Coding and Scoring

Following the bulk of task-based studies and to allow comparability of the results with previous research findings, the micro-measures employed to analyse data were three measures of L2 proficiency, i.e. complexity, accuracy, and fluency (CAF). Wolfe-Quintero et al. (1998) divide the methods by which CAF is calculated into three types: (a) frequency counts (i.e. researchers count the number of a measurement unit, such as words or T-units), (b) ratio measures (i.e. “the presence of one type of unit is expressed as a percentage of another type of unit, or one type of unit is divided by the total number of comparable units” [ibid, p.10]), and (c) index-based formulas (i.e. “[more complex] formulas that yield a numerical core” [ibid, p.10]). They further assert that among the three methods, ratio measures are more valid.

Pre- and post-writing data were also analysed in terms of macro-measures (i.e. textual structure and organisation). To this end, two data-driven writing scales were developed: (a) *general writing scale* which focused on general organisational features (i.e. idea expression and paragraph division and focus) and (b) *problem-solution structure writing scale* which focused on presence, order, and development of problem-solution structure elements (i.e. situation, problem, solution, and evaluation). The following sections provide a detailed explanation of how data were coded and scored. The procedures for developing the writing scales are also described.

4.6.1. Micro-measures

Complexity

Complexity is defined as “using a wide range of *structures* and *vocabulary* [emphasis added]” (Lennon, 1990, p.390 as cited in Wolfe-Quintero et al., 1998). Therefore, data were coded to measure both their structural complexity and lexical complexity. The structural complexity measure was adopted from previous written task repetition studies (e.g. Larsen-Freeman, 2006; Indrarathne, 2013; Jung, 2013; Pakbaz & Rezai, 2015) and calculated by dividing the total number of clauses by the total number of T-units. A T-unit is defined by Polio (1997) as “an independent clause and all its dependent clauses” (p. 138). A clause is “an overt subject and a finite verb” (ibid, p. 139).

Lexical complexity was measured employing Giraud Index (i.e. the total number of types divided by the square root of the total number of tokens). Giraud Index is a modified variant of the type-token ratio which corrects for the effect of text length (Vermeer, 2000). Types and tokens were counted using KWIC Concordance version 5.0 for Windows (available at <http://kwic-concordance.software.informer.com/5.0/>). Texts were first typed in Microsoft Word Document and all spelling errors were corrected. Spelling errors were corrected, so that the same word which is spelt differently on two occasions in a single text is not counted as two types (e.g. addiction and adiction). All two-word lexical items were typed with a space as in some cases it was difficult to detect whether there was originally a space between them or not. Hyphenated words were typed as they were written even if they had been hyphenated by mistake (e.g. drug-addiction). False starts and repetitions were excluded. Then the software was run for each text independently to count the number of the types and tokens. Two-word lexical items (e.g. each other) were counted as two words and hyphenated words were counted as one.

Accuracy

As defined by Skehan (1996a), accuracy refers to “how well the target language is produced in relation to the rule system of the target language” (p. 23). Accuracy was measured through mechanical accuracy and grammatical accuracy. Mechanical accuracy was measured in terms of the number of orthographically error-free T-units to total number of T-units and was judged based on spelling, punctuation, and capitalisation. In other words, a T-unit was orthographically error-free if it was correctly capitalised, had no spelling errors and was

punctuated correctly. Grammatical accuracy was measured in terms of the number of grammatically and lexically error-free clauses to total number of clauses. This measure of accuracy has been widely used in the literature on task-based research (e.g. Jung, 2013; Pakbaz & Rezai, 2015) and considered by Wolfe-Quintero et al. (1998) to be the best accuracy measure of written performance, especially in long term studies. Orthographical errors were treated separately since according to Polio's (1997) guidelines, spelling and capitalisation errors should not be counted as accuracy errors. T-units were chosen over clauses for measuring mechanical accuracy as T-units are better suited for considering punctuation and capitalisation errors, as capitalisation mostly takes place at the beginning of T-units and punctuation, commas in particular, is considered across clauses and within T-units.

Fluency

Fluency refers to "how comfortable the second language writer is with producing language" (Wolfe-Quintero et al., 1998, p.13). Abdel Latif (2014) divides written fluency measures into product- and process-based measures and questions the validity of the product-based measures that he believes have been derived from oral fluency measures. He further argues that process-based measures, such as the mean length of translating episodes, are the best measures to calculate written fluency. However, in the absence of the data required for process-based measures, this study adopted product-based measures to calculate fluency. Also, since nearly all research in the past has used product-based fluency measures, adopting these measures allows comparability of the results with the findings of the other studies.

As Indrarathne (2013) asserts, written language fluency is defined differently from oral language fluency. Larsen-Freeman (as stated in Wolfe-Quintero et al., 1998, p.14) states that a fluent writer is able to write longer texts confirming that length of the text can be used to determine fluency. Wolfe-Quintero et al. (1998) argue that fluency in writing can be measured by calculating the length of a measurement tool (e.g. Sentence or T-unit) or in other words, by counting the number of the words in a measurement tool, and suggest the 'number of words per T-unit' to be the best measure of written language fluency based on T-unit. Therefore, the number of words per T-unit was used in this study based on Wolfe-Quintero et al.'s suggestion and following Larsen-freeman (2006) and Indrarathne (2013).

Measure of number of words per minute had also been proposed to be used in this study prior to data collection since Wolfe-Quintero et al. (1998) argue that calculating the rate of production is another way of measuring written fluency and because Indrarathne (2013)

considers time factor to have an important role in length measurements. To this end, participants were required to write down the times when they started writing and when they stopped (task performances were not timed in this study). However, when data were being collected, it was observed that participants used the time they spent for writing in different ways. That is, some participants used the time to write several drafts before they finalised their texts, whereas others just wrote a single draft during that time. Therefore, this measure could not be operationalised.

Another measure of fluency which had been proposed was the ratio of disfluencies (Ellis & Yuan, 2005). Therefore, participants were asked to write their texts with pens and cross out their mistakes out. However, some participants ignored this and used pencils and erasers making it impossible to operationalise this measure. Hence, to compensate for this, another length measure (i.e. number of words per text) was employed to measure fluency.

A second rater – a native-like English teacher – coded 20 percent of the data for all measures excluding one of the fluency measures (number of words per text) and lexical complexity, which were calculated by the KWIC Concordance software. Since the inter-rater correlation coefficients did not reach a satisfactory level, the researcher and the second rater met to discuss the differences. Once the areas of discrepancies were resolved, the researcher checked all data one more time to make changes to the codings where necessary until a correlation coefficient of greater than 90% was achieved for all four measures.

4.6.2. Macro-measures

As stated earlier, two data-driven writing scales (i.e. general writing scale and problem-solution structure writing scale) were developed to measure macro-aspects of the written texts. A detailed explanation of this process is provided below.

Developing the Writing Scales

To start with, 20 pre-writing texts were randomly selected and analysed to discover their problems with regard to textual structure and organisation. The purpose was to find out what organisational problems the pre-writing texts had and to design a scale to investigate whether any improvements took place in post-writing texts. One problematic area was that some texts included ideas which were irrelevant to the task topic; that is, participants had digressed and diverged from task topic. Second, there were issues related to paragraph development. In other words, participants had either failed to divide their texts into paragraphs or they had written

paragraphs with more than one main idea and topic. Another problem was that the ideas had not been fully developed within the paragraphs; that is, not enough support had been provided for the main ideas in the paragraphs. It was also evident that not all of the problem-solution structure elements (i.e. situation, problem, solution, and evaluation) were included in the texts. As a result, the scale below was developed (Writing Scale stage 1).

To check the reliability of this scale, 10 texts were randomly selected and rated by two raters (the researcher and a native-like English teacher). The second rater received online (through e-mails) training and explanations with regard to what the ‘problem-solution’ discourse structure was and how the scale was supposed to be used. Since the correlation coefficients did not reach a satisfactory level, the two raters met to discuss the differences and determine the areas of discrepancies. The discussions revealed that the second rater had misinterpreted some of the rating criteria and therefore the criteria needed to be more fully explained. Moreover, the raters agreed that there was an overlap between criteria 3 and 4. That is, if the topic sentence was strongly supported throughout a paragraph, inevitably that support was achieved through logically connected ideas. In addition, it was discovered that rating would be more straight forward if scoring started from 4 instead of 5 since most of the texts had four paragraphs and also the ‘problem-solution’ discourse structure had four elements. The scale was then modified based on the raters’ discussions and sent again to the second rater who rated the same texts for a second time based on the new scale. The modified scale is presented below (Writing Scale stage 2).

Writing Scale (stage 1)

1. Ideas expressed through the text are related to the topic.

5 4 3 2 1

2. The text is appropriately divided into different paragraphs.

5 4 3 2 1

3. Every topic sentence is strongly supported throughout the paragraph.

5 4 3 2 1

4. Ideas are logically connected to each other in each paragraph.

5 4 3 2 1

5. Ideas are arranged to support the problem-solution discourse structure.

5 4 3 2 1

6. Paragraphs are ordered based on the problem-solution discourse structure.

5 4 3 2 1

In spite of the modifications made to the scale and the explanations given to the second rater, the correlation coefficients were still not satisfactory. Therefore, the two raters met again to discuss the discrepancies. As a result of this discussion, it was revealed that further explanations were required regarding the criteria to make sure that the texts were rated according to the criteria rather than the rater's own interpretations and experience of rating English texts. Hence, some explanations and guidelines were added to the scale, so that the criteria could easily be interpreted and the scale could be employed by any rater, not just the researcher. Moreover, the scale was split into two different scales; one measured the degree to which the text was appropriately organised when some general criteria were considered (General Writing Scale), and the other measured the degree to which the text reflected a 'problem-solution' textual structure (Problem-solution Structure Writing Scale). Consequently, the two scales, presented in Tables 4.2, 4.3, 4.4, and 4.5, were developed.

Writing Scale (Stage 2)

1. Ideas expressed through the text are related to the topic.

4 3 2 1

2. The text is appropriately divided into different paragraphs.

4 3 2 1

3. Every topic sentence is strongly supported in the paragraph through the expression of logically connected ideas.

4 3 2 1

4. Ideas are arranged to support the problem-solution structure.

4 3 2 1

5. Paragraphs are ordered based on the problem-solution structure.

4 3 2 1

Table 4.2 *General Writing Scale (Stage 3)*

Score	Idea Expression	Paragraph Division and Focus
-------	-----------------	------------------------------

4	All of the ideas expressed throughout the text are related to the topic.	The text is appropriately divided into different paragraphs. Each paragraph has a single focus.
3	Most of the ideas expressed throughout the text are related to the topic.	The text is appropriately divided into different paragraphs. At most, only one paragraph has more than one focus.
2	Some of the ideas expressed throughout the text are related to the topic.	The text is appropriately divided into different paragraphs. Most paragraphs have more than one focus.
1	The ideas expressed throughout the text are not related to the topic.	Either the text is not divided into paragraphs or it is, but all paragraphs have more than one focus.

Table 4.3 *Guidelines for Employing the General Writing Scale (Stage 3)*

Idea Expression	Paragraph Division and Focus
<p>Go through the text. Find the ideas that are expressed by the writer and check whether they are relevant to the topic of the text. On some occasions, the writer may digress and diverge and focus on different issues forgetting the main topic of the text. For example, a participant concluded her text on <i>Generation Gap</i> writing “In conclusion people must follow their favourites and interests because if we accept whatever other people want in this way we can’t be successful.”</p>	<p>First, consider paragraph division. That is, find out if it is clear where each paragraph starts and ends or the writer mainly wrote a single long paragraph without dividing it into short and clear paragraphs. If paragraph division is achieved, then check if each paragraph focuses on a single idea and the explanation of it or it includes several ideas.</p>

Table 4.4 *Problem-solution Structure Writing Scale (Stage 3)*

Score	Presence of the Elements of the Problem-solution Structure	Order of the Elements of the Problem-solution Structure	Development of the Elements of the Problem-solution Structure
4	All four elements of the problem-solution structure (situation, problem, solution, and evaluation) are present in the text	The four elements are clearly ordered according to the problem-solution structure	All of the elements are developed and expanded with appropriate details or examples
3	Only three elements of the problem-solution structure are present in the text	Three elements are ordered according to the problem-solution structure	Three of the elements are developed with appropriate details or examples
2	Only two elements of the problem-solution structure are present in the text	Two elements are ordered according to the problem-solution structure	Two of the elements are developed with appropriate details or examples
1	Only one or none of the elements of the problem-solution structure are present in the text	There is either one element present in the text or none of the existing elements are ordered according to the problem-solution structure	Only one or none of the elements are developed with appropriate details or examples

Table 4.5 *Guidelines for Employing the Problem-solution Structure Writing Scale (Stage 3)*

Presence of the Elements of the Problem-solution Structure

Go through the text and mark whenever you see one of the elements of the problem-solution structure (i.e. situation, problem, solution, and evaluation). Finally, assign a score based on the number of the elements you marked. That is, if you marked only a problem and a solution to it, score 2. Note, ignore the order in which the elements are organised.

Order of the Elements of the Problem-solution Structure

Go through the text and mark whenever you see one of the elements of the problem-solution structure. Then among the existing elements, consider the number of the elements that are in the right order (situation, problem, solution, and evaluation), and score the text based on that ignoring the number of the elements the text includes. For example,

- If the text includes four elements which are ordered as problem, situation, solution and evaluation, score 2 as only the ‘solution’ and the ‘evaluation’ elements are ordered correctly.
- If the text includes three elements ordered as problem, solution 1, evaluation 1, solution 2, and evaluation 2, score 3.
- If the text includes three elements ordered as situation, solution and evaluation, score 3 as all of the existing elements are ordered correctly.
- If the text includes two elements ordered as problem and solution, score 2.
- If the text includes three elements ordered as evaluation, solution and problem, score 1 as none of the elements are in the right order, and etc.

Development of the Elements of the Problem-solution Structure

Go through the text and mark whenever you see one of the elements of the problem-solution structure. Then take each element in isolation and see how it is explained and developed. In other words, if there are ample explanations or maybe examples through which the element is explicated and supported, consider it as a fully developed element. However, if the element is mainly expressed through a single sentence or very brief explanations are given making it very difficult to understand, consider it as an inadequately developed element.

After the above two scales were developed, another 10 pre-writing texts were randomly selected and rated by the researcher and the second rater. This time, the second rater’s scores correlated highly with the researcher’s scores for most of the criteria. Nonetheless,

discrepancies existed in ‘paragraph division and focus’ in the ‘General Writing Scale’ and ‘order of the elements’ in the ‘Problem-solution Structure Writing Scale’ and needed to be discussed. After the researcher met with the second rater for the third time, it transpired that the differences in the ratings of ‘paragraph division and focus’ stemmed from the fact that the second rater, based on his own experience of rating texts, took the number of the paragraphs into account and tended to give low marks to the texts that had fewer than four paragraphs. To avoid this misunderstanding, ‘the number of the paragraphs should be ignored while rating the texts on this criterion’ was added to the guidelines. One of the other problematic issues about ‘paragraph division and focus’ that caused discrepancies in the ratings was that in some texts participants had written about a single main idea in more than one paragraph. Therefore, it was also explained in the guidelines for the scale that ‘if a single main idea is explained in more than one paragraph, those paragraphs should be considered well-structured and receive a full score. However, if a main idea is discussed in a paragraph and then continued to a second paragraph which also focuses on another main idea, the first paragraph should receive a full score. The second paragraph, in contrast, receives no score as it focuses on two topics’.

Regarding the ‘order of the elements’, the discussions revealed that it was very difficult to draw a border between the elements of ‘situation’ and ‘problem’ in the texts. Moreover, it was difficult to rate the texts on this criterion in some cases. For example, when a text was ordered as ‘solution, evaluation, and problem’, it could be rated 2 as the ‘solution’ and the ‘evaluation’ were in order, but the ‘problem’ element at the end made the structure effectively disintegrate because it seemed that there was no order in the text. To solve these problems, the two criteria of ‘presence of the elements’ and ‘order of the elements’ were combined. A few other small adjustments were also made to the scales and the guidelines after the researcher and the second rater started to rate and score the texts; that is, since two texts were too short (i.e. not enough ideas had been provided) to be evaluated on ‘idea expression’, it was decided that the texts which were too short are scored 1 for ‘idea expression’. In addition, it was observed that in a few cases ideas provided in the texts were not clear enough, making the whole text or parts of it incomprehensible. Therefore, it was agreed that the ideas should not only be relevant, but also clear and concise and this was also added to the scale for ‘idea expression’. Finally, half scores were allowed in rating ‘idea expression’ and ‘development of the problem-solution structure elements’. The final scales are presented below (Tables 4.6, 4.7, 4.8, and 4.9).

Table 4.6 *General Writing Scale (Final)*

Score	Idea Expression	Paragraph Division and Focus
4	All of the ideas expressed throughout the text are related to the topic; clear, and concise.	The text is appropriately divided into different paragraphs. Each paragraph has a single focus.
3	Most of the ideas expressed throughout the text are related to the topic; clear, and concise.	The text is appropriately divided into different paragraphs. At most, only one paragraph has more than one focus.
2	Some of the ideas expressed throughout the text are related to the topic; clear, and concise.	The text is appropriately divided into different paragraphs. Most paragraphs have more than one focus.
1	The ideas expressed throughout the text are neither related to the topic nor concise, or the text is too short.	Either the text is not divided into paragraphs or it is, but all paragraphs have more than one focus.

Table 4.7 *Guidelines for Employing the General Writing Scale (Final)*

Idea Expression	Paragraph Division and Focus
<p>Go through the text. Find the ideas that are expressed by the writer and check if they are relevant to the topic of the text and comprehensible. Be careful of the instances where the writer digresses and focuses on different issues other than the main topic of the text. Please note that you can assign a half score if you believe the text is somewhere between two full scores. If the ideas expressed are related to the topic, but the text is too short, score it 1.</p>	<p>First, consider paragraph division. That is, find out if it is clear where each paragraph starts and ends or if the writer mainly wrote a single long paragraph without dividing it into short and clear paragraphs. If paragraph division is achieved, then check if each paragraph focuses on a single idea and the explanation of it or it includes several ideas without any supporting information. Moreover, if a single main idea is explained in more than one paragraph, those paragraphs should be considered well-structured and receive a full score. However, if a main idea is discussed in a paragraph and then continued to a second paragraph which also focuses on another main idea, the first paragraph should receive a full score. The second paragraph, in contrast, receives no score as it focuses on two topics</p>

Table 4.8 *Problem-solution Structure Writing Scale (Final)*

Score	Presence and Order of the Problem-solution Structure Elements	Score	Development of the Problem-solution Structure Elements
4	All four elements of the problem-solution structure (situation, problem, solution, and evaluation) are present in the text and they are ordered according to the problem-solution structure	4	All of the elements are developed and expanded with appropriate details or examples
3.5	All four elements of the problem-solution structure are present in the text, but they are not ordered according to the problem-solution structure		
3	Only three elements of the problem-solution structure are present in the text and they are ordered according to the problem-solution structure	3	Three of the elements are developed with appropriate details or examples
2.5	Three elements of the problem-solution structure are present in the text, but they are not ordered according to the problem-solution structure		
2	Only two elements of the problem-solution structure are present in the text and they are ordered according to the problem-solution structure	2	Two of the elements are developed with appropriate details or examples
1.5	Two elements of the problem-solution structure are present in the text, but they are not ordered according to the problem-solution structure		
1	Only one or none of the elements of the problem-solution structure are present in the text	1	Only one or none of the elements are developed with appropriate details or examples

Table 4.9 *Guidelines for Employing the Problem-solution Structure Writing Scale (Final)*

Presence and order of the Elements of the Problem-solution Structure	Development of the Elements of the Problem-solution Structure
Go through the text and mark whenever you see one of the elements of the problem-solution structure (i.e. situation, problem, solution, and evaluation). Score the text based on the number of the elements you marked. That is, if you marked only a problem and a solution, score 2. Remember this is not the final score. After that, find out whether or not the existing elements are in the right order according to the problem-solution structure. In other words, if the problem is followed by the solution, score 2, but if the solution precedes the problem, score 1.5.	Go through the text and mark whenever you see one of the elements of the problem-solution structure. Then take each element in isolation and see how it is explained and developed. In other words, if there is ample explanation or examples through which the element is explicated, consider it as a fully developed element. However, if the element is mainly expressed through a single sentence or a very brief explanation making it very difficult to understand, consider it to be an inadequately developed element. If you believe an element is neither fully nor briefly developed, you can assign a half score to it.

After the scales were developed, pre- and post-writing task performances were rated and scored according to both scales and by two raters – the researcher and a native English teacher and writer (not the same rater who contributed to the development of the scales). The second rater was given the scales to study and received some instruction with regard to problem-solution discourse structure. The researcher and the rater then met and rated a few texts together, so that any potential confusing areas were detected and discussed. When the rater was ready, the researcher and the second rater met on a daily basis for about two months and rated the texts together. This resulted in a more efficient rating since when they did not agree on a score for a criterion, they discussed it right away before they forgot why they had assigned a certain score. To assign scores, when the researcher's and the rater's scores were different only by one score (e.g. one assigned 3 and the other 4) the mean of the two scores were assigned (e.g. 3.5).

However, when the difference between the raters' scores was 2 or larger, the raters conferred until they reached agreement.

4.7. Statistical Analysis

Scores were fed into Statistical Package for Social Sciences (SPSS) software version 22 to perform the statistical analyses at the 95% confidence interval level. First, before the research questions were investigated, the homogeneity of the four groups were tested to specify whether the groups were significantly different from each other at the beginning of the study in terms of both micro- and macro-measures. To this end, the pre-test scores were compared across the four groups. Results of this investigation revealed that the groups were homogeneous in terms of all micro-measures and macro-measures excluding the 'presence and order of the problem-solution structure elements' (see Appendix G for detailed results). These results confirmed that the groups were highly homogenous and any significant results for research questions could be attributed to the variables of the study. Following this, each research question was investigated employing the tests and the procedures provided below.

Research Question 1: What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of the same task?

To answer research question 1, first, the gain/loss scores from Time 1 (T1) to Time 2 (T2) were computed for each micro-measure. Absolute z-scores associated with the gain/loss scores were calculated and any scores associated with z-scores larger than 2.5 were removed as outliers (Field, 2009). Scores were then checked for distribution and homogeneity employing Kolmogorov-Smirnov and Levene tests. When the results of these tests confirmed that the scores were both normally distributed (Kolmogorov-Smirnov $p > .05$), and homogeneous (Levene $p > .05$), a parametric one-way between groups ANOVA test was employed to compare the gain/loss scores across the three experimental groups (i.e. structured, structured + reformulation, and unstructured). However, when the assumptions of normality and homogeneity were not satisfied, a non-parametric Kruskal-Wallis H test was carried out. When the results of the ANOVA or the Kruskal-Wallis H test reached significance ($p < .05$), post-hoc analyses were performed. Bonferroni tests were carried out for post-hoc analyses of ANOVA test results. For post-hoc analyses of Kruskal-Wallis test results, further Kruskal-Wallis tests were carried out. Effect sizes were calculated for both significant and non-significant results using an online calculator (<http://www.uccs.edu/~lbecker/>). A d value of .40

to .69 was considered as a small effect size, .70 to .99 was considered as a medium effect size, and 1 or above was considered as a large effect size for between-groups comparisons. A d value of .60 to .99 was considered as a small effect size, 1 to 1.39 was considered as a medium effect size, and 1.40 or above was considered as a large effect size for within-groups comparisons (Plonsky and Oswald, 2014). Plonsky and Oswald's benchmarks were adopted as they are discipline-specific and allow better interpretation of d in second language research.

Research Question 2: What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of a new task of the same type?

To answer research question 2, the same statistical tests and procedures as above were adopted. The only difference was that the gain/loss scores were calculated from Time 1 to Time 3.

Research Question 3: What is the effect of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of a new task of a different type in terms of both micro- and macro-level measures?

To answer research question 3, both within-group and between-group analyses were performed. Within-group differences (i.e. time differences) from pre- to post-writing were investigated for the four groups of the study separately. Analyses included examining both micro- and macro- measures. After outliers were removed, scores were checked for their normality employing a Kolmogorov-Smirnov test. When scores were normally distributed, parametric paired-samples t-tests were carried out between pre- and post-writing performances. However, when the assumption of normality was not satisfied, non-parametric Wilcoxon Signed Ranks tests were performed. Effect sizes were calculated for both significant and non-significant results. To carry out between-group differences, the statistical tests and procedures employed for research question 1 were adopted. There were, however, three differences: (1) Gain/loss scores were computed from pre- to post-writing. (2) Analyses included investigating both micro- and macro-measures. (3) Group comparisons included the control group, too.

Chapter 5 Results and Discussion 1

This chapter addresses the first and the second research questions, which investigate the relative effects of task structure, task repetition, and reformulation on the repeat performance of the same task and a new task of the same type. Thus, the chapter is presented in three sections. The first section presents the results of the first research question: What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of the same task in terms of micro-measures? The second section focuses on the results of the second research question: What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of a new task of the same type in terms of micro-measures? And finally, the last section provides a discussion of the results of the two research questions.

5.1. Statistical Analysis

To answer the research questions, data were first coded by the researcher in order to calculate the structural and lexical complexity, grammatical and mechanical accuracy, and fluency measures. To calculate structural complexity, the total number of clauses was divided by the total number of T-units. Lexical complexity was measured by employing the Giraud Index; that is, the total number of types divided by the square root of the total number of tokens. Types and tokens were counted using KWIC Concordance version 5.0 for Windows (available at <http://kwic-concordance.software.informer.com/5.0/>). Texts were first typed in Microsoft Word Document. Then the software was run for each text independently to count the number of types and tokens. Grammatical accuracy was measured in terms of the number of error-free clauses to total number of clauses. Error-free clauses were the clauses that contained no grammatical and lexical errors. Mechanical accuracy was measured in terms of the number of orthographically error-free T-units to total number of T-units and was judged based on spelling, punctuation, and capitalisation. Fluency was measured both in terms of the total number of words per text and the number of words per T-unit. Then 20 percent of the data were also coded by a second rater – a native-like English teacher – to check for inter-rater reliability of all measures excluding fluency 1 (total number of words per text) and lexical complexity, which were calculated by the KWIC Concordance software. Since the inter-rater correlation coefficients did not reach a satisfactory level, the researcher and the second rater met to discuss

the differences. Once the areas of discrepancies were resolved, the researcher checked all data one more time to make changes to the codings where necessary until a correlation coefficient of greater than 90% was achieved for all four measures. After the inter-rater reliability check, scores were fed into SPSS (version 22) to carry out the statistical tests at the 95% confidence interval level across the three experimental groups – structured (Str.), structured + reformulation (Str. + ref.), and unstructured (Unstr.) – and between Time 1 (T1), Time 2 (T2), and Time 3 (T3) performances. To perform statistical analyses, first, the gain/loss scores from T1 to T2 and T1 to T3 were computed for each measure. Then the gain/loss scores were compared across the three groups employing either one-way between groups ANOVA or Kruskal-Wallis H tests. Effects sizes were also calculated for both significant and non-significant results and were interpreted adopting Plonsky and Oswald's (2014) benchmarks for between-group comparisons. Results of the statistical tests as well as the effect sizes are reported in the following sections.

5.2. Results for Research Question One

This section presents the results of the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of the same task in terms of micro-measures.

5.2.1. Structural Complexity

After the structural complexity gain/loss scores from T1 to T2 were computed in all groups, the absolute z-scores associated with the gain/loss scores were also calculated and any scores associated with the z-scores larger than 2.5 were removed as outliers (Field, 2009). The resulting number of participants was 26 in the structured group, 27 in the structured + reformulation group, and 25 in the unstructured group. Descriptive statistical tests were then carried out for all groups. Results are shown in Table 5.1 (see Appendix H for T1 and T2 descriptive statistical results).

Table 5.1 *Descriptive Statistics Results for T1 – T2 Structural Complexity Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	26	.04	.28	.52	-.46
Str. + ref.	27	-.06	.24	.61	-.60
Unstr.	25	.07	.19	.61	-.32

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Results of descriptive statistical tests revealed that both the structured and the unstructured groups improved in structural complexity from T1 to T2. However, this improvement was small in both groups (Str. $M = .04$, and Unstr. $M = .07$). The structured + reformulation group declined slightly from T1 to T2 ($M = -.06$).

The mean differences were examined among the three groups performing further statistical analyses. Hence, the scores in all groups were first checked for distribution and homogeneity employing Kolmogorov-Smirnov and Levene tests. Since the results of these tests confirmed that the scores were both normally distributed (Kolmogorov-Smirnov $p > .05$) and homogeneous (Levene $p > .05$) in all groups, a one-way between groups ANOVA test was carried out to find out whether or not the mean differences among the groups were statistically significant.

Results of the ANOVA test indicate that the differences among the three groups in terms of the gain/loss in structural complexity from T1 to T2 did not reach significance; $F(2, 76) = 2.32$, $p > .05$.

Although no significant results were found, effect sizes were calculated for group comparisons for better understanding of the effect of the variables on the repeat performance of the same task. The resulting d values were .38 for the structured *vs.* the structured + reformulation group comparison, .12 for the structured *vs.* the unstructured group comparison, and .60 for the structured + reformulation *vs.* the unstructured group comparison. These d values indicate no or a small effect for the variables.

5.2.2. Lexical Complexity

After computing the gain and loss scores and removing the outliers, the number of participants was 27 in both the structured and the structured + reformulation groups and 25 in the unstructured group. Descriptive statistical tests were then carried out for all groups. Results are indicated in Table 5.2 (see Appendix H for T1 and T2 descriptive statistical results).

Table 5.2 *Descriptive Statistics Results for T1 – T2 Lexical Complexity Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	-.12	.65	.89	-1.81
Str. + ref.	27	-.19	.55	1.11	-1.10
Unstr.	25	-.01	.73	1.17	-1.62

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Results of the descriptive statistics indicate that all three groups declined in lexical complexity from T1 to T2 with the largest decrease belonging to the structured + reformulation group ($M = -.19$). The unstructured group, however, showed the smallest decrease ($M = -.01$).

The differences among the groups were analysed employing a one-way between groups ANOVA test. Results of this test indicate that the differences among the groups in terms of the gain/loss in lexical complexity from T1 to T2 did not reach significance; $F(2, 77) = .55, p > .05$.

The d values were .11 for the structured vs. the structured + reformulation group comparison, .15 for the structured vs. the unstructured group comparison, and .27 for the structured + reformulation vs. the unstructured group comparison. These effect sizes indicate no effects for the variables.

5.2.3. Grammatical Accuracy

After computing the gain/loss scores and removing the outliers, the number of participants was 27 in the structured group, 28 in the structured + reformulation group, and 26 in the unstructured group. Descriptive statistical tests were then carried out for all groups. Results are reported in Table 5.3 (see Appendix H for T1 and T2 descriptive statistical results).

Table 5.3 *Descriptive Statistics Results for T1 – T2 Grammatical Accuracy Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.00	.10	.21	-.13
Str. + ref.	28	.01	.12	.20	-.30
Unstr.	26	-.03	.11	.17	-.25

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Results of descriptive statistical tests indicate that the structured group made no progress from T1 to T2 ($M = 0$). The structured + reformulation group improved in grammatical accuracy from T1 to T2 ($M = .01$), whereas the unstructured group declined ($M = -.03$). In terms of group differences, there are very small differences across the groups.

The above mean differences were analysed employing a one-way between groups ANOVA test. Results indicate no significant differences across the groups in terms of the extent to which they lost or gained in grammatical accuracy from T1 to T2; $F(2, 79) = 1.26, p > .05$.

The d values were .09 for the structured *vs.* the structured + reformulation group comparison, .28 for the structured *vs.* the unstructured group comparison, and .34 for the structured + reformulation *vs.* the unstructured group comparison. These effect sizes suggest that the variables had no effect on the repeat performance of the same task in terms of grammatical accuracy.

5.2.4. Mechanical Accuracy

After the gain/loss scores were computed and the outliers were removed, the number of participants was 27 in both the structured and the structured + reformulation groups and 25 in the unstructured group. Then the descriptive statistical tests were carried out for all groups. Results of these tests are presented in Table 5.4 (see Appendix H for T1 and T2 descriptive statistical results).

As indicated in Table 5.4, all groups showed a small improvement in mechanical accuracy from T1 to T2. The mean value was the same for both the structured and the structured + reformulation groups ($M = .03$ for both groups) while the unstructured group mean was smaller ($M = .02$).

Table 5.4 *Descriptive Statistics Results for T1 – T2 Mechanical Accuracy Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.03	.15	.34	-.33
Str. + ref.	27	.03	.16	.33	-.24
Unstr.	25	.02	.14	.29	-.24

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

The mean differences were analysed employing a one-way between groups ANOVA test. Results of this test indicate no significant differences among the groups in terms of the extent to which they gained in mechanical accuracy from T1 to T2; $F(2, 77) = .07$, $p > .93$.

The d values were 0 for the structured *vs.* the structured + reformulation group comparison, .07 for the structured *vs.* the unstructured group comparison, and .06 for the structured + reformulation *vs.* the unstructured group comparison. These d values indicate no effect for the variables.

5.2.5. Fluency 1 (total number of words per text)

After computing the gain/loss scores and removing the outliers, the number of participants was 27 in both the structured and the structured + reformulation groups and 26 in the unstructured group. Descriptive statistical tests were then carried out for all groups. Results are indicated in Table 5.5 (see Appendix H for T1 and T2 descriptive statistical results).

Table 5.5 *Descriptive Statistics Results for T1 – T2 Fluency1 Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.62	41.76	80.00	-83.00
Str. + ref.	27	-53.37	41.96	62.00	-125.00
Unstr.	26	.07	59.61	103.00	-137.00

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

As Table 5.5 illustrates, both the structured and the unstructured groups gained in fluency 1 from T1 to T2 (Str. M = .62, and Unstr. M = .07). However, the structured + reformulation group declined substantially from T1 to T2 (M = -53.37). Across the groups, there were considerable differences with the biggest difference belonging to the structured *vs.* the structured + reformulation group comparison.

The mean differences were examined across the groups employing a one-way between groups ANOVA test and results indicate that there were significant differences among the groups regarding the decline or the improvement in fluency 1 from T1 to T2; $F(2, 78) = 11.05$, $p < .05$. Therefore, Bonferroni tests were carried out for the purpose of post-hoc analyses. Results, shown in Table 5.6, revealed a significant difference between the structured and the structured + reformulation groups ($p < .05$, $d = 1.28$). Mean differences between the structured + reformulation and the unstructured groups also reached significance ($p < .05$, $d = 1.03$). However, no significant differences were found between the structured and the unstructured groups ($p > .05$).

Table 5.6 *Bonferroni Post Hoc Test Results for Fluency 1 Gain/Loss Scores*

Group	Group	Mean Difference	Std. Error	Sig.
Str.	Str. + ref.	54.00	13.15	.00*
Str.	Unstr.	.55	13.28	1.00
Unstr.	Str. + ref.	53.44	13.28	.00*

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

The effect size was also calculated for the structured vs. the unstructured group comparison and the resulting value was .01 which indicates no effect on the performance.

The above statistical analyses indicate that the degree to which the three groups lost or gained in fluency 1 from T1 to T2 significantly differed between the structured + reformulation group and the other two groups. In other words, the structured + reformulation group scores declined considerably from T1 to T2 compared to the other two groups which improved slightly in fluency from T1 to T2. The reported *d* values for these results show a large effect of the variables on the fluency of the performance of the same task.

5.2.6. Fluency 2 (number of words per T-unit)

After computing the gain/loss scores and removing the outliers, the number of participants was 27 in both the structured and the structured + reformulation groups and 26 in the unstructured group. The descriptive statistical tests were then carried out for all groups. Results are presented in Table 5.7 (see Appendix H for T1 and T2 descriptive statistical results).

Table 5.7 *Descriptive Statistics Results for T1 – T2 Fluency2 Gain/loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.24	2.63	6.29	-5.34
Str. + ref.	27	-.37	2.00	5.41	-6.06
Unstr.	26	1.00	2.57	5.72	-5.15

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Results of the statistical tests indicate that both the structured and the unstructured groups improved in fluency 2 with the unstructured group showing the largest increase ($M = 1.00$). However, the fluency of the structured + reformulation group dropped from T1 to T2 ($M = -.37$).

To examine the above mean differences, a Kruskal-Wallis H test was employed as the scores in the structured + reformulation group were not normally distributed (Kolmogorov-Smirnov $p < .05$). Results of the Kruskal-Wallis H test indicate that the differences in the mean values across the three groups were significant; $\chi^2(2, 78) = 7.25, p < .05$. Thus, post hoc analyses were performed employing further Kruskal-Wallis H tests. Results of these tests are shown in Table 5.8.

Table 5.8 *Kruskal-Wallis H Post Hoc Test Results for Fluency 2 Gain/loss Scores*

Group	Group	Chi-Square	df	Sig.
Str.	Str.+ref.	2.06	1	.15
Str.	Unstr.	1.75	1	.18
Str.+ref.	Unstr.	7.03	1	.01*

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

Results of the post hoc tests indicate that the only significant difference belongs to the structured + reformulation *vs.* the unstructured group comparison ($p < .05$, $d = .59$).

The d values for non-significant results were .26 for the structured *vs.* the structured + reformulation group comparison and .29 for the structured *vs.* the unstructured group comparison.

In summary, results indicate that the unstructured group outperformed the structured + reformulation group in terms of the number of words per T-unit. The reported d value for this effect was small. Effect sizes for non-significant results show no effect for the variables.

5.2.7. Summary of the Results for Research Question One

Results for research question one are summarised as below with regard to either significant differences or small to large effect sizes:

- (1) Compared to the unstructured group, the structured + reformulation group's structural complexity decreased with a small effect size. This effect was, however, not significant.
- (2) Compared to both the structured and the unstructured groups, the structured + reformulation group's fluency 1 (i.e. total number of words per text) decreased significantly with a large effect size.
- (3) Compared to the unstructured group, the structured + reformulation group's fluency 2 (i.e. number of words per T-unit) decreased significantly with a small effect size.

5.3. Results for Research Question Two

This section focuses on the results of the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of a new task of the same type in terms of micro-measures.

5.3.1. Structural Complexity

After computing the gain/loss scores and removing the outliers, the number of participants was 27 in both the structured and the structured + reformulation groups, and 26 in the unstructured group. Descriptive statistical tests were then carried out for all groups. Results are indicated in Table 5.9 (see Appendix H for T1 and T3 descriptive statistical results).

Table 5.9 *Descriptive Statistics Results for T1 – T3 Structural Complexity Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.11	.21	.54	-.42
Str. + ref.	27	.01	.20	.35	-.39
Unstr.	26	-.03	.25	.59	-.55

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Results of the descriptive statistics show that structural complexity improved from T1 to T3 in both the structured and the structured + reformulation groups. However, this increase is negligible with regard to the structured + reformulation group ($M = .01$). The unstructured group's structural complexity declined from T1 to T3 ($M = -.03$).

The above mean differences were analysed employing a one-way between groups ANOVA test. Results of this test indicate that the differences among the groups in terms of the structural complexity gain/loss from T1 to T3 were statistically significant; $F(2, 78) = 3.17, p < .05$. Hence, Bonferroni post hoc tests were carried out for further analyses. Results of these tests, indicated in Table 5.10, revealed that only the difference between the structured and the unstructured groups reached significance ($p < .05, d = .60$).

Table 5.10 *Bonferroni Post Hoc Test Results for Structural Complexity Gain/Loss Scores*

Group	Group	Mean Difference	Std. Error	Sig.
Str.	Str. + ref.	.10	.06	.30
Str.	Unstr.	.15	.06	.04*
Unstr.	Str. + ref.	.05	.06	1.00

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

Effect sizes for non-significant results were .48 for the structured vs. the structured + reformulation group comparison and .17 for the structured + reformulation vs. the unstructured group comparison.

In summary, results indicate that the structured group produced significantly more complex language than the unstructured group. The reported d value is small for this effect. For non-significant results, there are no or small effects for the variables.

5.3.2. Lexical Complexity

After computing the gain/loss scores and removing the outliers, the number of participants was 26 in the structured group, 28 in the structured + reformulation group, and 25 in the unstructured group. Descriptive statistical tests were then carried out for all groups. Results are indicated in Table 5.11 (see Appendix H for T1 and T3 descriptive statistical results).

Table 5.11 *Descriptive Statistics Results for T1 – T3 Lexical Complexity Gain/Loss Scores*

Group	N	Mean	Std.	Max value	Min Value
Str.	26	.27	.66	1.78	-1.19
Str. + ref.	28	.09	.55	1.41	-1.05
Unstr.	25	.41	.50	1.36	-.94

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

As the results show, all groups increased in lexical complexity from T1 to T3 with the largest increase belonging to the unstructured group ($M = .41$). The structured + reformulation group, however, showed the smallest improvement ($M = .09$).

The above mean differences were examined employing a one-way between groups ANOVA test. Results of this test indicate that the differences among the groups did not reach significance; $F(2, 77) = 2.12, p > .05$.

The d values were .29 for the structured vs. the structured + reformulation group comparison, .23 for the structured vs. the unstructured group comparison, and .60 for the structured + reformulation vs. the unstructured group comparison. These effect sizes indicate no or small effects for the variables.

5.3.3. Grammatical Accuracy

After computing the gain/loss scores and removing the outliers, the number of participants was 26 in both the structured and the unstructured groups and 27 in the structured + reformulation group. Descriptive statistical tests were then carried out for all groups. Results are presented in Table 5.12 (see Appendix H for T1 and T3 descriptive statistical results).

Table 5.12 *Descriptive Statistics Results for T1 – T3 Grammatical Accuracy Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	26	.07	.10	.28	-.16
Str. + ref.	27	.03	.09	.19	-.17
Unstr.	26	-.03	.09	.14	-.21

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

According to the results of the descriptive statistics, both the structured and the structured + reformulation groups gained in grammatical accuracy from T1 to T3 with the larger increase belonging to the structured group ($M = .07$). However, the unstructured group's grammatical accuracy decreased from T1 to T3 ($M = -.03$).

The above mean differences were examined employing a one-way between groups ANOVA test and results indicate significant differences among the groups in terms of the grammatical accuracy gain/loss scores from T1 to T3; $F(2, 77) = 7.42$, $p < .05$. Therefore, Bonferroni post hoc tests were carried out for further analyses. Results of the Bonferroni tests are presented in Table 5.13 and indicate that only the mean difference between the structured and the unstructured group was significant ($p < .05$, $d = 1.05$). Mean differences between the unstructured and the structured + reformulation group approached significance ($p = .07$, $d = .66$).

Table 5.13 *Bonferroni Post Hoc Test Results for Grammatical Accuracy Gain/Loss Scores*

Group	Group	Mean Difference	Std. Error	Sig.
Str.	Str. + ref.	.04	.03	.36
Str.	Unstr.	.10	.27	.00*
Unstr.	Str. + ref.	.06	.03	.07

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

In spite of the non-significant result, the effect size was also calculated for the structured vs. the structured + reformulation group comparison, and the resulting value was .42.

In summary, results indicate that the structured group significantly outperformed the unstructured group with a large effect size. The difference between the structured + reformulation and the unstructured groups is close to significance with a small effect. In other words, the structured + reformulation group produced more accurate language than the

unstructured group. The difference between the structured and the structured + reformulation groups did not reach significance, and the resulting d value shows a small effect.

5.3.4. Mechanical Accuracy

After the gain/loss scores were computed and the outliers were removed, the number of participants was 26 in the structured group, 28 in the structured + reformulation group, and 25 in the unstructured group. Descriptive statistical tests were then carried out for all groups. Results are presented in Table 5.14 (see Appendix H for T1 and T3 descriptive statistical results).

Table 5.14 *Descriptive Statistics Results for T1 – T3 Mechanical Accuracy Gain/Loss Scores*

Group	N	Mean	Std.	Max value	Min Value
Str.	26	.06	.14	.36	-.26
Str. + ref.	28	.03	.14	.32	-.18
Unstr.	25	-.03	.10	.16	-.34

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

As the above table shows, both the structured and the structured + reformulation groups gained in mechanical accuracy with the larger increase belonging to the structured group ($M = .06$). However, the unstructured group declined from T1 to T3 ($M = -.03$).

The above mean differences were analysed employing a one-way between groups ANOVA test and results indicate a significant difference among the groups in terms of the mechanical accuracy gain/loss from T1 to T3; $F(2, 77) = 3.78$, $p < .05$. Therefore, post hoc Bonferroni tests were carried out for further analyses. Results of these tests, presented in Table 5.15., revealed a significant difference between the structured and the unstructured groups ($p < .05$, $d = .73$).

The d values for non-significant results were .21 for the structured vs. the structured + reformulation group comparison, and .49 for the structured + reformulation vs. the unstructured group comparison.

Table 5.15 *Bonferroni Post Hoc Test Results for Mechanical Accuracy Gain/Loss Scores*

Group	Group	Mean Difference	Std. Error	Sig.
Str.	Str. + ref.	.03	.03	1.00
Str.	Unstr.	.09	.04	.03*
Unstr.	Str. + ref.	.07	.03	.16

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

In conclusion, results of the above analyses indicate that the structured group outperformed the unstructured group with regard to mechanical accuracy. The reported *d* value for this effect is medium-sized. No significant differences were found between the structured + reformulation group and the other two groups, and the effect sizes associated with these results show either no or a small effect.

5.3.5. Fluency 1 (total number of words per text)

After computing the gain/loss scores and removing the outliers, the number of participants was 26 in both the structured and the unstructured groups and 28 in the structured + reformulation group. Descriptive statistics were then produced for all groups. Results are presented in Table 5.16 (see Appendix H for T1 and T3 descriptive statistical results).

Results of the descriptive statistical tests revealed that all groups improved in fluency 1 from T1 to T3 with the greatest increase belonging to the unstructured group ($M = 32.23$). The structured + reformulation group, however, showed the smallest increase ($M = 19.85$).

Table 5.16 *Descriptive Statistics Results for T1 – T3 Fluency 1 Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	26	21.84	63.14	165.00	-.86.00
Str. + ref.	28	19.85	64.07	142.00	-99.00
Unstr.	26	32.23	53.00	119.00	-77.00

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

The mean differences were analysed performing a one-way between groups ANOVA test and no significant differences were found among the groups with regard to fluency 1 gain/loss from T1 to T3; $F(2, 78) = .32$, $p > .05$.

The *d* values were .03 for the structured vs. the structured + reformulation group comparison, .17 for the structured vs. the unstructured group comparison, and .21 for the structured +

reformulation vs. the unstructured group comparison. These d values show no effect for the variables.

5.3.6. Fluency 2 (total number of words per T-unit)

After computing the gain/loss scores and removing the outliers, the number of participants was 26 in the structured group, 28 in the structured + reformulation group, and 25 in the unstructured group. Descriptive statistics were then produced for all groups. Results are presented in Table 5.17 (see Appendix H for T1 and T3 descriptive statistical results).

Table 5.17 *Descriptive Statistics Results for T1 – T3 Fluency 2 Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min value
Str.	26	-.28	1.84	3.29	-3.77
Str. + ref.	28	-.69	1.63	2.17	-4.79
Unstr.	25	-.91	2.01	2.44	-5.53

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

As the results of the descriptive statistics indicate, all groups declined in fluency 2 from T1 to T3 with the greatest decrease belonging to the unstructured group ($M = -.91$). The structured group, however, showed the smallest decrease ($M = -.28$).

The mean differences were analysed performing a one-way between groups ANOVA test. Results of this test indicate that the differences among the groups in terms of the extent to which they lost in fluency 2 from T1 to T3 did not reach significance; $F(2, 77) = .79$, $p > .05$.

The d values were .23 for the structured vs. the structured + reformulation group comparison, .32 for the structured vs. the unstructured group comparison, and .12 for the structured + reformulation vs. the unstructured group comparison. These d values show no effect for the variables.

5.3.7. Summary of the Results for Research Question Two

Results for research question one are summarised as below with regard to either significant differences or small to large effect sizes:

- (1) The structured group significantly outperformed the unstructured group with regard to structural complexity (small effect), grammatical accuracy (large effect), and mechanical accuracy (medium effect).

- (2) The structured group outperformed the structured + reformulation group with a small effect in terms of structural complexity and grammatical accuracy. This effect was not significant, though.
- (3) The unstructured group outperformed the structured + reformulation group with a small effect in terms of lexical complexity. This effect was, however, not significant.
- (4) The structured + reformulation group outperformed the unstructured group with a small effect in terms of both grammatical and mechanical accuracy. However, these effects were not significant.

Table 5.18 summarises the results for structural complexity, lexical complexity, grammatical accuracy, mechanical accuracy, and fluency of the repeat performance of the same task and the performance of a new task of the same type.

Table 5.18 *Summary of the Main Results of Research Questions One and Two*

Measure	Repeat Performance of the Same Task	Performance of a New Task of the Same Type
SC	Unstr. > Str. + ref. ($d = .60$)	Str. > Str. + ref. ($d = .48$) Str. > Unstr.* ($d = .60$)
LC	-	Unstr. > Str. + ref. ($d = .60$)
GA	-	Str. > Unstr. * ($d = 1.05$) Str. > Str. + ref. ($d = .42$) Str. + ref. > Unstr. ($d = .66$)
MA	-	Str. + ref. > Unstr. ($d = .49$) Str. > Unstr. * ($d = .73$)
F1	Str. > Str. + ref.* ($d = 1.28$) Unstr. > Str. + ref.* ($d = 1.03$)	-
F2	Unstr. > Str. + ref.* ($d = .59$)	-

Note. SC = Structural complexity, LC = Lexical complexity, GA = Grammatical accuracy, MC = Mechanical accuracy, F1 = Fluency 1, F2 = Fluency 2, Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

5.4. Discussion of the Results of Research Questions One and Two

The first two research questions investigated the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of the same task and the performance of a new task of the same type in terms of micro-measures. There was no evidence that repeating the structured task + reformulation had any significant effect on the repeat performance of the same task and the performance of a new task of the same type. Repeating the structured task as opposed to the unstructured task had no notable effects on the repeat performance of the same task, but increased the structural complexity, grammatical accuracy, and mechanical accuracy of the performance of a new task of the same type. This section will focus on the explanation of these results. Learners' written texts and their responses to the questionnaires will be used to support the explanations.

When reformulation is considered, findings indicate that in general, repeating the structured task + reformulation had little effect on the repeat performance of the same task and performance of a new task of the same type. In fact, reformulation resulted in less fluent and a less structurally complex repeat performance of the same task and a less lexically complex performance of a new task of the same type. These results are not surprising where the performance of a new task of the same type is concerned, as giving learners only one chance to compare their texts with their reformulated versions does not seem to be sufficient for any transfer of learning to take place, especially as the learners had no prior experience of reformulation and were completely unfamiliar with this kind of feedback. Regarding the repeat performance of the same task, although findings reflect the results of the study by Jung (2013), who found no clear effect of feedback on the repeat performance, it was expected that the repeat performance in this study would improve following reformulation as reported by previous studies (Qi & Lapkin, 2001; Swain & Lapkin, 2002; Adams, 2003). These unexpected results can be explained by referring to the way texts were reformulated, the way reformulation was implemented in class, and the way the learners' written performance was analysed.

In the previous studies (Qi & Lapkin, 2001; Swain & Lapkin, 2002; Adams, 2003; Sachs & Polio, 2007), texts were reformulated mainly in terms of sentence structure and lexis with no changes made to the order of the sentences, whereas in this study texts were reformulated not only in terms of syntax and lexis, but also in terms of paragraph and textual organisation. That is, the text was divided into paragraphs where previously it had consisted of a single paragraph,

and/or sentences which were irrelevant to the topic of the paragraph were removed and placed in the paragraph where they belonged, and if no relevant paragraphs existed in the text to place the sentences in, the sentences were repositioned in the text as an independent paragraph. For example, in the following paragraph, learner A writes about both ‘the reasons why young people get addicted’ and ‘the problem of drug addiction and its effects’. In the reformulated text, this paragraph was divided into two different paragraphs; one focusing on the reasons and the other on the effects of addiction.

Original Paragraph:

Nowadays the drug problem is one of the most important concerns of lots of people, especially parents. Their worries and anxiousness about their children isn’t so weird, Because teens are the most effective part of society and it’s because of their age. Youth in these ages are curious about everything and this is not a bad thing until it causes trouble for them. Usually people will be attracted to things that they’re not informed enough about it. So youth would be attracted more than others to those things because of their curiosity. One of problems that have been spreaded in the society is drug addiction. And some young people really don’t know the affects of it. Drug addiction doesn’t just effect the addicted person, It’ll effect all the people in the society. If one person in a family has faced this problem, all member of family will be involved in it and may help him/her or by their wrong treats make him/her to not be comfortable with the family and find other people to open up and maybe spread the problem to the society and not find a suitable cure for it.

Reformulated text:

Nowadays, the problem of drugs is one of the major concerns of a lot of people, especially parents. It is not so weird that parents are worried about and anxious for their children because teens are inexperienced and make up the most important part of the society. Youth are curious about everything, which is not bad as long as it does not cause trouble for them. Besides, people are usually attracted to things that they are not informed about. So, youth are attracted to these things more than other age groups because of their curiosity.

Although drug addiction has existed in the society for a long time, some young people do not really know about its effects. Drug addiction does not affect just the addict, but all of the people in the society. If someone in a family has this problem, all of the other members of that family will also be involved with it. Families may apply wrong treatments to help the addicts which may make the addicts feel uncomfortable with their families and turn to other people to open up. They may also infect the whole society and fail to find an effective cure for it.

In another example, learner B proposes two different solutions to a problem in the same paragraph. The original paragraph was therefore divided into two paragraphs in the reformulated text, so that each paragraph focused on one solution.

Original text:

Another solution for this big problem is public education. by educating people we can increase their awareness and decrease the number of addictions. also parents by controlling their childrens relationships and by educating them can prevent this problem.

Reformulated text:

Another solution for this big problem is public education. By educating people, their awareness will increase and as a result, the number of addicts will decrease.

Parents can also prevent this problem by controlling their children's relationships and educating them.

In the following example, the final part of a paragraph by learner C is irrelevant to the topic of the paragraph. Thus, in the reformulated text, that part was removed and positioned as an independent paragraph because it did not belong to any other paragraphs in the text.

Original text:

If the government want to find solution, it should obligate a lot of police to catch the drug dealers and addicts. so they can reduce the main reason of this problem. Until they don't stop drug dealers, they can't manage to reduce

this problem, because it became more and more. *By the way, families with financial burden can't help their children. They must get help from especial clinic which their expertise is addiction.*

Reformulated text:

If governments want to find a solution, they should hire a lot of police officers to catch drug dealers and addicts, so that they can eliminate the main reason for this problem. This problem cannot be reduced and will get worse if drug dealers are not stopped.

By the way, families with financial burden can't help their children. They must get help from especial clinic which their expertise is addiction.

Therefore, learners needed to compare and notice not only the syntactical and lexical differences, but also the textual and organisational differences between their original performance and the reformulated texts. The conclusion therefore is that considering that the texts were quite long and that the learners had limited time (i.e. 15 minutes) to compare their texts with their reformulated version, it must have been very difficult for them to notice all the differences and address them when they rewrote their texts. Thus, they were not able to correct their errors in the subsequent performance. Moreover, the way texts were reformulated may have given the learners the impression that some of the content they had included in their texts was irrelevant and did not fit into any paragraph, so that content was better to be removed. This therefore resulted in considerably shorter texts and less fluent performance at the second performance. This explanation is also supported by learners' responses to the questionnaire; more than half of the participants in the structured + reformulation group said that they had written shorter texts the second time, which indicates that they thought one way to improve their texts was to remove information. For example, learner D said:

Writing long texts does not mean writing better texts.

Learner E mentioned:

Quality better than quantity! So, I wrote less.

Learner F responded:

I wrote less and better, so that it is native-like.

A few participants also mentioned that they deleted content and avoided focusing on details.

It should also be noted that in this study reformulation was implemented in a different way from previous studies (Qi & Lapkin, 2001; Swain & Lapkin, 2002; Adams, 2003; Sachs & Polio, 2007). That is, unlike previous studies, learners wrote their first texts and also compared their original texts with the reformulated texts *individually*. This is important because when learners write and compare their texts with their reformulated version collaboratively, they have a chance to *discuss* the differences. This may result in ‘substantive’ noticing (i.e. noticing with reasoning [Qi & Lapkin, 2001; Swain & Lapkin, 2002]), which in turn leads to improvement in the subsequent performance. Moreover, unlike previous studies, when performing the task for the second time, the learners did not have access to their original texts but rather were asked to write a whole new text. This could have also affected the results as rewriting while having access to the original text is a completely different process from writing a whole new text.

The way the repeat performance was analysed in this study was different, too. That is, while previous reformulation studies counted the number of the reformulations that were applied in the repeat performance, this study looked at error-free clauses. Thus, most of the repeat performance structures that incorporated changes from the reformulated texts were not counted because the clauses they were embedded in were erroneous. For example, a sentence from learner G’s original and repeat performances shows that the learner successfully changed ‘all the time’ to ‘always’ and used the correct collocation for the word ‘problem’ in the repeat performance. However, because the subject-verb agreement was violated, the clause was counted as erroneous.

Example 1: Original sentence: ... lack of awareness *all the time makes* a problem for people.

Reformulated version: ... lack of awareness *always causes* problems for people.

Repeat performance: ... lack of awareness *always cause* a problem for people.

Another example from learner H’s texts also shows that this learner was successful in using correct collocations for *establishing camps* and *rehabilitation camps*, but since there were other errors in the clause (i.e. *extreeming*, *other solution*), the clause was counted as erroneous.

Example 2: Original sentence: *Making some medical places, public education, increasin the awareness and ... are other ways that help the people to distinguish them and stop using these drugs.*

Reformulated sentence: *Establishing some rehabilitation camps, providing public education, increasing awareness about drugs, and are some other ways that can help people recognise addictive drugs and stop using them.*

Repeat performance: Public education, extreeming the awarness and *establishing some rehabilitation camps* are some other solution that help people to stop abusing drugs.

These as well as many other similar instances show that the reformulated texts enabled the participants to notice and correct their errors in the repeat performance, but did not result in accurate clauses. It is possible that if the effects of reformulation had not been measured in terms of accurate clauses (a general measure of accuracy) but in terms of ‘reformulations incorporated’, positive effect for reformulation would have been found for accuracy. However, it is clear that reformulation did not enable the learners to improve their overall accuracy.

Finally, it is worth mentioning that when writing the repeat performance, the participants in the reformulation group had less time than the other groups. That is, the other groups had a whole session to write, whereas the reformulation group had to first compare their original performance with their reformulated version and then perform the task for a second time, which gave them less time for writing. They therefore might have been under pressure when writing or have had insufficient time to plan their performance prior to writing, which could have affected their performance (Ellis & Yuan, 2005).

With regard to task structure, there was no evidence that repeating the structured task as opposed to the unstructured task had any effect on the repeat performance of the same task. This is at odds with the findings of the written task repetition studies by Larsen-Freeman (2006), Indrarathne (2013), and Jung (2013), who reported positive effects of task repetition on the repeat performance of the same task. The explanation for this can be that these learners were not familiar with a task-based approach to the teaching of writing. In other words, participating in this study was the learners’ first experience with task-based language teaching

methodology, and therefore they were not only unfamiliar with task repetition as a type of task planning but also unfamiliar with note-expansion tasks. Moreover, in the Iranian educational system, learners are always prevented from repeating the same content in their essays when they write on the same topic on two different occasions. It is believed that a good piece of writing does not repeat the same content and good writers are always able to produce new ideas and content for their texts. It can thus be concluded that learners' unfamiliarity with task repetition and note-expansion tasks as well as their educational background affected their second performance of the same task. That is, instead of repeating the same task, learners may have tried to write a whole new text with different content. This explanation is also supported by the participants' responses to the questionnaires; 22 out of 53 participants in the structured and the unstructured groups mentioned that task repetition helped them to better understand the notes and find ways to make connections between the notes within and across the paragraphs, which confirms that learners were unfamiliar with note-expansion tasks and how they were required to perform them. Furthermore, 36 participants reported that when repeating the task for the second time, they either tried to change the whole of the content in order to avoid repeating the content, or they changed some parts or paragraphs, or they kept the initial content but tried to add new content. For example, learner I wrote the following two texts when performing the task for the first and the second time.

Time 1

Now a day "Drugs" is a big problem between families specially among who have child (youth). now a day the youth want to experience the new things, in that point I mean they active their sense of curiosity, and because of lack of awareness they want to use and experience drugs for fun and etc.

We call to this people Dug (Drug) addiction who have a bad effect on their society and their families too. With processing or progresing in science the new products of drugs is increasing. because of that most of the countries have a big program to solving of this problem for example:

by using of police manpower against drug dealers, producing of places like camps for addict peoples and etc.

near of this ways some educations, like public education for example in tv. and by bookes are using by governments.

in conclusion, management of this should be start from families by health educating of their chields, after that the governments can help to the families in this issues.

Time 2

Now a day the drugs have became a big problem almost for most of the countries, this disaster has spreaded specially between youth, because of their curiosity sense and lack of awareness, because of that they want to experience new things in their life.

effects of drugs on societies and families are so clear us. for example destroying of health, wasting of the money and etc. this materials have a bad effects on body but addicted people don't know anything about this issue because of that they use this drugs for fun.

for solving of this problem, families have a big influence on their children and methods of their bring up is so important. and society can help them in this field, by making some camps for remedication of addicted people.

Police can manage this problem too, with making of a strong protection in borders of their countries and reducing of addicted people population in their town.

in conclusion, awareness of drug should be increase in countries. and families have a big influence in this field.

As can be seen, learner I added some new content when repeating the task. That is, a paragraph was added about the effects of addiction (see *italics*), and the two solutions proposed are different from the ones listed in the initial performance (see underlined). Similarly, in the following example, learner J changed the content of the second paragraph and listed different effects when repeating the task (see *italics*). A new solution was also added in paragraph three (see underlined).

Time 1

first of all I want to talk about why mostly young population of our society start to use drugs and become addicted; when there isn't enough education so there is lack of awareness about what is really these drugs are and what are

they really for? They don't really know its effects and harms to them. Young generation is always trying to find ways of fun. They want to experience new things, new feelings and senses, they're full of energy and want to use it the way that gives them satisfaction so when there isn't easy reach to some pleasing habits it causes repetitive use of it and it ends with addiction.

When a person addicted to drugs while it harm him/her health, it could probably cause some mental problems which means that, that person cannot communicate well with his/her family or friends, either cannot have a good and benefitable work in his/her occupation; so there we have an useless person either in family and society. This person could also make others life full of problems.

We could offer some solutions such as strict laws of government about controlling drug dealers, having full control on strips of country to know what is coming or going; our government could have make police mans in new branches of drug dealers stop or drug users stop; any way, these ways might not be efficient and police mans or governors should have to do different operation about stopping the raising use of drugs, such as estimating the addictions and founding companies to help them not to use and abandon.

Another solution to decrease the use of drugs and addicted people is to make some public courses for each member of society classified in ages like classes for childrens which has information useful for their age and classes for young or old generations which their goal is the same. To make every one aware of drugs and its different harms and effects to both users and society's when there are educated and aware peoples they are less useless persons too.

Time 2

Nowadays we can see that a large group of young generation in every society are addicted to different types of drugs and sometimes alcoholic stuff, which means that this unfortunate event happens mostly because of sense of curiosity which leads them to test and become addicted or this may happens because these people do not have enough education, or they do not know about the result of using those stuff, so that may become a major problem.

In most of the cases in different societies, when a person is addicted; while addiction harms that persons health and life, it could easily ruins his family members life too; that person could easily discard the honor of his family or may sometimes attempts to behave the way that no one expects him to do. In some cases addiction could even lead the person to end up with hurting others in society for gaining the money of drugs, this could also be the beginning for some crimes or riots in the society.

Here are some ways being offered to prevent the new generations from being addicted, which is believed that it's inefficient in general; for example, if there wasn't extra money so there wasn't addicted people which just become addicted because of their extra money and extra time. If government trie to control reach peoples money and it's source and know how they're using it, this might not happens, but this is a difficult task for police manpower and I don't think it could result in good, another solution is that polices could controll the strips of the country and try to be more conscious about smugglers to stop them from bringing drugs and stop distributing those stuffs, although the police manpower group of drugs and narcotics could try to focus on people whom use drugs and prevent them or make some charities of doctors to help them abandon drugs.

We could also offer some other basic solution which might be more efficient, like public education for every age group of society, like classes which is being classified for youngers and olders, the purpose is the same but the ways is different, this way everyone could attempts to go those free calasses and get information about the dangers of addiction and to know that it's not really worth even to try it once, I think if there are more educated people in society so the society will improve in every major and the rate of crime would become lower because of a cultural change.

The fact that learners changed the content of their texts on the second performance can also be attributed to the nature of note-expansion tasks. In their studies, Bygate (1996, 2001), Gass et al. (1999), Larsen-Freeman (2006), and Indrarathne (2013) used narrative tasks with which there is less possibility of changing the content when repeating the task, whereas with note-

expansion tasks learners have the opportunity to add their own ideas to the text even if they stay loyal to the task notes and do not change them.

Therefore, it can be inferred that when learners engaged in performing the same task for the second time, their attentional resources were mainly allocated to either discovering how the task should be performed or finding new content to avoid repeating the same content or both, and since attentional resources are limited when learning a new language (Skehan, 1998b), learners were unable to allocate attention to improve their performance in terms of complexity, accuracy, and fluency.

There was, however, strong evidence that repeating the structured as opposed to the unstructured tasks resulted in significantly greater gains in structural complexity, grammatical accuracy, and mechanical accuracy of the performance of a new task of the same type. Considering the discussion above with regard to the effect of task repetition on the performance of the same task, it can be concluded that these gains are a result of the interaction between task structure and task repetition. In other words, on one hand, repeating note-expansion tasks helped learners discover this task type and figure out how connections should be made between the notes to expand them into full paragraphs and texts. This consequently released attention to be allocated to different areas of performance (i.e. complexity and accuracy) (Skehan, 1998b). On the other hand, performing tasks where the text is logically structured (i.e. one paragraph presents the situation leading to a statement of the problem followed first by a solution that is negatively evaluated and then a second solution that is positively evaluated) allowed learners to focus their attention on the micro aspects of the texts by easing the processing burden of the task and setting attentional resources free to be allocated to different areas of performance. However, these results are not in line with the results of previous studies of oral task performance (Skehan & Foster, 1999; Tavakoli & Skehan, 2005; Tavakoli & Foster, 2008; Tavakoli, 2009) which reported positive effects of task structure on accuracy and fluency but not on complexity. This is not surprising in that writing is different from speaking as it does not happen in real time and gives the writer more time to formulate, execute, and monitor the written text (Ellis & Yuan, 2005; Manchon, 2014) leading to both greater complexity and accuracy. This is also supported by the results of the study by Ellis and Yuan (2005) who found written production to be more complex and accurate but less fluent than oral production. Therefore, it can be concluded that for task repetition effects to be carried over to a new task of the same type (i.e. for learning to take place) the tasks being repeated should have a clear structure.

Trade-offs are also evident between structural and lexical complexity in the repeat performance of the structured task. Results show an increase in structural complexity but not in lexical complexity. This could perhaps be explained by referring to Kellogg's (1996) model of writing. According to this model, writers attend to syntactical properties of their message and select appropriate lexical units at the 'translation' stage. There is thus a competition between the structural and lexical complexity of the language to be produced. In the case of this study, the learners appear to have prioritised syntax over lexis by focusing on producing more structurally complex language.

Overall, the results of this study show that reformulation had no notable effect on the learners' writing no matter whether they were repeating the same task or performing a new task of the same type. There was no difference in the effects of the structured and unstructured task repetition on the repeat performance of the same task but a clear advantage was found for the structured task repetition with regard to structural complexity as well as grammatical and mechanical accuracy when it came to the performance of a new task of the same type.

Chapter 6 Results and Discussion 2

This chapter focuses on the results of the post-writing performance, investigating the third research question: what is the effect of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of a new task of a different type in terms of both micro- and macro-level measures? Results are presented in two sections: within-group differences (i.e. time differences) and between-group differences.

6.1. Micro-measures

As mentioned in the previous chapter, micro-measures in this study include structural and lexical complexity, mechanical and grammatical accuracy, and fluency. To calculate structural complexity, the total number of clauses was divided by the total number of T-units. Lexical complexity was measured by employing the Giraud Index (see 5.1 for more details). Grammatical accuracy was measured in terms of the number of grammatically and lexically error-free clauses to total number of clauses. Mechanical accuracy was measured in terms of the number of orthographically error-free T-units to total number of T-units and was judged based on spelling, punctuation, and capitalisation. Fluency was measured both in terms of the number of words per T-unit and the total number of words per text. Then 20 percent of the data was checked for the inter-rater reliability of all measures excluding fluency¹ and lexical complexity, which were calculated by the KWIC Concordance software (see 5.1), and correlation coefficients of greater than 90% were achieved for all four measures.

6.2. Macro-measures

To measure the organisational features of a text, two data-driven writing scales were developed: one for the general organisational features and the other for the problem-solution structural elements. The general scale focused on two elements: the relevance of the ideas to the topic and the division of the text into well-organised paragraphs. The problem-solution structure scale focused on the presence of the problem-solution structure elements (i.e. situation, problem, solution, and evaluation) and their order as well as the degree to which the elements are developed in the text. The scales and the guidelines for the raters are presented in Tables 6.1 and 6.2, 6.3, and 6.4.

Table 6.1 *General Writing Scale*

Score	Idea Expression	Paragraph Division and Focus
4	All of the ideas expressed throughout the text are related to the topic; clear, and concise.	The text is appropriately divided into different paragraphs. Each paragraph has a single focus.
3	Most of the ideas expressed throughout the text are related to the topic; clear, and concise.	The text is appropriately divided into different paragraphs. At most, only one paragraph has more than one focus.
2	Some of the ideas expressed throughout the text are related to the topic; clear, and concise.	The text is appropriately divided into different paragraphs. Most paragraphs have more than one focus.
1	The ideas expressed throughout the text are neither related to the topic nor concise, or the text is too short.	Either the text is not divided into paragraphs or it is, but all paragraphs have more than one focus.

Table 6.2 *Guidelines for Employing the General Writing Scale*

Idea Expression	Paragraph Division and Focus
<p>Go through the text. Find the ideas that are expressed by the writer and check if they are relevant to the topic of the text and comprehensible. Be careful of the instances where the writer digresses and focuses on different issues other than the main topic of the text. Please note that you can assign a half score if you believe the text is somewhere between two full scores. If the ideas expressed are related to the topic, but the text is too short, score it 1.</p>	<p>First, consider paragraph division. That is, find out if it is clear where each paragraph starts and ends or if the writer mainly wrote a single long paragraph without dividing it into short and clear paragraphs. If paragraph division is achieved, then check if each paragraph focuses on a single idea and the explanation of it or it includes several ideas without any supporting information. Moreover, if a single main idea is explained in more than one paragraph, those paragraphs should be considered well-structured and receive a full score. However, if a main idea is discussed in a paragraph and then continued to a second paragraph which also focuses on another main idea, the first paragraph should receive a full score. The second paragraph, in contrast, receives no score as it focuses on two topics</p>

Table 6.3 *Problem-solution Structure Writing Scale*

Score	Presence and Order of the Problem-solution Structure Elements	Score	Development of the Problem-solution Structure Elements
4	All four elements of the problem-solution structure (situation, problem, solution, and evaluation) are present in the text and they are ordered according to the problem-solution structure	4	All of the elements are developed and expanded with appropriate details or examples
3.5	All four elements of the problem-solution structure are present in the text, but they are not ordered according to the problem-solution structure		
3	Only three elements of the problem-solution structure are present in the text and they are ordered according to the problem-solution structure	3	Three of the elements are developed with appropriate details or examples
2.5	Three elements of the problem-solution structure are present in the text, but they are not ordered according to the problem-solution structure		
2	Only two elements of the problem-solution structure are present in the text and they are ordered according to the problem-solution structure	2	Two of the elements are developed with appropriate details or examples
1.5	Two elements of the problem-solution structure are present in the text, but they are not ordered according to the problem-solution structure		
1	Only one or none of the elements of the problem-solution structure are present in the text	1	Only one or none of the elements are developed with appropriate details or examples

Table 6.4 *Guidelines for Employing the Problem-solution Structure Writing Scale*

Presence and order of the Elements of the Problem-solution Structure	Development of the Elements of the Problem-solution Structure
Go through the text and mark whenever you see one of the elements of the problem-solution structure (i.e. situation, problem, solution, and evaluation). Score the text based on the number of the elements you marked. That is, if you marked only a problem and a solution, score 2. Remember this is not the final score. After that, find out whether or not the existing elements are in the right order according to the problem-solution structure. In other words, if the problem is followed by the solution, score 2, but if the solution precedes the problem, score 1.5.	Go through the text and mark whenever you see one of the elements of the problem-solution structure. Then take each element in isolation and see how it is explained and developed. In other words, if there is ample explanation or examples through which the element is explicated, consider it as a fully developed element. However, if the element is mainly expressed through a single sentence or a very brief explanation making it very difficult to understand, consider it to be an inadequately developed element. If you believe an element is neither fully nor briefly developed, you can assign a half score to it.

After the scales were developed (see 4.6.2 for more details), pre- and post-writing performances were rated according to both scales and by two raters (i.e. the researcher and a native English teacher and writer). To facilitate rating, the researcher and the second rater met and rated the texts together, so that they could confer if they did not agree on the scores. When the two raters' scores were different only by one score (i.e. one rater agreed on 3 and the other on 4) the mean of the two scores were assigned (i.e. 3.5). However, when the difference between the raters' scores was 2 or larger, the raters conferred until they reached agreement.

6.3. Statistical Analyses

Scores were fed into SPSS (version 22) to carry out the statistical tests at the 95% confidence interval level across the four groups and between pre- and post-writing performances employing 2 x 4 mixed ANOVA tests. To this end, first, the absolute z-scores associated with the pre- and the post-writing scores were calculated and any scores associated with the z-scores

larger than 2.5 were removed as outliers (Field, 2009). The scores in every group and at both times were then tested for their distribution employing Kolmogorov-Smirnov tests. Since the scores for mechanical accuracy, fluency 2 (i.e. total number of words per text), idea expression, paragraph division and focus, presence and order of problem-solution elements, and development of problem-solution elements were not normally distributed (Kolmogorov_Smirnov $p < .05$), the base 10 logarithm, the square root, and the reciprocal (Field, 2009) compute functions in SPSS were used to transform the scores. However, none of these functions were successful in correcting the non-normality of the scores and the results of the Kolmogorov-Smirnov tests were significant for all converted sets of scores ($p < .05$). Therefore, since the normality assumption of the mixed ANOVA test was not met, statistical analyses were performed based on time (i.e. within-group) and group (i.e. between-group) differences by independently employing paired-samples t-tests (within-group differences for normal scores), Wilcoxon Signed Ranks tests (within-group differences for non-normal scores), one-way ANOVAs (between-group differences for normal scores), and Kruskal-Wallis H tests (between-group differences for non-normal scores). Mixed ANOVA tests were, however, run and their results are presented in Appendix I. Effect sizes were calculated for both significant and non-significant results and were interpreted adopting Plonsky and Oswald's (2014) benchmarks for within-group and between-group comparisons. Plonsky and Oswald argue that their benchmarks are discipline-specific and allow better interpretation of d in second language research.

6.4. Within-group Differences

6.4.1. Structured Group

Table 6.5 presents the results for both descriptive and inferential statistics for the structured group's pre- and post-writing performances. The structured group improved in structural complexity, lexical complexity, fluency 2, idea expression, and presence and order of the problem-solution elements. However, it declined in terms of grammatical accuracy, fluency 1, paragraph division and focus, and development of the problem-solution elements. Mechanical accuracy remained stable from pre- to post-writing.

Inferential statistical results indicate that the mean differences from pre- to post-writing were significant with regard to idea expression, paragraph division and focus, and presence and order of problem-solution elements ($p < .05$); that is, there were significant improvements from pre-

to post-writing with regard to idea expression and presence and order of the problem-solution elements while paragraph division and focus decreased considerably. Effect sizes were small for idea expression ($d = .73$), medium for paragraph division and focus ($d = 1.10$), and negligible for presence of the problem-solution elements ($d = .47$).

Effect sizes were also calculated for the non-significant results for better understanding of the effects of the structured task repetition on different aspects of the performance (see Table 6.5). The resulting values show no effect for the treatment.

Table 6.5 *Descriptive and Inferential Statistics for the Structured Group's Pre-W – Post-W Performance*

Measure	N	Pre-writing		Post-writing		t/z	Sig.	d
		Mean	Std.	Mean	Std.			
SC	26	1.51	.22	1.56	.31	-.60	.55	.18
LC	27	7.75	.88	7.79	.83	-.33	.74	.04
GA	26	.33	.12	.28	.11	1.54	.13	.43
MA	27	.26	.17	.26	.14	-.01	.98	0
F1	25	12.39	2.67	12.07	1.62	-.01	.98	.14
F2	26	211.81	52.15	217.00	49.62	-.65	.51	.10
IE	27	3.46	.85	3.92	.26	-2.57	.01*	.73
PDF	27	3.12	.75	2.11	1.05	-3.48	.00*	1.10
P&OoE	27	2.50	.75	2.87	.81	-2.03	.04*	.47
DoE	27	1.61	.67	1.53	.60	-.50	.61	.12

* Significance reached

SC = Structural Complexity; LC = Lexical Complexity; GA = Grammatical Accuracy; MA = Mechanical Accuracy; F1 = Fluency 1; F2 = Fluency 2; IE = Idea Expression; PDF = Paragraph Division and Focus; P&OoE = Presence and Order of Problem-solution Structure Elements; DoE = Development of Problem-Solution Structure Elements

6.4.2. Structured + Reformulation Group

Table 6.6 reports the results for both descriptive and inferential statistics for the structured + reformulation group's pre- and post-writing performances. The structured + reformulation group improved in lexical complexity, fluency 2, idea expression, presence and order of the problem-solution elements, and development of the problem-solution elements. However, there was a decrease in structural complexity, grammatical accuracy, fluency 1, and paragraph division and focus. Mechanical accuracy did not change from pre- to post-writing.

Table 6.6 *Descriptive and Inferential Statistics for the Structured + Reformulation Group's Pre-W – Post-W Performance*

Measure	N	Pre-writing		Post-writing		t/z	Sig.	d
		Mean	Std.	Mean	Std.			
SC	28	1.71	.36	1.56	.23	2.12	.04*	.50
LC	27	7.50	.85	7.73	.79	-1.51	.14	.28
GA	28	.32	.11	.31	.12	.49	.62	.08
MA	26	.30	.15	.30	.15	-.17	.86	0
F1	26	13.46	3.33	12.86	2.11	-1.13	.25	.27
F2	26	205.23	54.01	243.69	64.12	-3.11	.00*	.64
IE	28	3.71	.51	3.96	.13	-2.35	.02*	.67
PDF	28	2.87	.99	2.35	1.01	-1.98	.05*	.51
P&OoE	28	2.89	.89	3.01	.61	-.79	.42	.15
DoE	28	1.58	.66	1.87	.71	-1.60	.11	.42

* Significance reached

SC = Structural Complexity; LC = Lexical Complexity; GA = Grammatical Accuracy; MA = Mechanical Accuracy; F1 = Fluency 1; F2 = Fluency 2; IE = Idea Expression; PDF = Paragraph Division and Focus; P&OoE = Presence and Order of Problem-solution Structure Elements; DoE = Development of Problem-Solution Structure Elements

Results of the inferential statistics indicate that the mean differences from pre- to post-writing reached significance with regard to structural complexity, fluency 2, idea expression, and paragraph division and focus ($p < .05$). In other words, the structured + reformulation group improved significantly in terms of fluency 2 and idea expression, but declined in structural complexity and paragraph division and focus. Resulting d values were either small or showed no effect for the treatment: $d = .50, .64, .67$, and $.51$ for structural complexity, fluency 2, idea expression, and paragraph division and focus respectively.

Effect sizes for the non-significant results (see Table 6.6) show no effect for the treatment.

6.4.3. Unstructured Group

Table 6.7 presents the results for both descriptive and inferential statistics for the unstructured group's pre- and post-writing performances. The descriptive statistics show that the unstructured group increased in structural complexity, fluency 2, idea expression, presence and order of the problem-solution elements, and development of the problem-solution elements.

However, there was a decline in grammatical accuracy, mechanical accuracy, fluency 1, and paragraph division and focus. Lexical complexity did not change from pre- to post-writing.

Table 6.7 *Descriptive and Inferential Statistics for the Unstructured Group's Pre-W – Post-W Performance*

Measure	N	Pre-writing		Post-writing		t/z	Sig.	d
		Mean	Std.	Mean	Std.			
SC	26	1.53	.30	1.58	.23	-.84	.40	.18
LC	25	7.65	.88	7.65	.79	.04	.96	0
GA	27	.35	.14	.29	.11	2.38	.02*	.47
MA	27	.29	.15	.28	.16	-.16	.87	.06
F1	26	12.46	2.46	12.15	2.18	-.24	.80	.13
F2	24	223.92	49.86	231.12	55.39	-.72	.47	.13
IE	27	3.09	1.01	3.77	.57	-2.68	.01*	.82
PDF	27	2.74	.91	2.57	1.19	-.26	.79	.16
P&OoE	27	2.38	.90	3.24	.57	-3.63	.00*	1.14
DoE	27	1.38	.57	1.64	.70	-1.57	.11	.40

* Significance reached

SC = Structural Complexity; LC = Lexical Complexity; GA = Grammatical Accuracy; MA = Mechanical Accuracy; F1 = Fluency 1; F2 = Fluency 2; IE = Idea Expression; PDF = Paragraph Division and Focus; P&OoE = Presence and Order of Problem-solution Structure Elements; DoE = Development of Problem-Solution Structure Elements

Results of the inferential statistics indicate that the mean differences in this group were significant with regard to grammatical accuracy, idea expression, and presence and order of the problem-solution elements ($p < .05$); that is, idea expression increased significantly from pre- to post-writing while grammatical accuracy and presence and order of the problem-solution elements decreased. The reported d values were negligible for grammatical accuracy ($d = .47$), small for idea expression ($d = .82$), and medium for presence and order of the problem-solution elements ($d = 1.14$).

Effect sizes for non-significant results (see Table 6.7) show no effect for the treatment.

6.4.4. Control Group

Table 6.8 shows the results for both descriptive and inferential statistics for the control group's pre- and post-writing performances. The control group improved in idea expression, paragraph

division and focus, presence and order of the problem-solution elements, and development of the problem-solution elements, but decreased in lexical complexity, grammatical accuracy, mechanical accuracy, and fluency. Structural complexity did not change from pre- to post-writing.

Inferential statistical results indicate that the mean differences from pre- to post-writing were significant with regard to idea expression and development of the problem-solution elements ($p < .05$); that is, the control group improved significantly both in terms of idea expression and developments of the problem-solution elements from pre- to post-writing. The resulting d values show no effect for the treatment in terms of idea expression ($d = .56$) and is small with regard to development of the problem-solution elements ($d = .66$).

Effect sizes for non-significant results (see Table 6.8) show no effect for the treatment.

Table 6.8 *Descriptive and Inferential Statistics for the Control Group's Pre-W – Post-W Performance*

Measure	N	Pre-writing		Post-writing		t/z	Sig.	d
		Mean	Std.	Mean	Std.			
SC	22	1.61	.23	1.61	.26	-.02	.98	0
LC	24	7.53	.84	7.50	.80	.21	.83	.03
GA	23	.36	.17	.33	.09	.91	.37	.22
MA	24	.31	.12	.30	.13	-.25	.79	.07
F1	22	12.69	1.95	12.52	2.47	-.34	.73	.07
F2	23	215.35	63.06	203.86	49.37	1.10	.28	.20
IE	24	3.66	.70	3.95	.20	-2.12	.03*	.56
PDF	24	2.83	1.10	3.14	.99	-1.17	.24	.29
P&OoE	24	2.20	.76	2.33	.76	-.28	.77	.17
DoE	24	1.29	.44	1.62	.55	-2.81	.00*	.66

* Significance reached

SC = Structural Complexity; LC = Lexical Complexity; GA = Grammatical Accuracy; MA = Mechanical Accuracy; F1 = Fluency 1; F2 = Fluency 2; IE = Idea Expression; PDF = Paragraph Division and Focus; P&OoE = Presence and Order of Problem-solution Structure Elements; DoE = Development of Problem-Solution Structure Elements

6.4.5. Summary

Table 6.9 summarises the main results for within-group differences from pre- to post-writing performance for the three experimental groups and the control group. As the table shows, results are summarised in terms of significant results.

Table 6.9 *Summary of the Main Results for Pre- vs. Post-writing Performance*

Measure	Structured	Structured+ reformulation	Unstructured	Control
SC	-	Post-W<Pre-W ($d = .49$) *	-	-
LC	-	-	-	-
GA	-	-	Post-W<Pre-W ($d = .47$) *	-
MA	-	-	-	-
F 1	-	-	-	-
F 2	-	Post-W>Pre-W ($d = .64$) *	-	-
IE	Post-W>Pre-W ($d = .73$) *	Post-W>Pre-W ($d = .67$) *	Post-W>Pre-W ($d = .82$) *	Post-W>Pre-W ($d = .56$) *
PDF	Post-W<Pre-W ($d = 1.10$) *	Post-W<Pre-W ($d = .51$) *	-	-
P&OoE	Post-W>Pre-W ($d = .47$) *	-	Post-W>Pre-W ($d = 1.14$) *	-
DoE	-	-	-	Post-W>Pre-W ($d = .60$) *

* Significance reached

SC = Structural Complexity; LC = Lexical Complexity; GA = Grammatical Accuracy; MA = Mechanical Accuracy; F1 = Fluency 1; F2 = Fluency 2; IE = Idea Expression; PDF = Paragraph Division and Focus; P&OoE = Presence and Order of Problem-solution Structure Elements; DoE = Development of Problem-Solution Structure Elements

As Table 6.9 indicates, repeating the structured tasks did not have any notable effects on the micro measures, but significantly increased idea expression and presence and order of the problem-solution elements and led to a decrease in paragraph division and focus. The structured + reformulation group improved significantly with regard to fluency 2 and idea

expression but declined in structural complexity and paragraph division and focus. Repeating the unstructured tasks significantly increased idea expression and presence and order of elements but resulted in a decrease in grammatical accuracy. No notable changes were found in the performance of the control group with regard to micro measures. However, idea expression and development of the problem-solution elements increased.

6.5. Between-group Differences

Group differences were analysed in terms of gain/loss scores from pre- to post-writing. For this purpose, first, the gain/loss scores from pre-writing (pre-W) to post-writing (post-W) were computed for each measure. Then the absolute z-scores associated with the gain/loss scores were calculated and any scores associated with the z-scores larger than 2.5 were removed as outliers (Field, 2009). Then the scores for each measure were compared across the four groups employing one-way between groups ANOVA tests where scores were normally distributed (Kolmogorov-Smirnov $p > .05$) and homogeneous (Levene $p > .05$), and Kruskal Wallis H tests where scores were not normally distributed (Kolmogorov-Smirnov $p < .05$). Results of these tests as well as the effect sizes for both significant and non-significant differences are provided below.

6.5.1. Structural Complexity

The first variable investigated was structural complexity. After computing the gain/loss scores and removing the outliers, the number of participants was 27 in the structured group, 28 in the structured + reformulation group, 26 in the unstructured group, and 24 in the control group. The descriptive statistics, shown in Table 6.10, indicate that the structured group did not change in structural complexity from pre- to post-writing ($M = 0$). Both the structured + reformulation and the control group showed a decrease (str. + ref. $M = -.14$ and control $M = -.02$). However, the unstructured group showed an improvement from pre- to post-writing ($M = .04$).

Table 6.10 *Descriptive Statistics Results for Pre-W – Post-W Structural Complexity Gain/Loss Scores*

Group	N	Mean	Std.	Max value	Min Value
Str.	27	.00	.41	.89	-1.00
Str. + ref.	28	-.14	.36	.50	-1.04
Unstr.	26	.04	.29	.60	-.58
Control	24	-.02	.35	.64	-.85

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

The above mean differences were further examined employing a one-way between groups ANOVA test. Although the ANOVA test results indicate that the above mean differences across the groups did not reach significance; $F(3, 103) = 1.48, p > .05$, d values were calculated for group differences for better understanding of the effect of the treatments. The resulting values were .36 for the structured *vs.* the structured + reformulation group, .11 for the structured *vs.* the unstructured group, .05 for the structured *vs.* the control group, .55 for the structured + reformulation *vs.* the unstructured group, .33 for the structured + reformulation *vs.* the control group, and .18 for the unstructured *vs.* the control group. These d values indicate no effects for the treatments except for the structured + reformulation *vs.* the unstructured group comparison for which a small effect was found.

6.5.2. Lexical Complexity

The second variable investigated was lexical complexity. After computing the gain/loss scores and removing the outliers, the number of participants was 27 in the three experimental groups, and 24 in the control group. The descriptive statistics, presented in Table 6.11, indicate that the three experimental groups improved in lexical complexity from pre- to post-writing with the greatest improvement belonging to the structured + reformulation group ($M = .22$) and the smallest improvement belonging to the structured group ($M = .04$). The control group's lexical complexity, however, decreased from pre- to post-writing ($M = -.03$).

The mean differences were further examined performing a one-way between groups ANOVA test. The ANOVA test results indicate no significant differences among the groups in terms of the change in lexical complexity from pre- to post-writing; $F(3,102) = .55, p > .05$.

Table 6.11 *Descriptive Statistics Results for Pre-W – Post-W Lexical Complexity Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.04	.70	1.23	-1.30
Str. + ref.	27	.22	.76	2.05	-1.18
Unstr.	27	.13	.80	2.08	-1.86
Control	24	-.03	.75	1.13	-1.62

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Effect sizes were calculated for group differences and the resulting values were .24 for the structured *vs.* the structured + reformulation group, .11 for both the structured *vs.* the unstructured group and the structured + reformulation *vs.* the unstructured group, .09 for the

structured *vs.* the control group, .33 for the structured + reformulation *vs.* the control group, and .20 for the unstructured *vs.* the control group. The reported *d* values indicate no effects for the treatments.

6.5.3. Grammatical Accuracy

The third variable investigated was grammatical accuracy. After computing the gain/loss scores and removing the outliers, the number of participants was 27 in both the structured and the unstructured groups, 28 in the structured + reformulation group, and 24 in the control group. The descriptive statistics presented in Table 6.12, indicate that all four groups declined in grammatical accuracy from pre- to post-writing with the biggest decrease belonging to the unstructured group ($M = -.06$) and the smallest decrease belonging to the structured + reformulation and the control groups ($M = -.01$ for both groups).

Table 6.12 *Descriptive Statistics Results for Pre-W – Post-W Grammatical Accuracy Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	-.04	.15	.28	-.37
Str. + ref.	28	-.01	.14	.31	-.38
Unstr.	27	-.06	.13	.21	-.32
Control	24	-.01	.17	.34	-.32

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Mean differences among the four groups were further analysed employing a one-way between groups ANOVA test. Results of the ANOVA test indicate that groups were not significantly different from each other in terms of the grammatical accuracy gain/loss from pre- to post-writing; $F(3,103) = .65$, $p > .05$. The *d* values were .20 for the structured *vs.* the structured + reformulation group, .14 for the structured *vs.* the unstructured group, .18 for the structured *vs.* the control group, .37 for the structured + reformulation *vs.* the unstructured group, 0 for the structured + reformulation *vs.* the control group, and .33 for the unstructured *vs.* the control group. These *d* values indicate no effects for the treatments.

6.5.4. Mechanical Accuracy

Mechanical accuracy was the fourth variable investigated. After computing the gain/loss scores and removing the outliers, the number of participants was 27 in both the structured and the unstructured groups, 28 in the structured + reformulation group, and 23 in the control group.

As presented in Table 6.13, the descriptive statistics show that the structured group did not change in mechanical accuracy from pre- to post-writing ($M = .00$). The structured + reformulation group showed a slight increase ($M = .01$), whereas the unstructured and the control groups declined (Unstr. $M = -.01$, and Control $M = -.02$).

Table 6.13 *Descriptive Statistics Results for Pre-W – Post-W Mechanical Accuracy Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.00	.11	.23	-.19
Str. + ref.	28	.01	.19	.40	-.32
Unstr.	27	-.01	.17	.30	-.39
Control	23	-.02	.19	.28	-.36

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Mean differences among the four groups were further analysed employing a one-way between groups ANOVA test. Results of the ANOVA test indicate that the above mean differences were not statistically significant; $F(3, 102) = .17$, $p > .05$. The d values were .06 for the structured *vs.* the structured + reformulation group as well as the structured *vs.* the unstructured group, .12 for the structured *vs.* the control group, .11 for the structured + reformulation *vs.* the unstructured group, .15 for the structured + reformulation *vs.* the control group, and .05 for the unstructured *vs.* the control group. These effect sizes indicate no effects for the treatments.

6.5.5. Fluency 1 (total number of words per T-unit)

Fluency in terms of the total number of words per T-unit was the fifth variable investigated. After computing the gain/loss scores and removing the outliers, the number of participants was 27 in the structured group, 26 in both the structured + reformulation and the unstructured groups, and 24 in the control group. The descriptive statistics, presented in Table 6.14, indicate that the three experimental groups declined in fluency 1 from pre- to post-writing with the structured + reformulation group showing the largest decrease ($M = -.60$). However, the control group improved in fluency 1 ($M = .20$).

The mean differences were further examined performing a one-way between groups ANOVA test. The ANOVA test results indicate that the mean differences among the groups did not reach significance; $F(3, 100) = .33$, $P > .05$. The d values were .18 for the structured *vs.* the structured + reformulation group, .08 for the structured *vs.* the unstructured group as well as the structured *vs.* the control group, .10 for the structured + reformulation *vs.* the unstructured group, .27 for

the structured + reformulation vs. the control group, and .17 for the unstructured vs. the control group. These d values indicate no effects for the treatments.

Table 6.14 *Descriptive Statistics Results for Pre-W – Post-W Fluency 2 Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	-.06	3.16	6.08	-.8.60
Str. + ref.	26	-.60	2.77	4.43	-7.49
Unstr.	26	-.31	2.82	4.55	-7.64
Control	24	.20	3.09	5.57	-5.37

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

6.5.6. Fluency 2 (total number of words per text)

Fluency in terms of the total number of words per text was the last micro-measure investigated. After computing the gain/loss scores and removing the outliers, the number of participants was 27 in both the structured and the structured + reformulation groups, 26 in the unstructured groups, and 24 in the control group. The descriptive statistics, presented in Table 6.15, indicate that the three experimental groups improved in fluency 2 from pre- to post-writing with the structured + reformulation group showing the greatest improvement ($M = 29.00$). However, the control group declined in fluency 2 ($M = -13.58$).

Table 6.15 *Descriptive Statistics Results for Pre-W – Post-W Fluency 1 Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	10.51	48.40	149.00	-93.00
Str. + ref.	27	29.00	56.81	154.00	-112.00
Unstr.	26	7.84	47.70	97.00	-96.00
Control	24	-13.58	49.91	97.00	-125.00

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

The above mean differences were analysed employing a one-way between groups ANOVA test, and since the results of the ANOVA test revealed a significant difference across the groups; $F(3, 101) = 2.97$, $p < .05$, Bonferroni post hoc tests were carried out to find out exactly where the differences lay. Results, presented in Table 6.16, indicate that only the difference between the structured + reformulation and the control group reached significance ($p < .05$, $d = .79$).

Table 6.16 *Bonferroni Post Hoc Test Results for Fluency 1 Gain/Loss Scores*

Group	Group	Mean Difference	Std. Error	Sig.
Str.	Str. + ref.	-18.48	13.85	1.00
Str.	Unstr.	2.67	13.98	1.00
Str.	Control	24.10	14.27	.56
Str. + ref.	Unstr.	21.15	13.98	.80
Str. + ref.	Control	42.58	14.27	.02*
Unsr.	Control	21.42	14.40	.84

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

Effects sizes were also calculated for non-significant results and the d values were .35 for the structured *vs.* the structured + reformulation group, .05 for the structured *vs.* the unstructured group, .49 for the structured *vs.* the control group, .40 for the structured + reformulation *vs.* the unstructured group, and .43 for the unstructured *vs.* the control group.

In summary, the results indicate that the structured + reformulation group significantly outperformed the control group with a medium effect size. No significant differences were found among the other groups and the reported d values show no effect for the treatments.

6.5.7. Idea Expression

The conciseness and relevance of the expressed ideas to the topic of the task was the first macro-measure investigated. After the gain/loss scores were computed and the outliers were removed, the number of participants was 27 in the structured group, 28 in the structured + reformulation group, 24 in the unstructured group, and 23 in the control group. The descriptive statistics, presented in Table 6.17, show that all four groups improved in idea expression from pre- to post-writing with the unstructured group showing the largest increase ($M = .60$) and the control group showing the smallest increase ($M = .17$).

Table 6.17 *Descriptive Statistics Results for Pre-W – Post-W Idea Expression Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	.46	.80	2.00	-1.00
Str. + ref.	28	.25	.55	1.50	-.50
Unstr.	24	.60	.73	2.00	.00
Control	23	.17	.38	1.00	.00

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Mean differences among the four groups were analysed employing a Kruskal Wallis H test as the scores were not normally distributed within the groups (Kolmogorov-Smirnov $p > .05$). Kruskal Wallis H test results indicate that the above mean differences were not significant across the groups; $\chi^2(3, 99) = 5.91, p > .05$. The d values were .30 for the structured *vs.* the structured + reformulation group, .18 for the structured *vs.* the unstructured group, .46 for the structured *vs.* the control group, .54 for the structured + reformulation *vs.* the unstructured group, .16 for the structured + reformulation *vs.* the control group, and .73 for the unstructured *vs.* the control group.

Although the above results show no significant differences among the groups in terms of idea expression, the effect sizes indicate small effects for the structured *vs.* the control group and the structured + reformulation *vs.* the unstructured group comparisons. Medium effects were also found for the unstructured *vs.* the control group comparison.

6.5.8. Paragraph Division and Focus

The second macro-measure investigated was the division of the text into well-organised paragraphs. After the gain/loss scores were fed into SPSS, no outliers were detected as the absolute z-scores associated with the gain/loss scores were all smaller than 2.5 (Field, 2009). Therefore, the number of participants was 27 in both the structured and the unstructured groups, 28 in the structured + reformulation group, and 24 in the control group. The descriptive statistics in Table 6.18 show that the three experimental groups declined in paragraph division and focus from pre- to post-writing while the control group improved ($M = .31$). Among the experimental groups, the largest decrease belonged to the structured group ($M = -1.01$), whereas the smallest decrease was shown by the unstructured group ($M = -.16$).

Table 6.18 *Descriptive Statistics Results for Pre-W – Post-W Paragraph Division and Focus Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	-1.01	1.16	1.00	-3.00
Str. + ref.	28	-.51	1.40	2.00	-3.00
Unstr.	27	-.16	1.49	2.00	-3.00
Control	24	.31	1.56	3.00	-3.00

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Mean differences among the four groups were examined employing a Kruskal Wallis H test as the scores were not normally distributed within the groups (Kolmogorov-Smirnov $p > .05$).

Results of the Kruskal Wallis H test revealed significant differences among the groups in terms of the change in paragraph division and focus from pre- to post-writing; $\chi^2(3, 103) = 12.18$, $p < .05$. Therefore, Kruskal Wallis H post hoc analyses were performed to find out where the differences lay. Results, presented in Table 6.19, indicate a significant difference for the structured *vs.* both the unstructured and the control groups ($p < .05$, $d = .63$, and $.96$ respectively) and for the structured + reformulation *vs.* the control group ($p < .05$, $d = .55$).

Table 6.19 *Kruskal Wallis H Post Hoc Test Results for Paragraph Division and Focus*

Group	Group	Chi-Square	df	Sig.
Str.	Str. + ref.	1.78	1	.18
Str.	Unstr.	5.51	1	.02*
Str.	Control	10.52	1	.00*
Str. + ref.	Unstr.	1.24	1	.26
Str. + ref.	Control	4.13	1	.04*
Unsr.	Control	1.31	1	.25

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

The d values for non-significant results were $.38$ for the structured *vs.* the structured + reformulation group, $.24$ for the structured + reformulation *vs.* the unstructured group, and $.30$ for the unstructured *vs.* the control group.

In conclusion, results indicate that the structured group significantly declined in terms of paragraph division and focus compared to the unstructured and the control groups. The effect sizes for these differences were small and medium, respectively. The control group also significantly outperformed the structured + reformulation group with a medium-sized effect. The reported d values for the non-significant results show no effects for the treatments.

6.5.9. Presence and Order of the Problem-solution Structure Elements

The third macro-measure investigated was the presence and order of the four elements of the problem-solution structure, namely, situation, problem, solution, and evaluation. After the outliers were detected and removed, the number of participants was 26 in the structured group, 28 in the structured + reformulation group, 27 in the unstructured group, and 24 in the control group. The descriptive statistics, presented in Table 6.20, indicate that all four groups improved from pre- to post-writing with the unstructured group showing the largest increase ($M = .85$)

and the structured + reformulation as well as the control groups showing the smallest improvement ($M = .12$ for both groups).

Table 6.20 *Descriptive Statistics Results for Pre-W – Post-W Presence and Order of Elements Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	26	.50	.87	2.00	-1.00
Str. + ref.	28	.12	.89	2.00	-1.50
Unstr.	27	.85	.96	2.50	-1.00
Control	24	.12	1.20	2.50	-1.50

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Results of the Kruskal Wallis H test carried out to investigate the mean differences revealed significant differences among the groups in terms of the change in the presence and order of the problem-solution structure elements from pre- to post-writing; $\chi^2(3, 102) = 8.94$, $p < .05$. Thus, Kruskal Wallis H post hoc analyses were performed to find out where the differences lay. Results of the post hoc analyses, presented in Table 6.21, indicate a significant difference for the unstructured group *vs.* both the structured + reformulation and the control groups ($p < .05$, $d = .78$, and $.67$ respectively).

The d values for the non-significant results were $.43$ for the structured *vs.* the structured + reformulation group, $.38$ for the structured *vs.* the unstructured group, $.36$ for the structured *vs.* the control group, and 0 for the structured + reformulation *vs.* the control group.

Table 6.21 *Kruskal Wallis H Post Hoc Test Results for Presence and Order of Elements*

Group	Group	Chi-Square	df	Sig.
Str.	Str. + ref.	2.35	1	.12
Str.	Unstr.	1.53	1	.21
Str.	Control	1.80	1	.17
Str. + ref.	Unstr.	6.72	1	.01*
Str. + ref.	Control	.00	1	.97
Unsr.	Control	5.30	1	.02*

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

In summary, the unstructured group significantly outperformed the structured + reformulation and the control groups with medium and small effects, respectively. With regard to the effect

sizes for non-significant results, the structured group outperformed the structured + reformulation group with a small effect. In other cases, no differential effect for the treatments was found.

6.5.10. Development of the Problem-Solution Structure Elements

The last macro-measure investigated was the degree to which the elements of the problem-solution structure were developed in the text. After the outliers were removed, the number of participants was 27 in both the structured and the structured + reformulation groups, 26 in the unstructured group, and 24 in the control group. The descriptive statistics, presented in Table 6.22, indicate that the structured group declined from pre- to post-writing ($M = -.07$), whereas the other three groups showed improvement with the greatest increase belonging to the structured + reformulation group ($M = .37$).

Table 6.22 *Descriptive Statistics Results for Pre-W – Post-W Development of Elements Gain/Loss Scores*

Group	N	Mean	Std.	Max Value	Min Value
Str.	27	-.07	.89	2.00	-1.50
Str. + ref.	27	.37	.88	2.00	-1.00
Unstr.	26	.34	.74	2.00	-1.00
Control	24	.33	.50	1.00	-.50

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

A Kruskal Wallis H test was carried out to compare the mean differences. Results of this test indicate that no significant differences existed across the groups in terms of the change in the development of the elements of the problem-solution structure from pre- to post-writing; $\chi^2(3, 101) = 4.96, p > .05$.

The d values were .50 for both the structured vs. the structured + reformulation group and the structured vs. the unstructured group, .55 for the structured vs. the control group, .03 for the structured + reformulation vs. the unstructured group, .05 for the structured + reformulation vs. the control group, and .01 for the unstructured vs. the control group.

The reported d values indicate small effects for the structured group's decline in comparison to the other three groups. In other cases, no differential effect for the treatments was found.

6.5.11. Summary

Tables 6.23 and 6.24 summarise the main results for micro- and macro-measures in terms of significant results and/or effect sizes greater than .39.

Table 6.23 *Summary of the Main Results for Pre-W – Post-W Gain/Loss*

Measure	Comparisons
Structural Complexity	Unstr. > Str. + ref. ($d = .55$)
Lexical Complexity	-
Grammatical Accuracy	-
Mechanical Accuracy	-
Fluency 1	-
Fluency 2	Str. + ref. > Control* ($d = .79$) Str. > Control ($d = .49$) Str. + ref. > Unsr. ($d = .40$) Unstr. > Control ($d = .43$)

* Significance reached

As presented in Table 6.23, no notable differential effects were found for the treatments when micro measures were considered among the groups. The only significant difference was between the structured + reformulation and the control groups for fluency 2 (i.e. total number of words per text).

Table 6.24 *Summary of the Main Results for Pre-W – Post-W Gain/Loss with*

Measure	Comparisons
Idea Expression	Unstr. > Str. + ref. ($d = .54$) Str. > Control ($d = .46$) Unstr. > Control ($d = .73$)
Paragraph Division and Focus	Unstr. > Str. *($d = .63$) Control > Str. + ref.*($d = .55$) Control > Str. *($d = .96$)
Presence and Order of Elements	Str. > Str. + ref. ($d = .43$) Unstr. > Control* ($d = .67$) Control > Str. *($d = .96$)
Development of Elements	Str. + ref. > Str. ($d = .50$) Unstr. > Str. ($d = .50$) Control > Str. ($d = .55$)

* Significance reached

According to Table 6.24, group differences with regard to macro-measures were mostly in favour of the unstructured and the control groups. The unstructured group significantly outperformed the structured group in terms of paragraph division and focus and significantly improved in comparison to both the structured + reformulation and the control groups in terms of presence and order of the problem-solution structure elements. This group also improved more than the structured + reformulation and the control groups with regard to idea expression, and more than the structured group in terms of development of the problem-solution structure elements. These results were, however, not significant.

The control group significantly gained more in paragraph division and focus in comparison to both the structured and the structured + reformulation groups. This group also outperformed the structured group with regard to development of the problem-solution structure elements.

6.6. Discussion of the Results of Research Question Three

Research question three investigated the effects of (a) a structured written task repetition, (b) a structured written task repetition + reformulation, and (c) an unstructured written task repetition on the performance of a new task of a different type (i.e. free writing) in terms of both micro-measures (i.e. structural complexity, lexical complexity, grammatical accuracy, mechanical accuracy, and fluency) and macro-measures (i.e. idea expression, paragraph division and focus,

presence and order of elements, and development of elements). Results were investigated in terms of both within-group differences from pre- to post-writing and between-group differences with regard to the amount of gain/loss from pre- to post-writing.

Considering significant differences and effect sizes, results indicate that both within and among the groups, the different treatments did not have a strong positive effect on either the micro or macro aspects of the performance of a new task of a different type (i.e. free writing). However, a few statistical differences were found which were mostly indicative of significant losses or were in favour of the control group. Almost all of the effect sizes were negligible, small or medium.

Regarding within-group differences, findings indicate that the statistical differences were mainly found in terms of macro measures, whereas very little change occurred with regard to micro measures. With regard to micro measures, grammatical accuracy decreased from pre- to post-writing within the unstructured group. Within the structured + reformulation group, structural complexity dropped significantly, while fluency 2 (i.e. total number of words per text) showed a significant increase. Considering macro-measures, all four groups improved significantly from pre- to post-writing in terms of idea expression. Both the structured and the structured + reformulation groups declined significantly from pre- to post-writing with regard to paragraph division and focus. Both the structured and the unstructured groups increased from pre- to post-writing with regard to presence and order of the problem-solution structure elements. Finally, considering the development of the elements, the only significant increase was in the control group.

With regard to between-group differences, the structured + reformulation group significantly outperformed the control group in terms of fluency 2 (i.e. total number of words per text) but decreased in structural complexity in comparison to the unstructured group. In terms of macro measures, in general, the unstructured and the control groups performed better than the other groups; that is, the unstructured group outperformed the structured + reformulation and the control groups in terms of idea expression, performed significantly better than the structured group in terms of paragraph division and focus, showed significantly greater gains than the structured + reformulation and the control groups with regard to presence and order of the problem-solution structure elements, and gained more than the structured group with regard to the development of the problem-solution structure elements. The control group significantly outperformed the structured and the structured + reformulation groups in terms of paragraph

division and focus and performed better than the structured group with regard to development of the problem-solution structure elements.

In the following sections, the above findings will be discussed providing some examples from students' texts where needed to flesh out the discussion. The discussion will first focus on the overall results of research question three. Then the within-group differences will be discussed in terms of both micro- and macro-measures. Afterwards, the results of between-group differences will be discussed with regard to micro- and macro-measures. Before I start discussing the findings, it is worth mentioning that the pre- and post-writing tasks were counter-balanced which removes the potential task effect on the findings.

6.6.1. Overall Results

As mentioned above, the results provided little evidence of any effect for the experimental treatments both within and among the groups. This finding in fact replicates the findings of the oral task repetition studies which failed to show any effects or found negligible effects on the performance of a new task (Bygate, 2001; Gass et al. 1999; Kim & Tracy-Ventura, 2013; Patansorn, 2010). Therefore, as Ellis (2009b) also suggested, for task repetition effects to be transferred to a new task, some type of intervention might be required between the repeated performances. However, since the intervention employed in this study (i.e. reformulation as a form of written feedback) was not beneficial either, it can be speculated that a more explicit intervention where the linguistic features as well as the textual structure is directly explained with some examples is needed for the task repetition to have any effect on learners' writing proficiency. Previous research indicates that explicit teaching of linguistic features and textual structure benefits written performance both in terms of micro (Yasuda, 2011) and macro aspects (Bacha, 2010; Henry & Roseberry, 1998; Khodabandeh, 2014; Yasuda, 2011). For example, during a fifteen-week task-based email writing course, Yasuda (2011) found that through performing email-writing tasks and receiving explicit instruction on email-writing genre, students improved significantly with regard to task fulfilment and appropriacy, cohesion and organisation, grammatical control, fluency and language sophistication.

However, it should be noted that all the above task repetition studies are on oral task performance whereas in this study task repetition effects were investigated on written performance. In a recent study on the effects of written task repetition on language performance, Nitta and Baba (2014) found limited effects of task repetition on the performance of the same task, but marked effects on the lexical and grammatical aspects of the performance of a new task of the same type; that

is, completely opposite to the findings of the task repetition studies on oral performance. These findings are in line with the findings of research questions 1 and 2 (see chapter 5) as no effect was found for task repetition on the performance of the same task, but positive effects were reported for structured task repetition on both lexical and grammatical aspects of the performance of a new task of the same type which suggested that with more repetitions, particularly structured task repetitions, over an extended period of time, the performance of a new task is likely to improve. Nitta and Baba's study supports this conclusion, as their study is a semi-longitudinal study over a one-year period and suggests that task repetition effects may not show up in the short term but can lead to positive effects over a course of time. It can therefore be concluded that as also Bygate (2001) suggested, if learners (especially the learners in this study who were unfamiliar with task-based language teaching) have more repetition opportunities over a long period of time and also receive some kind of explicit intervention between repeated performances, it is likely that they can also improve their performance of a new task of a different type.

6.6.2. Within-group Differences

With regard to within-group differences and considering micro-measures, results show that the grammatical accuracy of the unstructured group declined from pre- to post-writing. The explanation for this might be that performing unstructured tasks where paragraphs are not logically ordered is demanding on working memory, so attention is not available to be allocated to improve micro aspects of the performance. The findings of the previous chapter where the structured group outperformed the unstructured group in terms of grammatical accuracy of the performance of a new task of the same type also support this explanation. Other task structure studies (Skehan & Foster, 1999; Tavakoli & Skehan, 2005; Tavakoli & Foster, 2008; Tavakoli, 2009) also showed positive effects for performing structured tasks on grammatical accuracy, but no effect for performing unstructured tasks. It should be noted that no studies have investigated the effect of task structure on written performance and all of the studies referred to above are on oral performance. Although the present findings suggest that written performance is affected in the same way by task structure as the oral performance is, it should be noted that these findings are obtained by repeating the structured and the unstructured tasks, whereas the oral performance studies found positive effects of task structure on a single performance.

For the structured + reformulation group, there was a positive effect on fluency in terms of total number of words per text but a negative effect on structural complexity. Trade-offs are evident

in accordance with Skehan's (1998b) limited attentional resources theory. In other words, learners could only attend to one aspect of performance (i.e. fluency) having no attentional resources free to improve the complexity or accuracy of their performance.

Regarding macro measures, there were negative effects for paragraph division and focus within the structured and the structured + reformulation groups. The explanation for this can be found in the students' educational background and the way Farsi composition writing is taught in Iranian educational settings. That is, in Farsi composition writing, the main focus is on teaching sentence structure and almost no instruction is offered in terms of idea organisation and paragraph development (Skype interview with Mr. Hossein Mehrang, an experienced composition teacher in Iran, 25th May, 2015). It can thus be concluded that since learners had no background knowledge about how to organise their ideas into paragraphs and as mentioned above, no explicit teaching of textual structure was offered to them in this study, they not only failed to benefit from the treatments of the study with regard to idea organisation and paragraph development but also developed even less organised paragraphs at post-writing. In other words, mere exposure to the treatment tasks without any follow-up awareness raising activities or class discussions might have confused learners.

There is evidence that both the structured and the unstructured groups benefited from repeating the treatment tasks. They manifested a more effective use of the problem-solution structure in terms of presence and order of the elements in the post-writing task. This is not surprising for the structured group as it was expected that the nature of the structured tasks would assist learners in figuring out what elements should be included in a problem-solution text and how these elements should be ordered. However, this is clearly not the case when the unstructured group is considered. The explanation for this might be that performing unstructured tasks where one paragraph does not lead to the other may encourage the students to figure out how to impose a structure on the disordered elements in the unstructured task. Students therefore allocate attention to macro aspects of their writing and produce better organised texts.

For example, learner K from the structured group wrote the two texts below in the pre- and post-writing tasks:

Pre-writing Performance

In our society generation gap is one of the important problems between parents and children especially teenagers. Sometimes it's because of technology.

In the past, for example parents didn't have computer or cellphones and they play outside or with other children in their same age but now when the children play a lot of game in front of computers parent's doesn't let them because they think it is harmful device for children health especially for their eyes.

These days teenagers want to be more independent and don't want their parents limited them in anything but parents don't accept such a thing because they can't understand sometimes their children.

In past days people weren't knowledgeable enough and it was not common to read books, magazine even there were not useful programs on TV but these days people can get information about everything they like from many sources and they can answer easily to their childrens questions.

We can see many differences in Generation and continually we will have them in future.

Post-writing Performance

Nowadays we have more unemployment people in our society especially when they graduated from university they can't find proper job.

These days most of the people don't have good income or sometimes they have but they don't like their job. They can't reach their goals because of these problem. Because of unemployment we faced with big problems such as rubbing. Also the young people can't get married or have their own apartment because expenses are high.

One solution can be create more job for young people but I think it can be inefficient because educated people want professional jobs related to their majors and also sometimes we see middle aged or old people don't have good jobs and good income.

Another useful solution can be the government should listed the graduations number of people and also their major and create useful and professional jobs for them. I think this solution can help the society eliminate the unemployment.

In the pre-writing text, learner K wrote only about the two elements of the problem-solution structure (i.e. situation [modern technology and more knowledgeable people] and problem [conflicts between parents and children]), whereas in the post-writing text, the learner starts with

explaining the situation (low-paid jobs, and lack of job satisfaction) dominant in the society and the problems that result from it (theft, marriage and buying houses being impossible). He/she then suggests a solution (creating more jobs) and positively evaluates it (unemployment can be eliminated). It is therefore clear that the learner developed understanding of the elements to include in a problem-solution text and the way to order them.

In another example, learner L from the unstructured group wrote the following two texts in the pre- and post-writing tasks:

Pre-writing Performance

In a society like ours, unemployment is a major problem for all of the people and mostly the young ones who graduated from university. I think that the main goal of going to university is to have a related job with your major and do something you like in your job. So unemployment after graduation can be an annoying problem.

Employment is a process that many organisations are involved with it. One of these organisations is the university. Universities must know the various needs in the society and based on these needs accept students in the related majors. For example, in a country that needs a lot of nurses universities should expand nursing major in order to have more nurses in the future.

In the other hand the high school student must be very careful in the time of choosing a university major for themselves. First they have to be aware of job conditions in the society and, then choose a major which could offer them a suitable profession in the future. Unfortunately, nowadays most of the students choose their major based on very low information and then in the future regret their choice.

Overall there are many other factors which affect the employment but in my point of view those two ones that write about are the most important. So first the government and the universities and then the students themselves should try to solve the unemployment issue.

Post-writing Performance

Generation gap is known as one of the problems in our society. It most cause problems between parents and children because they belong to different generations. The first question is which factors cause generation gap? In my point of view development of technology is one of the important reasons. To be honest the speed of this development was very high that past generations couldn't keep up with it. What's more in the way of living in the past was a lot different with today's lifestyles.

One of the problems that generation gap cause is the two old and new generation can't understand each other sometimes and it causes disagreements and arguments. Also sometimes two generations have strenge expectations from each other. In order to prevent these kinds of problems either of generations should get aware about the other generation. The past generation's people must be more updated about today's technology, youthes problems and challenges and the new generation had to learn about their parents and grand parents beliefes and lifestyle in the past. I think both of the generations should be more openminded.

If young and old people know each other better ad respect each other's values and beliefs they would fill the gap between generations and therefor they could tolerant opposite ideas better. If this gap fills old and young people will do best. I think diffrent generations can complete each other provided that they try to learn each other any good point.

In the pre-writing performance, learner L mainly focused on giving solutions to the problem of unemployment (students to choose their majors carefully and universities to offer what is needed in the society). The learner also very briefly points to the situation dominant in society (nowadays most of the students choose their major based on very low information and then in the future become completly regretful). The situation and the solution elements are not in the correct order. In the post-writing performance, however, learner L makes clear gains and includes all four elements of the problem-solution structure in his/her writing. The learner starts with explaining the situation (technology development, different lifestyles from past to present) which leads to the problem (arguments and conflicts between parents and adults). He/she then proposes a solution (different generations should learn about their differences) and evaluates it positively (gaps can be filled, generations can tolerate opposite ideas, both old and young people can perform better). All four elements are also correctly ordered.

Finally, there was some evidence that the control group improved in terms of the development of the individual problem-solution structure elements. Since the control group did not change with regard to presence and order of the elements, it can be concluded that due to lack of treatment, this group did not develop any understanding of the elements to include in a problem-solution text and so focused on developing one or two elements of the problem-solution structure that they had included in their texts. For example, learner M from the control group wrote the following two texts in the pre- and post-writing tasks:

Pre-writing Performance

The unemployment is rising in our society. It is a calamity for young people who are nearly graduated from university and that is the reason of affliction and deviation of young people. It is difficult to find a job for young people because of the situation in our country these days.

The unemployment causes to many catastrophe is the society like smuggling which is a serious problem for country in financial case.

Like every other major problem of country, unemployment has different kinds too and this variation makes it hard to solve this problem. The most dangerous kind of unemployment is the fake one in which the unemployed person has a job but this job doesn't have any positive results for society like cigar selling which is the current kind of unemployment in big cities and also the major problem and dilemma that politicians should solve it.

In conclusion, unemployment causes hard hurts to the society. If politicians don't solve it, many catastrophe may wait for us in the future. Like insecurity. But politicians can't do many things lonely. Young people should help the too.

Post-writing Performance

Generation gap is usually seen between young adults and parents which causes to many problems and slopes in our society and the reason of many of rubbing and stealing scandals.

The problem starts with teenager when he or she wants to have another life with her own rules and a life than teenager plans about. But on the other side, parents don't really want it to happen. They want a child who obeys all of the orders that parents give to them and they want that their children have a life in a style which their parents want.

The view of point of parents which is mentioned on paragraph below, makes the life to young adults person. So young adults like every one wants to escape from prison but he doesn't know that this action will cause his death.

Solution of this problem is in both parents and young adult hands. Parents should accept adults life style and they should respect the decisions of their young adults and young adults should let their parents to help them too. They should know that the advices that their parents give them are only for their successes.

In my opinion, if we can solve the generation gap, maybe most of the problems in our country will solve and we will get a nice life.

In the pre-test performance, the learner included only two elements of the problem-solution structure (i.e. situation [the situation in our country makes it difficult for young people to find jobs], and problem [the rest of the text is about unemployment and its consequences]). It is clear that the element of situation is not developed as no explanation is given about the situation dominant in the country and the reason why it results in unemployment. In the post-test performance, the learner included two elements of problem (first two paragraphs) and solution (paragraph 4) and developed both to an acceptable degree. It is clear from these two texts that in the control group, learners were not aware of the elements that they should include in a problem-solution text and focused mostly on developing their ideas in their texts, probably to make sure they meet the word limit required in the task.

6.6.3. Between-group Differences

When group comparisons are considered, results show that the unstructured group outperformed the structured + reformulation group in terms of structural complexity, and the structured + reformulation group outperformed the control group with regard to fluency in terms of total number of words per text. These results reflect the gains and losses that occurred within the structured + reformulation group from pre- to post-writing. As discussed above, the structured + reformulation group declined with regard to structural complexity but increased in terms of fluency, indicating a trade-off effect (Skehan, 1998b). That is, according to Skehan's limited attentional resources model, this group could only attend to one aspect of performance and prioritised meaning (i.e. fluency) over form (i.e. complexity and accuracy). This group, therefore, outperformed the control group with regard to fluency and declined in structural complexity in comparison to the unstructured group.

Considering macro aspects, group comparisons revealed that the unstructured group manifested better control of the macro aspects in their post-writing performance even for the presence and order of the problem-solution structure elements. The explanation is that, as mentioned above, requiring students to perform tasks with disorganised paragraphs is more effective than giving them the structure, as unstructured tasks induce attention to the organisation and structure of the ideas even though the students have not been asked to order the paragraphs or structure their texts. As a result, it leads to the production of better organised and structured texts (see 6.6.2. for examples from data).

With regard to paragraph division and focus, the control group outperformed both the structured and the structured + reformulation groups. As mentioned above, within-group comparisons show

that the structured and the structured + reformulation groups both lost considerably from pre- to post-writing in terms of paragraph division and focus while the control group showed a small increase. Thus, the control group outperformed these two groups, which suggests that students tend to organise better paragraphs in the absence of any treatment than when they are given the structure and asked to make use of that structure without being explicitly taught.

In conclusion, findings indicate that in the absence of explicit teaching of the linguistic features and the textual structure or an explicit intervention, learners largely failed to benefit from task repetition or the experimental treatments to improve their post-writing texts. Lack of extensive task repetition is also believed to have affected the performance of a new task of a different type. Repeating structured tasks was found to affect written performance in the same way as task structure affects oral performance; that is, repeating structured tasks facilitated grammatical accuracy. Trade-off effects were evident in the structured + reformulation group performance and resulted in increased fluency but decreased structural complexity. Performing unstructured tasks led to more structurally organised texts even in terms of the presence and order of the problem-solution structure elements. Learners' educational background (no prior knowledge of textual structure) might have prevented them from benefiting from the treatments of the study, and as a result they developed messy paragraphs focusing on multiple ideas, while learners in the control group produced better organised paragraphs. The lack of any experimental intervention in the control group appears to have been beneficial for the development of the problem-solution structure elements as learners did not allocate attention to include the different elements, but developed their ideas in their texts.

Chapter 7 Conclusion

This final chapter provides a summary of the design and the main findings of the study followed by the presentation of its theoretical and pedagogical implications. The chapter concludes with outlining the limitations of the study and suggesting directions for future research.

7.1. Summary of Objectives and Design

The research reported in this study examined the effects of task structure as a task design variable, task repetition as a task implementation variable, and reformulation as an intervention variable on the written performance of English as a foreign language in two language schools in Iran. The purpose was to fill the gap in the field of task-based language teaching with regard to writing and to discover how written performance is affected by certain task design and task implementation variables.

The study adopted an experimental design (i.e. pre-writing, task repetition, post-writing) with a task repetition study woven into it. One-hundred and six participants were divided into three experimental groups (i.e. structured, structured + reformulation, unstructured) and a control group. The whole process of data collection took place over a five-week period: in week 1, all groups attended a pre-writing session where they performed a free-writing task; in week 2, the structured and the structured + reformulation groups completed a structured note-expansion task, whereas the unstructured group performed an unstructured note-expansion task; in week 3, the structured and the unstructured groups repeated the same note-expansion tasks as week 2, but the structured + reformulation group first received feedback on their week 2 performance (i.e. reformulation) and then performed the same task; in week 4, the structured and the structured + reformulation groups completed a new structured note-expansion task and the unstructured group performed a new unstructured note-expansion task; finally in week 5, all four groups engaged in a post-writing session where they completed a free-writing task. Pre- and post-writing sessions adopted a counter-balanced design to avoid the task effect. The experimental groups also completed a questionnaire at week 3 after they repeated the same task to elicit their accounts of how task repetition and/or reformulation were beneficial for them to improve their texts.

The design summarised above allowed the investigation of the three research questions of the study which aimed to discover the effects of task structure, task repetition, and reformulation

on the performance of the same task, a new task of the same type, and a new task of a different type. Pre- and post-writing data were analysed in terms of both micro-measures (i.e. complexity, accuracy, and fluency) and macro-measures (i.e. textual structure and organisation). Task repetition data (i.e. performance of note-expansion tasks) were analysed only with regard to micro-measures. The key findings for each research question are outlined in the next section.

7.2. Summary of the Key Findings

RQ1. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the repeat performance of the same task in terms of micro-measures?

- Repeating the same structured task as opposed to the unstructured task had no effects on the repeat performance.
- Reformulation significantly disadvantaged the fluency of the performance in the structured + reformulation group.

RQ2. What are the relative effects of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of a new task of the same type in terms of micro-measures?

- In comparison to the unstructured group, the structured group produced a more structurally complex and more grammatically and mechanically accurate performance of a new task of the same type.
- Reformulation had no notable effects on the performance of a new task of the same type.

RQ3. What is the effect of (a) structured written task repetition, (b) structured written task repetition + reformulation, and (c) unstructured written task repetition on the performance of a new task of a different type in terms of both micro- and macro-level measures?

Within-group differences with regard to micro-measures:

- Within the unstructured group, grammatical accuracy of the performance declined significantly over time.
- Within the structured + reformulation group, the structural complexity declined significantly over time, whereas fluency showed a significant increase.

Within-group differences with regard to macro-measures:

- All four groups progressed over time with regard to idea expression.
- Within the structured and the structured + reformulation groups, paragraph division and focus was significantly lower at post-writing than at pre-writing.
- Both the structured and the unstructured groups had better presentation and order of the problem-solution structure elements at post-writing than at pre-writing.
- The control group developed their ideas and the elements of the problem-solution structure better at post-writing than at pre-writing.

Between-group differences with regard to micro-measures:

- The structured + reformulation group produced a significantly more fluent performance in comparison to the control group.

Between-group differences with regard to macro-measures:

- The unstructured group showed better performance in terms of both idea expression as well as presence and order of problem-solution structure elements in comparison to the structured + reformulation and the control groups.
- The unstructured group gained more in both paragraph division and focus as well as development of the problem-solution structure elements than the structured group.
- The control group performed better than the structured and the structured + reformulation groups in terms of paragraph division and focus.
- The control group better developed their ideas and the problem-solution structure elements in their texts in comparison to the structured group.

7.3. Research Implications

7.3.1. Theoretical Implications

One major finding of this study is that repeating the structured tasks as opposed to the unstructured tasks results in improvements in the structural complexity, grammatical accuracy, and mechanical accuracy of the performance of a new task of the same type. This finding provides support to Skehan's (1998b) limited attentional resources model and trade-off hypothesis suggesting that repeating the structured tasks eases the processing burden of the task through task repetition on the one hand and task structure on the other. This in turn, sets

attentional resources free to be allocated to both complexity and accuracy of the performance. However, improvements in complexity and accuracy of the performance as a result of repeating structured tasks are at the expense of fluency. In other words, repeating structured tasks directs attention to form and leads learners to retrieve from the rule-based system rather than the exemplar-based system, and therefore, produce a more complex and more accurate performance. A further trade-off effect, which is explicable by referring to Kellogg's (1996) model of writing, was observed between structural and lexical complexity in the above finding. According to Kellogg, writers attend to both syntactical and lexical properties of their message at the 'translation' stage which is highly demanding on working memory, and there is thus a competition between the two. The above finding suggests that in this study, syntax is prioritised over lexis resulting in a more structurally complex performance.

The finding that repeating the unstructured tasks led to a less grammatically accurate, but more organised performance of a new task of a different type, is further evidence to support the limited attentional resources model and trade-off effects. This result suggests that performing tasks without a clear macro-structure and tasks that are not familiar to the writer's mind could be highly demanding on working memory and directs attention, which is limited, to one aspect of performance at the expense of the other. In this case, macro-measures were prioritised over micro-measures, which led to a loss in terms of grammatical accuracy but gains in terms of macro-measures.

Trade-offs were also evident between the structural complexity and fluency of the performance of a new task of a different type within the structured + reformulation group; that is, the structural complexity declined while fluency improved. This further suggests that there is competition between rule-based and exemplar-based systems and thus prioritising one is at the expense of the other.

Findings of this study also provide support to Ellis and Yuan's (2005) and Manchon's (2014) claims about the differences between oral and written performance. In contrast to oral language studies (Skehan & Foster, 1999; Tavakoli & Skehan, 2005; Tavakoli & Skehan, 2008; Tavakoli, 2009) which reported positive effects of task structure on accuracy and fluency, in this study repeating structured tasks increased the complexity and accuracy of performance. According to both Ellis and Yuan (2005) and Manchon (2014), unlike speech, writing does not take place in real time and so gives the writer more time to focus on formulation, execution, and monitoring and produce a more complex and more accurate written performance. Since

promoting both complexity and accuracy of performance is not always easily achieved, this finding is crucial and has strong implications for teaching practice which will be discussed in the next section.

Although findings indicate that reformulation did not promote the overall accuracy of the performance, a detailed analysis of participants' texts revealed that reformulation led to corrections of some of the errors that had been reformulated. This supports Swain's (1985) "output hypothesis" and Schmidt's (1990) "noticing hypothesis". In other words, performing the task for the first time gave learners the opportunity to develop awareness of the gaps between their interlanguage and the target language forms and the gaps they needed to fill to successfully communicate their intended meaning. Learners therefore looked for ways to overcome their interlanguage deficiencies in the subsequent input (i.e. reformulated texts). The search for these ways promoted noticing by leading participants to actively compare their own texts with the reformulated texts.

7.3.2. Pedagogical Implications

The findings of the study also have a number of implications for teaching practice. Firstly, teachers need to realise that certain task design features can affect task performance by either promoting or limiting development of different performance areas (i.e. complexity, accuracy, and fluency). The present study indicates that performing tasks with a clear macro-structure which is familiar to the learners promotes both complexity and accuracy of the performance, whereas unstructured tasks tend to direct learners' attention towards macro aspects of written performance and facilitate improvement in textual structure and organisation. Teachers thus need to be aware of their students' needs and incorporate more structured tasks in their teaching if it is to facilitate micro aspects of performance. On the other hand, if development of macro aspects of performance is desired, more unstructured tasks should be integrated into teaching to ensure noticing towards text structure is achieved and ultimately better structured texts are produced by learners. Obviously, maintaining a balance between structured and unstructured tasks is necessary if the aim is to develop both micro and macro aspects of performance.

Although lack of familiarity with note-expansion tasks influenced the way task repetition was expected to affect performance, it was evident from participants' responses to the questionnaires that they benefited from task repetition in working out how to perform note-expansion tasks. The conclusion thus is that implementing tasks under certain conditions, for instance, having opportunities to practise task performance before the actual performance of

the task, can ease the processing burden of tasks and allow for more attention to be allocated to performance areas. It should, however, be noted that it is also implied from the findings of the study that more repetitions over an extended period of time are required for task repetition effects to be internalised and for learning to take place.

In a similar vein, since repeating the structured tasks resulted in a more complex and more accurate performance of a new task of the same type, it can be argued that the interaction between task structure and task repetition creates a rare condition where both complexity and accuracy can improve. Since complexity and accuracy are in competition, and attending to one is at the expense of the other (Skehan, 1998b), teachers can consider providing opportunities for task repetition along with implementing structured tasks in their teaching to make sure both performance areas are attended to and the likelihood of development of both complexity and accuracy is maximised.

Teachers also need to make decisions about the types of tasks they employ in their teaching. It is evident from the findings that task type affected the way task repetition was expected to affect the performance of the same task. It was anticipated that during their first encounter with the task, the participants would work out the content they wanted to convey, and so the second encounter with the task would allow them to mainly focus on the language and the forms they needed to match to their intended meaning to successfully communicate their ideas. This would therefore lead to improvements in the second performance of the task (Bygate, 1999, 2001). However, these results were not yielded and a careful analysis of data indicated that participants included new content and ideas in their second performance, which hindered task repetition effects. This can partly be attributed to task type effect since tasks used in Bygate's (1999, 2001) studies were narrative tasks, which do not allow for inclusion of new content regardless of how many times they are completed, whereas the tasks used in this study (i.e. note-expansion tasks) made it possible for the participants to include new ideas when performing the task for the second time. Extra care thus needs to be taken in choosing pedagogic tasks as task type might affect implementation conditions and in turn impede the achievement of the instructional goals.

Providing opportunities to learners to monitor their output and discover their interlanguage deficiencies by comparing their own texts to the reformulated texts is also worthy of consideration since it promotes noticing of problematic areas in learners' language and directs learners to pay more attention to those areas in subsequent performance. However, teachers

should consider a few issues with regard to reformulation. First and foremost is how texts are reformulated; that is, whether they should be reformulated with regard to linguistic errors, or organisational errors, or both. In this study, texts were reformulated with regard to both linguistic and organisational errors but the results suggest that participants failed to notice all the changes and reformulations, especially in longer texts. It is therefore proposed that teachers decide which errors they want to be improved, and so they only reformulate those errors to avoid imposing a heavy processing burden on learners' cognitive resources and allow them space to focus on intended reformulations and attend to them in the later production. The second issue concerns implementation of reformulation. Firstly, the duration of the time that should be given to learners to compare their texts and the reformulated texts should be taken into consideration. Due to logistics, in the current study only 15 minutes were allocated for comparisons which seemed to be too short for learners to notice all the differences and discover the reasons for the changes because they had written long texts and also had to notice both linguistic and organisational errors. It is therefore important that teachers ensure the time they give their learners is sufficient. Text length, reformulated errors, and learners' proficiency level should be considered in determining the duration of this time. Secondly, teachers need to decide whether or not participants should write their texts and compare them with the reformulated texts individually or in collaboration with their peers. In most of the previous studies investigating reformulation (e.g. Qi & Lapkin, 2001; Swain & Lapkin, 2002; Adams, 2003), learners wrote their texts and compared them with the reformulated version collaboratively but wrote individually after reformulation. In the present study, however, participants went through all stages individually. A comparison of the results of the present study with the previous studies showed that collaboration, especially at the comparison stage, could be more beneficial as it creates opportunities for discussions which might promote deeper noticing of the differences. The last issue with regard to reformulation involves the difficulty of reformulating learners' texts, particularly for non-native English teachers. Apart from how time-consuming reformulating learners' texts is, it is a demanding task for non-native teachers to ascertain they make error-free and native-like changes to the learners' texts while preserving their intended meanings. Teachers can circumvent this problem in two ways. They can either reformulate one learner's text or use a model text (a text written by a native-speaker in response to the written task of the study) instead of a reformulated text to make comparisons and conduct class discussions (Yang and Zhang, 2010).

A further implication has to do with the educational context in which the study was conducted and the learners' educational background. As mentioned earlier, task-based language teaching methodology was not being practised in Iran at the time of data collection, and so the participants were not familiar with tasks and task repetition. They also had never experienced reformulation as a type of written feedback. This seemed to have largely affected the results of the study since learners included new content and ideas in their texts when repeating the same task for the second time. This was not favourable because as mentioned above, the theory behind repeating the same task is that on the second performance of the same task learners have already worked out the content of their output and therefore they have attention free to mainly retrieve the forms to match to that content. In addition, as participants' responses to the questionnaires indicated, they used the free attention from task repetition in working out how to tackle the note-expansion tasks instead of using it to improve their performance. Also, a lack of familiarity with reformulation could have hindered learners' improvement on the second performance. Thus, familiarity with teaching methods, tasks, and practice methods plays a crucial role in how learners understand the instructions given to them and how they benefit from teaching. This is worthy of consideration in settings in which new teaching methods are introduced and perhaps sufficient support needs to be provided to learners to make sure that teaching practices would be beneficial and the educational goals would be achieved.

Another implication which is closely related to the learning environment and learners' educational background is the importance of explicit instruction. Findings indicate that participants did not benefit from the treatments of the study to improve their performance with regard to macro-measures; that is, the gains were in favour of the unstructured and the control groups with the structured task groups experiencing losses. This implies that implicit instruction which indirectly guides students to use their cognitive resources in improving their performance may not be adequate and cannot produce the desired outcome, especially in contexts where there is an imbalance in teaching language skills with an emphasis placed on listening, reading, and speaking skills and the writing skills being largely ignored. In Iran, in spite of the recent revolutionary transformations in the education system, which includes writing education in primary school, teaching both Farsi and English writing suffers from an absence of a formal syllabus. In fact, the main focus in teaching writing is on sentence structure with no instruction taking place when textual structure and organisation is concerned. Hence teaching writing is perceived as teaching grammar. In Iran and similar contexts therefore there is a need to include explicit instruction (possibly by introducing either pre-task or post-task

activities or both to the syllabus) to directly address the areas which require specific attention since there is a possibility that favourable results would not occur with implicit instruction alone.

7.4. Limitations of the Study

Although this study makes potential contributions to task-based writing research, it is limited in a number of ways.

The first limitation concerns the potential contextual influences attributable to the setting in which data were collected (i.e. Iran). Participants' lack of prior knowledge in terms of textual organisation and structure, their unfamiliarity with task-based language teaching, note-expansions tasks, task repetition, and reformulation as a type of written feedback might have affected the results of the study and further research in a different setting might produce different results.

Issues with operationalisation of reformulation could also be a limitation of the study. Although it was arranged for a native English writer to reformulate participants' texts, she could not complete the task by the deadline and so, the researcher, who is not a native English speaker, had to reformulate the texts herself. The researcher consulted several dictionaries to ascertain if the reformulated structures were native-like and error-free. A further limitation involving reformulation is the fact that there were restrictions on the amount of time available during the reformulation session. Due to the restraints in the language schools where data were collected, the maximum time every session could last was 90 minutes. During all other sessions, this time was spent only for writing, whereas during the reformulation session, this time had to be spent on teaching reformulation, comparing original texts and the reformulated texts, *and* writing. Therefore, no more than 15 minutes was spent on comparing the original texts and the reformulated texts which was not sufficient for the long texts that the participants had written. These are assumed to have contributed to the effects of reformulation on writing.

The fact that the study did not adopt a counter-balanced design in implementing the task repetition study may also be regarded as a limitation. In the present study, all participants performed the *Drug Addiction* task in weeks 2 and 3 and then completed the *Divorce* task in week 4. Even though both tasks were of the same type and their topics were culturally familiar to the participants, it could be possible that practice effect might have contributed to the task repetition findings. Practice effect might have been avoided if a counter-balanced design had

been followed. However, with a mortality problem identified and another counter-balanced design adopted for pre- and post-writing tasks, a second counter-balanced design would have caused a lot of logistical problems.

Another limitation involves measuring fluency. As mentioned in the methodology chapter, the two measures of number of words per minute and ratio of disfluencies which were proposed prior to data collection could not be operationalised and hence fluency was measured employing two length measures of number of words per T-unit and number of words. Attempts were made during data collection to include a time measure by asking participants to write down the time they started writing and the time they finished but since it was observed that different participants used this time in a different way (some wrote drafts, whereas others just wrote a final draft), time measures could not be used. This can be seen as a limitation to the study because time is an important factor in measuring fluency (Indrarathne, 2013).

Another possible concern relates to the fact that the participants were recruited from two different learning environments. However, the two language schools are among the leading schools in Tabriz, Iran and are therefore similar in several ways, including the way they are run and managed, their teacher recruitment process, their syllabi, and the teaching materials they cover. Additionally, the pre-writing task results, analysed among the four groups, indicated that the groups were highly homogeneous in terms of their English writing proficiency (the only statistical difference among the groups was in presence and order of the problem-solution structure elements) (see Appendix G). It is also worth mentioning that to avoid the potential mortality effect identified in the pilot study, participants had to be recruited from the two language schools to ensure the sample size was large enough not to be affected by mortality.

Employing data driven scales instead of well-established writing scales (e.g. scales used in IELTS or TOEFL test) to rate the textual organisation of participants' written performance might also be considered a limitation of the study. However, since the purpose of the study was to mainly measure to what extent participants' texts reflected the problem-solution textual structure, there was a need to develop scales for this purpose as the well-established scales only measure general organisational features such as coherence and cohesion.

The last limitation has to do with task and text type. The two types of tasks selected for the purposes of this study were free-writing and note-expansions tasks involving the problem-solution textual structure. The results thus may not be generalisable to other types of tasks and texts, and further research might yield different results for a variety of task and text types.

7.5. Suggestions for Future Research

The present study is among the first studies which investigated the effect of task repetition on written performance. More research is thus needed to gauge the effects of task repetition on written performance. When doing this, it is worthwhile to consider investigating different task and text types. Adding more task repetitions to the design of the study (i.e. a longitudinal study) is also worth scrutinizing.

Similarly, this study is among the very first studies exploring the effect of task structure on written performance. This calls for more studies to better determine how written language is affected by task structure. The current study employed note-expansion tasks to operationalise task structure and investigated the effect of task structure in interaction with task repetition. Therefore, to allow comparisons with the oral task structure studies, all of which employed narrative tasks and examined the effect of task structure on single task performance, it is important to study the effect of task structure on single written performance employing narrative tasks.

In this study, reformulation was employed to provide feedback between the repeated performances. However, since the expected results were not produced and reformulation seemed not to have any notable effects on performance, it is suggested that in future research other types of written feedback be investigated. It is also proposed that more feedback sessions are introduced to the design of the study. One further direction for future research is to reformulate texts in a different way; that is, texts could be reformulated with regard to either micro-aspects or macro-aspects of texts, not both. Pair or class discussions can also be conducted at the comparison stage.

To analyse the effects of the independent variables, this study adopted measures which would allow comparability of the results with the findings of the previous writing task repetition studies and which Wolfe-Quintero et al. (1998) claimed to be the best measures for written accuracy and fluency. However, as noted in the limitations, no time measure could be included to analyse fluency of the performance. Future research could examine timed fluency or adopt other micro-measures employed in the writing research literature. This practice can help develop a deeper appreciation of the effects of the treatments on written performance.

The present study only adopted questionnaires to collect information about participants' accounts of whether they benefited from repetition and reformulation and if yes, how and what

changes they made to their texts when they repeated the task for the first and the second time. It is also suggested that more qualitative data in the form of interviews or think aloud are collected in future research to better understand the underlying cognitive process that learners go through when they engage in reformulation or task repetition.

As a final suggestion, the current study could be replicated or a similar study could be conducted in a different educational context. Since the findings of the present study indicate that the participants' educational background could have had an impact on their writing, replicating the current study or conducting a similar study in a different learning environment where learners have received a decent writing education and can to some extent demonstrate mastery of producing written texts both in their first language and the language under investigation, or a context where task-based language teaching is being practised and learners have gained familiarity with tasks and task implementation conditions, could shed light on this.

References

- Abdel Latif, M. M. M. (2014). Recent developments in writing fluency measurement. In T. Muller, J. Adamson, P. Shigeo Brown & S. Herder (Eds.), *Exploring EFL fluency in Asia* (pp. 196-212). Hampshire: Palgrave Macmillan.
- Abedifirouzjaie, J. (2014). The effect of different strategic planning foci on the accuracy of writing task performance. *Texas Papers in Foreign Language Education*, 16(1), 43-58.
- Adams, R. (2003). L2 output, reformulation and noticing: Implications for IL development. *Language Teaching Research*, 7(3), 347-376.
- Ahmadian, M., & Tavakoli, M. (2011). The effects of simultaneous use of careful on-line planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 15(1), 35-5.
- Ahmadian, M. J., Tavakoli, M., & Dastjerdi, H. V. (2015). The combined effects of online planning and task structure on complexity, accuracy, and fluency of L2 speech. *The Language Learning Journal*, 43(1), 41-56.
- Aliakbari, M. (2002). *Culture in ELT programs: An evaluation of the issue of culture in the Iranian ELT program in the high school level* (Unpublished doctoral thesis). Esfahan University, Esfahan, Iran.
- Bacha, N. N. (2010). Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes*, 9, 229-241.
- Baddeley, A. (1986). *Working Memory*. Oxford: Oxford University Press.
- Baleghizadeh, S., & Derakhshesh, A. (2012). The effect of task repetition and noticing on EFL learners' oral output. *International Journal of Instruction*, 5, 141-152.
- Bygate, M. (1996). Effects of task repetition: Appraising the developing language of learners. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 136-146). Oxford: Heinemann.
- Bygate, M. (1999). Task as the context for the framing, re-framing and unframing of language. *System*, 27, 33-48.
- Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 23-48). Harlow: Longman.

- Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 37-74). Amsterdam: John Benjamins.
- Bygate, M., Skehan, P., & Swain, M. (Eds.). (2001). *Researching pedagogic tasks, second language learning, teaching and testing*. Harlow: Longman.
- Byrnes, H., & Manchon, R. M. (Eds.). (2014). *Task-based language learning: Insights to and from writing*. Amsterdam: John Benjamins.
- Carter, R., & Nunan, D. (Eds.). (2001). *The Cambridge guide to teaching English to speakers of other languages*. Cambridge: Cambridge University Press.
- Chenoweth, A., & Hays, J. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, 18, 80-98.
- Cohen, A. (1983). Reformulating compositions. *TESOL Newsletter*, 17(6), 1-5.
- Cook, G., & Seidhofer, B. (Eds.). (1995). *Principles and practice in applied linguistics*. Cambridge: Cambridge University Press.
- Crookes, G., & Gass, S. (Eds.). (1993). *Tasks and language learning: Integrating theory and practice*. Clevedon: Multilingual Matters.
- Davari, H., & Aghagolzadeh, F. (2015). To teach or not to teach? Still an open question for the Iranian education system. In C. Kennedy (Ed.), *English language teaching in the Islamic Republic of Iran: Innovations, trends and challenges* (pp. 11-19). London: British Council.
- Day, R. (Ed.). (1986). *Talking to learn: Conversation in second language acquisition*. Rowley, MA: Newbury House.
- De Bot, K. (1992). A bilingual production model: Levelt's 'speaking model' adapted. *Applied Linguistics*, 13, 1-24.
- De Larios, J., Marin, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning*, 51, 497-538.
- Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, 9, 1-20.
- Ellis, R. (1995). Interpretation tasks for grammar teaching. *TESOL Quarterly*, 29(1), 87-106.
- Ellis, R. (2001). Non-reciprocal tasks, comprehension and second language acquisition. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 49-75). Harlow: Longman.
- Ellis, R. (2003). *Task-based language teaching and learning*. Oxford: Oxford University Press.

- Ellis, R. (Ed.). (2005). *Planning and task performance in a second language*. Amsterdam: John Benjamins.
- Ellis, R. (2009a). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, 19(3), 221-246.
- 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(4), 474-509.
- Ellis, R., & Shintani, N. (2014). *Exploring language pedagogy through second language acquisition research*. New York: Routledge.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26, 59-84.
- Ellis R., & Yuan, F. (2005). The effects of careful within-task planning on oral and written task performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 167-192). Amsterdam: John Benjamins.
- Farhady, H., Sajadi Hezaveh, F., & Hedayati, H. (2010). Reflections on foreign language education in Iran. *The Electronic Journal for English as a Second Language*, 13(4). Retrieved from <http://www.tesl-ej.org/wordpress/issues/volume13/ej52/ej52a1/>
- Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll)* (3rd Ed.). London: SAGE Publications.
- Foster-Cohen, S., García-Mayo, M. P., & Cenoz, J. (Eds.). (2005). *Eurosla yearbook*. Amsterdam: John Benjamins.
- Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition*, 18(3), 299-324.
- Foster, P., & Skehan, P. (1999). The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*, 3(3), 215-247.
- Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency and lexical diversity. *Language Learning*, 59(4), 866-896.
- García Mayo, M. P. (Ed.). (2007). *Investigating tasks in formal language learning*. Clevedon: Multilingual Matters.
- Gass, S. M., & Mackey, A. (Eds.). (2012). *The Routledge handbook of second language acquisition*. London: Routledge.
- Gass, S., & Madden, C. G. (Eds.). (1985). *Input in second language acquisition*. Rowley, MA: Newbury House.

- Gass, S., Mackey, A., Alvarez-Torres, M. J., & Fernandez-Garcia, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, 49(4), 549-581.
- Ghavamnia, M., Tavakoli, M., & Esteki, M. (2010). The effect of pre-task and online planning conditions on complexity, accuracy, and fluency on EFL learners' Written Production. *Porta Linguarum*, 20, 31-43.
- Ghorbani, M. R. (2009). ELT in Iranian high schools in Iran, Malaysia and Japan. *Reflections on English Language Teaching*, 8(2), 131-139.
- Gilabert, R. (2005). *The effects of task complexity on oral L2 narrative speech production* (Unpublished doctoral thesis). University of Barcelona, Spain.
- Givon, T. (1989). *Mind, code and context: Essays in pragmatics*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Grabe, W., & Kaplan, W. (1996). *Theory and Practice of writing: An applied linguistic perspective*. Harlow UK: Longman.
- Guchte, M., Braaksma, M., Rijlaarsdam, G., & Bimmel, P. (2016). Focus on form through task repetition in TBLT. *Language Teaching Research*, 20(3), 300-320.
- Hawkes, M. L. (2011). Using task repetition to direct learner attention and focus on form. *ELT Journal*, 12, 1-10.
- Henry, A., & Roseberry, R. L. (1998). An evaluation of a genre-based approach to the Teaching of EAP/ESP writing. *TESOL Quarterly*, 32(1), 147-156.
- Hoey, M. (1983). *On the surface of discourse*. London: Allen and Unwin.
- Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. London: Routledge.
- Indrarathne, B. (2013). Effects of task repetition on written language production in task based language teaching. *Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching*. Retrieved from <http://www.lancaster.ac.uk/fass/events/laelpgconference/papers/v08/Bimali-Indrarathne.pdf>
- Johnson, M. D., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity and lexical complexity. *Journal of Second Language Writing*, 21, 264-282.
- Jung, S. (2013). The effect of task repetition and corrective feedback in L2 writing: A pilot study. *MSU Working Papers in SLS*, 4, 24-38.

- Kassen, M. A. (1988). Native and non-NS teacher response to foreign language learner writing: A study of intermediate level French. *Texas Papers in Foreign Language Education*, 1, 1-15.
- Kellog, R. (1987). Effects of topic knowledge on the allocation of processing time and cognitive efforts to writing processes. *Memory and Cognition*, 15, 256-266.
- Kellog, R. (1988). Attentional overload and writing performance: Effects of rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 355-365.
- Kellog, R. (1990). Effectiveness of prewriting strategies as a function of task demands. *American Journal of psychology*, 103, 327-342.
- Kellog, R. (1996). A model of working memory in writing. In C. Levy & S. Ransdell (Eds.), *The science of writing* (pp. 57-71). Mahwah NJ: Lawrence Erlbaum.
- Kennedy, C. (Ed.). (2015). *English language teaching in the Islamic Republic of Iran: Innovations, trends and challenges*. London: British Council.
- Khodabandeh, F. (2014). Argumentation across L1 and L2: Examination of three instructional treatments of genre-based approach to teaching writing. *Procedia - Social and Behavioural Sciences*, 98, 968-975.
- Kim, Y. (2013). Promoting attention to form through task repetition in a Korean EFL context. In K. McDonough & A. Mackey (Eds.), *Second language interaction in diverse educational settings* (pp. 3-24). Amsterdam: John Benjamins.
- Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System*, 41, 829-840.
- Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20(2), 148-161.
- Kuiken, F., & Vedder, I. (2007a). Cognitive task complexity and linguistic performance in French L2 writing. In M.P. García Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 117-135). Clevedon: Multilingual Matters.
- Kuiken, F., & Vedder, I. (2007b). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics*, 45(3), 261-284.
- Kuiken, F., & Vedder, I. (2007c). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17(1), 48-60.
- Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task*

- complexity: Researching the cognition hypothesis of language learning and performance* (pp. 91-104). Amsterdam: John Benjamins.
- Kuiken, F., Mos, M., & Vedder, I. (2005). Cognitive task complexity and second language writing performance. In S. Foster-Cohen, M. P. García-Mayo & J. Cenoz (Eds.), *Eurosla yearbook* (Vol. 5., pp. 195-222). Amsterdam: John Benjamins.
- Kuiken, F., & Vedder, I. (2012). Speaking and writing tasks and their effects on second language performance. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 364-377). London: Routledge.
- Lapkin, S., Swain, M., & Smith, M. (2002). Reformulation and the learning of French pronominal verbs in a Canadian French immersion context. *The Modern Language Journal*, 86(4), 485-507.
- 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(4), 590-619.
- Lazaro Ibarrola, A. (2009). Reformulation and self-correction: Testing the validity of correction strategies in the classroom. *Revista Espanola de Linguistica Aplicada*, 22(22), 189-215.
- Lazaro Ibarrola, A. (2013). Reformulation and self-correction: Insights into correction strategies for EFL writing in a school context. *VIAL Vigo International Journal of Applied Linguistics*, 10, 29-49.
- Leaver, B., & Willis, J. (Eds.). (2004). *Task-based instruction in foreign language education: Practices and programs*. Washington, D. C.: Georgetown University Press.
- Lennon, P. (1990) Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge MA: The MIT Press.
- Levy, C., & Ransdell, S. (Eds.). (1996). *The science of writing*. Mahwah NJ: Lawrence Erlbaum.
- Long, M. (1989). Task, group, and task-group interaction. *University of Hawai'i Working Papers in English as Second Language*, 8(2), 1-26.
- Long, M. H. & Douhty, C. J. (Eds.). (2009). *The handbook of language teaching*. Malden: Wiley-Blackwell.
- Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4, 221-250.

- Lynch, T., & Maclean, J. (2001). Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 141-163). Harlow: Longman.
- McDonough, K., & Mackey, A. (Eds.). (2013). *Second language interaction in diverse educational settings*. Amsterdam: John Benjamins.
- Mackey, A., Kaganas, A. P., & Oliver, R. (2007). Task familiarity and interactional feedback in child ESL classrooms. *TESOL Quarterly*, 40, 107-132.
- Manchon, R. M. (2014). The distinctive nature of task repetition writing. Implications for theory, research and pedagogy. *Estudios de Lingüística Inglesa Aplicada*, 14, 13-41.
- Matsumura, M., Kawamura, K., & Affricano, A. (2008). Narrative task-type repetition and changes in second language use in a classroom environment: A case study. *Journal of the Faculty of Education and Human Sciences Yokohama National University*, 10, 124-45.
- Navon, D. (1989). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system. *European Journal of Cognitive Psychology*, 1, 191-238.
- Nitta, R., & Baba, K. (2014). Task repetition and L2 writing development: A longitudinal study from a dynamic systems perspective. In H. Byrnes & R. M. Manchón (Eds.), *Task-based language learning: Insights to and from writing* (pp. 107-136). Amsterdam: John Benjamins.
- Norris, J. M. (2009). Task-based teaching and testing. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 578-594). Malden: Wiley-Blackwell.
- Neumann, O. (1996). Theories of attention. In O. Neumann & A. Sanders (Eds.), *Handbook of perception and action: Vol. 3. Attention* (pp. 389-446). San Diego: Academic Press.
- Neumann, O., & Sanders, A. (Eds.). (1996). *Handbook of perception and action: Vol. 3. Attention*. San Diego: Academic Press.
- Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*, 19, 218-233.
- Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 77-109). Amsterdam: John Benjamins.

- Pakbaz, R., & Rezai, M. J. (2015). The effect of task repetition and consciousness-raising on Iranian L2 learners' writing performance. *International Journal of Language and Applied Linguistics*, 1(2), 49-57.
- Patanasorn, C. (2010). *Effect of procedural content and task repetition on accuracy and fluency in an EFL context* (Unpublished doctoral dissertation). Northern Arizona University, Arizona, USA.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912.
- Plough, L., & Gass, S. (1993). Interlocutor and task familiarity: Effects on interaction structure. In G. Crookes & S. Gass (Eds.), *Tasks and language learning: Integrating theory and practice* (pp. 35-56). Clevedon: Multilingual Matters.
- Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
- Prabhu, V. S. (1987). *Second language pedagogy*. Oxford: Oxford University Press.
- Qi, D. S., & Lapkin, S. (2001). Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing*, 10, 277-303.
- Rahimpour, M., Mehrang, H., & Hosseini, P. (2011). Effects of task design on EFL learners' written language performance. *The Journal of Asia TEFL*, 8(2), 95-117.
- Revesz, A., Kourtali, N., & Mazgutova, D. (2016). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*. Retrieved from <http://onlinelibrary.wiley.com.ezproxy.auckland.ac.nz/doi/10.1111/lang.12205/epdf>
- Robinson, P. (2001). Task complexity, task difficulty and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27-57.
- Robinson, P. (2003). The cognition hypothesis, task design and adult task-based language learning. *Second Language Studies*, 21(2), 45-107.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 43(1), 1-32.
- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on speech production, interaction, uptake of premodified input and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 45(3), 195-215.

- Robinson, P. (2011a). Task-based language learning: A review of issues. *Language Learning*, 61, 1-36.
- Robinson, P. (Ed.). (2011b). Second language task complexity: *Researching the cognition hypothesis of language learning and performance*. Amsterdam: John Benjamins.
- Ruiz-Funes, M. (2014). An exploration of task complexity in foreign language writing at the intermediate level and measures of linguistic production. *Journal of Linguistics and Language Teaching*, 5(2), 181-206.
- Sachs, R., & Polio, C. (2007). Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition*, 29(1), 67-100.
- Sadeghi, K., & Richards, J. C. (2015). The idea of English in Iran: An example from Urmia. *Journal of Multilingual and Multicultural Development*, 37(4), 419-434.
- Saeedi, M., & Kazerooni, S. R. (2014). The influence of task repetition and task structure on EFL learners' oral narrative retellings. *Innovation in Language Learning and Teaching*, 8, 116-131.
- Selinker, L., & Douglas, D. (1985). Wrestling with context in interlanguage theory. *Applied Linguistics*, 6, 190-204.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.
- Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.
- Schneider, W., & Shiffrin, R. (1977). Controlled and automatic human information processing I: Detection, search and attention. *Psychological Review*, 84(1), 1-66.
- Sheppard, C. (2006). *The effects of instruction directed at the gaps second language learners noticed in their oral production* (Unpublished Doctoral thesis). University of Auckland, Auckland. New Zealand.
- Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing II: Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84(2), 127-190.
- Skehan, P. (1996a). Second language acquisition research and task-based instruction. In J. Willis, & D. Willis (Eds.), *The challenge and change in language teaching* (pp. 17-30). Oxford: Heinemann.

- Skehan, P. (1996b). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17(1), 38-62.
- Skehan, P. (1998b). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (1998a). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268-286.
- Skehan, P. (2003). Task based instruction. *Language Teaching*, 36(1), 1-14.
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1, 185-211.
- Skehan P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93-120.
- Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 193-218). Amsterdam: John Benjamins.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidhofer (Eds.), *Principles and practice in applied linguistics* (pp. 64-81). Cambridge: Cambridge University Press.
- Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, 16, 371-391.
- Swain, M., & Lapkin, S. (2002). Talking it through: Two French immersion learners' response to reformulation. *International Journal of Educational Research*, 37, 285-304.
- Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, 4, 143-163.
- Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System*, 37, 482-495.
- Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, 58(2), 439-473.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-273). Amsterdam: John Benjamins.

- Thornbury, S. (1997). Reformulation and reconstruction: Tasks that promote 'noticing'. *ELT Journal*, 51(4), 326-335.
- Tocalli-Beller, A., & Swain, M. (2005). Reformulation: The cognitive conflict and L2 learning it generates. *International Journal of Applied Linguistics*, 15(1), 5-28.
- Tomlin, R. (1984). The treatment of foreground-background information in the on-line descriptive discourse of second language learner. *Studies in Second Language Acquisition*, 6, 115-142.
- Van de Guchte, M., Braaksma, M., Rojlaarsdam, G., & Bimmel, P. (2016). Focus on form through task repetition in TBLT. *Language Teaching Research*, 20(3), 300-320.
- Vanpatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood: NJ: Ablex.
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17(1), 65-83.
- Vygotsky, L. S. (1962). *Thought and language* (E. Haufmann & G. Vakar, Eds. and Trans.). Cambridge: MIT Press.
- Widdowson, H. (2003). Pedagogic design. In H. Widdowson (Ed.), *Defining issues in English language teaching* (pp. 111-134). Oxford: Oxford University Press.
- Widdowson, H. (2003). *Defining issues in English language teaching*. Oxford: Oxford University Press.
- Willis, J. (1996). *A framework for task-based learning*. Harlow: Longman.
- Willis, J. (2004). Perspectives on task-based instruction: Understanding our practices, acknowledges different practitioners. In B. Leaver, & J. Willis (Eds.), *Task-based instruction in foreign language education: Practices and programs* (pp. 3-44). Washington, D. C.: Georgetown University Press.
- Willis, J., & Willis, D. (Eds.). (1996). *Challenge and change in language teaching*. Oxford: Heinemann.
- Willis, D., & Willis, J. (2001). Task-based language learning. In R. Carter & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 173-179). Cambridge: Cambridge University Press.
- Witte, S. P. (1987). Pretext and composing. *College Composition and Communication*, 38, 397-425.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Center.

- Yang, L., & Zhang, L. (2010). Exploring the role of reformulations and a model text in EFL students' writing performance. *Language Teaching Research*, 14(4), 464-484.
- Yasuda, S. (2011). Genre-based tasks in foreign language writing: Developing writers' genre awareness, linguistic knowledge and writing competence. *Journal of Second Language Writing*, 20, 111-133.
- Zhang, L. J. (2010). A dynamic metacognitive systems account of Chinese university students' knowledge about EFL reading. *TESOL Quarterly*, 44, 320-353.
- Zimmerman, R. (2000). L2 writing: Subprocesses, a model of formulating and empirical findings. *Learning and Instruction*, 10, 73-99.

Appendix A: Ethics Forms



THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Department of Applied Language Studies and Linguistics
Level 4, Arts 1, Building 206, 14a Symonds Street
Phone: +64 9 373 7599 ext. 86252

The University of Auckland
Private Bag 92019
Auckland, New Zealand

PARTICIPANT INFORMATION SHEET

Project title: The Effect of Task Structure, Task Repetition and Reformulation on Second Language Written Performance

Name of Researcher: Faezeh Mehrang

Researcher introduction

My name is Faezeh Mehrang and I am a PhD candidate in Language Teaching and Learning at The University of Auckland, Department of Applied Language Studies and Linguistics. I am currently undertaking a research project for my dissertation. The purpose and nature of this research will be described below. Please take your time to read through this form.

Project description and invitation

The main purpose of my research is to explore the effectiveness of task-based language teaching. This study investigates how task design and task conditions affect language acquisition. I would like to invite you to participate in this project because your proficiency level is ideal for the implementation of the study. Moreover, the study will have benefits for your language learning experience, both in providing opportunities to receive writing instruction, as well as practicing writing tasks in your second language, which will be beneficial to your university studies. She will provide feedback on individual task performance at the end of the study to those who would like to receive a summary of results.

Project Procedures

This project mainly involves asking students to write about selected topics. There will be four groups of participants: three experimental groups and a control group. The study has five sessions. In the beginning and final sessions, all participants will be given two topics to write about. The participants will take a pre-test and a post-test respectively by completing a writing task that is estimated to take 45 minutes. During the middle sessions, participants in the experimental groups will be given a note to expand. This will take about 30 minutes and will be repeated twice. One of the experimental groups will receive some feedback before they repeat the task for the second time. The feedback involves students comparing their own texts with the reformulated version of them and will take about 15 minutes. After the repeated performance of the tasks, participants in the experimental groups will be given a questionnaire to complete. This will take about 10 minutes. The control group will only attend the testing sessions.

Data storage/retention/destruction/future use

The data collected will include the written performance and completed questionnaires of the participants. The data will be used for the purpose of this project, and the results will be presented at conferences and published in academic journals. The data will be stored in a locked cabinet on university premises. The data will be destroyed after 6 years by means of shredding in a way that makes it impossible to be recovered.

Right to Withdraw from Participation

Your participation in this study is entirely voluntary. You have the right to withdraw from the study at any time. If you wish to withdraw from the study, you will be able to withdraw any data associated with you up to 20th of June, 2013.

Anonymity and Confidentiality

Every effort will be made to keep the collected data confidential. Students' names or identity will not be revealed at any time during the study. Students will use pseudonyms instead of their own names. That is, all students will be asked to select a pseudonym and use it for any item that they submit to the researcher.

The data will not be viewed by any other person or organisation except for the researchers, the reformulator and the second rater of the data who have both signed the confidentiality agreement form. The reformulator has the responsibility of reading the participants' texts and rewriting them so that they reflect native-like usage of English. The second rater will rate about 20% of the data to ensure that the marks awarded on the tasks can be considered to reveal as accurately as possible what participants can do.

Contact Details and Approval Wording

Researcher (New Zealand contact details):

Faezeh Mehrang
Room 436,
Level 4, Department of Applied Language Studies and Linguistics,
Faculty of Arts, Building 206,
14a Symonds Street,
Auckland, New Zealand
Email: fmeh210@aucklanduni.ac.nz
Phone: +64 9 373 7599 ext. 86252

Researcher (local contact details):

Faezeh Mehrang
13 Sharabiani Alley,
Haj Naeb Alley,
New 17 Shahrivar Street,
Tabriz, Iran
Phone: +98 914 412 6181

Supervisor:

Prof. Rod Ellis
Room 403,
Level 4, Department of Applied Language Studies and Linguistics,
Faculty of Arts, Building 206,
14a Symonds Street,

Auckland, New Zealand
Email: r.ellis@auckland.ac.nz
Phone: +64 9 373 7599 ext. 84876

Head of the Department:
Dr. Rosemary Wette
Room 550,
Level 5, Department of Applied Language Studies and Linguistics,
Faculty of Arts, Building 206,
14a Symonds Street,
Auckland, New Zealand
Email: r.wette@auckland.ac.nz
Phone: +64 9 373 7599 ext. 88538

Chair contact details: "For any queries regarding ethical concerns you may contact the Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Research Office, Private Bag 92019, Auckland 1142. Telephone 09 373-7599 extn. 87830/83761. Email: humanethics@auckland.ac.nz."

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 20-March-2012 for (3) years, Reference Number 7911.

CONSENT FORM
THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project title: The Effect of Task Structure, Task Repetition and Reformulation on Second Language Written Performance

Name(s) of Researcher(s): Faezeh Mehrang

- I have read the Participant Information Sheet and have understood the nature of the research. I have had the opportunity to ask questions and have them answered to my satisfaction.
- I agree to take part in this research.
- I understand that I am free to withdraw participation at any time, and to withdraw any data traceable to me up to 20th of June, 2013.
- I understand that my participation or non-participation will not affect my course grades and my relationship with the school.
- I wish / do not wish to receive the summary of findings. (Please circle)
- I understand that a third party who has signed a confidentiality agreement will reformulate my texts.
- I understand that a third party who has signed a confidentiality agreement will rate my texts.
- I understand that data will be kept for 6 years, after which they will be destroyed.

Name _____

Signature _____ Date _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 20-March-2012 FOR (3) YEARS REFERENCE NUMBER 7911.

Appendix B: Reproduction Tasks

Spam Task (structured)

Spam emails are considered to be one of the major problems for email users. Read the text below about spam emails. Use the ideas in the text to write an article on this topic to be published in a science magazine. Write at least 250 words.

Email is an effective means of communication for business and personal use. People use emails because they have several advantages over traditional post. Emails are delivered extremely fast and they can be sent 24 hours a day, 365 days a year. They are very cheap. That is, when using high speed internet connection, each email which is sent is free. Dial up users are charged at local call rates but it only takes a few seconds to send an email. Finally, they can be sent and received from any computer, anywhere in the world, that has an internet connection.

Although emails have several benefits for internet users, they can also cause problems. One of the most common problems with emails is Spam. Spam is unwanted mail which is sent to millions of people each day. It may seem that sending spam has increased just recently, but the fact is it has been around as long as the internet has. Actually, the first case of spam happened in 1978, when a computer company sent out 400 emails using Arpanet (a primary kind of modern internet). Now, spam emails make up more than two-thirds of all the email sent over the internet, and for some unlucky users, it makes up 80 percent of the messages they receive.

Since spam has been annoying to all its recipients, there have been attempts to stop and filter them. Most spam filters work using finger printing system. In this system, a program analyzes several spam messages and finds out common characteristics in them. Any arriving emails that match these characteristics are deleted and cannot enter peoples' inboxes. A second possible solution is to use "smart filters". These filters stop spams by looking for words and phrases that are usually used in a spam message and thus deleting any email which includes these words and phrases. Another spam stopper uses a proof system. With this system, a user must first confirm that he or she is a person before the email is sent. To do so, he or she should solve a simple puzzle or answer a question. This system prevents Automatic spam systems from sending out large numbers of emails because computers are not able to pass confirmation tests.

The "finger printing" system is easy for spammers to defeat. To confuse the program, a spammer simply has to include a series of random character or numbers. These character or numbers change its fingerprint, so that the spam cannot be detected. Concerning the smart filters, again the spammers have learned how to avoid them. That is, they hide words and phrases by using numbers or other characters instead of letters. For example, the word "money" might be written by a zero instead of the letter "O". However, with

the proof system, spam no longer becomes cheap to send because each email would have to be confirmed by a person before it could be sent which stops sending large numbers of emails at once. So far, spammers have not been able to defeat proof systems.

Spam Task (unstructured)

Spam emails are considered to be one of the major problems for email users. Read the text below about spam emails. Use the ideas in the text to write an article on this topic to be published in a science magazine. Write at least 250 words.

The “finger printing” system is easy for spammers to defeat. To confuse the program, a spammer simply has to include a series of random character or numbers. These character or numbers change its fingerprint, so that the spam cannot be detected. Concerning the smart filters, again the spammers have learned how to avoid them. That is, they hide words and phrases by using numbers or other characters instead of letters. For example, the word “money” might be written by a zero instead of the letter “O”. However, with the proof system, spam no longer becomes cheap to send because each email would have to be confirmed by a person before it could be sent which stops sending large numbers of emails at once. So far, spammers have not been able to defeat proof systems.

Since spam has been annoying to all its recipients, there have been attempts to stop and filter them. Most spam filters work using finger printing system. In this system, a program analyzes several spam messages and finds out common characteristics in them. Any arriving emails that match these characteristics are deleted and cannot enter peoples’ inboxes. A second possible solution is to use “smart filters”. These filters stop spams by looking for words and phrases that are usually used in a spam message and thus deleting any email which includes these words and phrases. Another spam stopper uses a proof system. With this system, a user must first confirm that he or she is a person before the email is sent. To do so, he or she should solve a simple puzzle or answer a question. This system prevents Automatic spam systems from sending out large numbers of emails because computers are not able to pass confirmation tests.

Although emails have several benefits for internet users, they can also cause problems. One of the most common problems with emails is Spam. Spam is unwanted mail which is sent to millions of people each day. It may seem that sending spam has increased just recently, but the fact is it has been around as long as the internet has. Actually, the first case of spam happened in 1978, when a computer company sent out 400 emails using Arpanet (a primary kind of modern internet). Now, spam emails make up more than two-thirds of all the email sent over the internet, and for some unlucky users, it makes up 80 percent of the messages they receive.

Email is an effective means of communication for business and personal use. People use emails because they have several advantages over traditional post. Emails are delivered extremely fast and they can be sent 24 hours a day, 365 days a year. They are very cheap. That is, when using high speed internet connection, each email which is sent is free. Dial up users are charged at local call rates but it only takes a few seconds to send an email. Finally, they can

be sent and received from any computer, anywhere in the world, that has an internet connection.

Sea Level Task (structured)

The rise in sea levels is a threat to coastal parts of the world. Read the text below about this problem. Use the ideas in the text to write an article on this topic to be published in a science magazine. Write at least 250 words.

Since 1979, it has been agreed that a doubling of carbon dioxide would raise global temperatures 1.5 to 4.5 degrees Celsius. Emissions of methane and other gases that absorb infrared radiation could speed up this process. Although attention has been given to methods to limit global warming, most experts feel there will be an average temperature increase of one to two degrees Celsius.

The potential danger of the increase in worldwide temperature was first understood in the early 1970s. Scientists predicted that the level of the seas all over the world would go up about 6 meters because the ice in South Pole is melting. This rise in sea levels would lead to flooding, and as a result, coastal parts of the world would be under water. Scientists predict a 15- to 30- meters loss of land in New Jersey, and up to 1,000 feet of shore areas lost in Florida. According to some studies, the rise in water levels would result in the loss of 5- to 90 percent of U.S. wetlands.

At the moment, two major ways have been considered by coastal communities to deal with this problem. The first is called the no-protection method. Communities in coastal parts simply identify those coastal areas which are expected to be lost within the next 30 to 60 years. No new buildings are permitted to be built in these areas and the structures that already exist are deserted and destroyed. The second option is to raise the land level along the shore. That is, the entire land mass is raised to protect it from the ocean. To raise the land, sand must be pumped onto the beach (including the underwater part of the beach) until the land level gradually rises. Furthermore, roads, houses, and other structures must be gradually raised again.

Communities that take a no-protection method approach to coming danger are often unable to afford the finances of removing and pulling down the beachfront buildings. However, it should be noted that the communities that adopt the no-protection approach place the financial burden on the government to buy the land which is destroyed by floods. The second option (raising the land level) is a better solution. First, there is no need to destroy buildings. Second, the government does not have to pay for flood insurance or buy the land. This approach, however, is very expensive and involves great amount of hard work.

Sea Level Task (unstructured)

The rise in sea levels is a threat to coastal parts of the world. Read the text below about this problem. Use the ideas in the text to write an article on this topic to be published in a science magazine. Write at least 250 words.

Communities that take a no-protection method approach to coming danger are often unable to afford the finances of removing and pulling down the beachfront buildings. However, it should be noted that the communities that adopt the no-protection approach place the financial burden on the government to buy the land which is destroyed by floods. The second option (raising the land level) is a better solution. First, there is no need to destroy buildings. Second, the government does not have to pay for flood insurance or buy the land. This approach, however, is very expensive and involves great amount of hard work.

At the moment, two major ways have been considered by coastal communities to deal with this problem. The first is called the no-protection method. Communities in coastal parts simply identify those coastal areas which are expected to be lost within the next 30 to 60 years. No new buildings are permitted to be built in these areas and the structures that already exist are deserted and destroyed. The second option is to raise the land level along the shore. That is, the entire land mass is raised to protect it from the ocean. To raise the land, sand must be pumped onto the beach (including the underwater part of the beach) until the land level gradually rises. Furthermore, roads, houses, and other structures must be gradually raised again.

The potential danger of the increase in worldwide temperature was first understood in the early 1970s. Scientists predicted that the level of the seas all over the world would go up about 6 meters because the ice in South Pole is melting. This rise in sea levels would lead to flooding, and as a result, coastal parts of the world would be under water. Scientists predict a 15- to 30- meters loss of land in New Jersey, and up to 1,000 feet of shore areas lost in Florida. According to some studies, the rise in water levels would result in the loss of 5- to 90 percent of U.S. wetlands.

Since 1979, it has been agreed that a doubling of carbon dioxide would raise global temperatures 1.5 to 4.5 degrees Celsius. Emissions of methane and other gases that absorb infrared radiation could speed up this process. Although attention has been given to methods to limit global warming, most experts feel there will be an average temperature increase of one to two degrees Celsius.

Appendix C: Questionnaires (Farsi Version)

Structured and Unstructured Groups:

سوالات زیر را به دقت بخوانید و پاسخ دهید.

- آیا به نظر شما تکرار این فعالیتهای زبانی مفید بود؟
الف) خیلی مفید ب) تا حدی مفید ج) کمی مفید د) غیر مفید
- تکرار این فعالیتهای زبانی چگونه به شما کمک کرد که نوشته ی نهایی خود را بهبود ببخشید؟
- آیا فکر می کنید در نتیجه تکرار این فعالیتهای زبانی نکته زبانی جدیدی اعم از لغت و گرامر و یا نکته ای در مورد ساختار نوشتاری زبان انگلیسی آموختید؟ اگر بله، لطفا بگویید چه چیزهایی آموختید.
- در نتیجه ی تکرار این فعالیت زبانی چه تغییراتی در زمینه های زیر به نوشته ی نهایی خود اعمال کردید:
الف. تغییرات در زمینه محتوا
ب. تغییرات زبانی (لغت، گرامر، املا)
ج. تغییرات ساختاری

Structured + Reformulation Group:

سوالات زیر را به دقت بخوانید و پاسخ دهید.

- آیا به نظر شما تکرار این فعالیتهای زبانی مفید بود؟
الف) خیلی مفید ب) تا حدی مفید ج) کمی مفید د) غیر مفید
- تکرار این فعالیتهای زبانی چگونه به شما کمک کرد که نوشته ی نهایی خود را بهبود ببخشید؟
- آیا فکر می کنید در نتیجه تکرار این فعالیتهای زبانی نکته زبانی جدیدی اعم از لغت و گرامر و یا نکته ای در مورد ساختار نوشتاری زبان انگلیسی آموختید؟ اگر بله، لطفا بگویید چه چیزهایی آموختید.
- در نتیجه ی تکرار این فعالیت زبانی چه تغییراتی در زمینه های زیر به نوشته ی نهایی خود اعمال کردید:
الف. تغییرات در زمینه محتوا
ب. تغییرات زبانی (لغت، گرامر، املا)
ج. تغییرات ساختاری
- آیا به نظر شما مقایسه ی نوشته ی خود با نسخه ی اصلاح شده ی آن مفید بود؟
- مقایسه ی نوشته ی خود با نسخه ی اصلاح شده ی آن چگونه به شما کمک کرد که نوشته ی نهایی خود را بهبود بخشید؟
- آیا فکر می کنید در نتیجه ی مقایسه ی نوشته ی خود با نسخه ی اصلاح شده ی آن نکته زبانی جدیدی اعم از لغت و گرامر و یا نکته ای در زمینه ی ساختار نوشتاری زبان انگلیسی یاد گرفتید؟ اگر بله، لطفا بگویید چه چیزی یاد گرفتید.
- چه تغییراتی در زمینه های زیر در نتیجه ی مقایسه ی نوشته ی خود با نسخه ی اصلاح شده ی آن در نوشته نهایی خود اعمال کردید؟
الف. تغییرات در زمینه محتوا
ب. تغییرات زبانی (لغت، گرامر، املا)
ج. تغییرات ساختاری

Appendix D: Pilot Study Results

Research Question One:

Table 1 *Results for the Structured Group's T1 – T3 Performance*

Measure	N	T1		T3		Sig.
		Mean	Std.	Mean	Std.	
SC	8	1.34	.28	1.24	.11	.34
LD	8	48.28	5.38	46.03	17.95	1.0
A	8	.17	.15	.22	.11	.54
F	8	9.38	2.83	8.56	1.27	.43
HR	8	1.63	1.18	2.63	1.18	.17
PSSR	8	3.00	.80	3.06	.41	.70

SC = Structural Complexity; LD = Lexical Density; A = Accuracy; F = Fluency; HR = Holistic Rating; PSSR = Problem-Solution Structure Rating

Research Question Two:

Table 2 *Results for the Structured Group's T1 – T4 Performance*

Measure	N	T1		T4		Sig.
		Mean	Std.	Mean	Std.	
SC	8	1.34	.28	1.50	.31	.20
LD	8	48.28	5.38	44.31	4.86	.23
A	8	.17	.15	.04	.05	.17
F	8	9.38	2.83	13.50	4.22	.02*
HR	8	1.63	1.18	2.38	1.50	.27
PSSR	8	3.00	.80	2.56	.82	.19

* Significance reached

SC = Structural Complexity; LD = Lexical Density; A = Accuracy; F = Fluency; HR = Holistic Rating; PSSR = Problem-Solution Structure Rating

Research Question Three:

Table 3 *Results for the Structured + Reformulation Group's T1 – T3 Performance*

Measure	N	T1		T3		Sig.
		Mean	Std.	Mean	Std.	
SC	13	1.38	.26	1.36	.18	.88
LD	13	49.97	8.87	45.97	12.93	.65
A	13	.19	.14	.19	.14	.97
F	13	10.17	2.09	10.94	1.23	.23
HR	13	2.62	.96	2.54	1.66	.96
PSSR	13	2.76	.72	3.23	.72	.20

SC = Structural Complexity; LD = Lexical Density; A = Accuracy; F = Fluency; HR = Holistic Rating; PSSR = Problem-Solution Structure Rating

Research Question Four:

Table 4 *Results for the Structured + Reformulation Group's T1 – T4 Performance*

Measure	N	T1		T4		Sig.
		Mean	Std.	Mean	Std.	
SC	13	1.38	.26	1.48	.40	.46
LD	13	49.97	8.87	43.48	3.84	.01*
A	13	.19	.14	.11	.16	.20
F	13	10.17	2.09	13.44	3.77	.01*
HR	13	2.62	.96	2.00	1.58	.20
PSSR	13	2.76	.72	3.19	1.07	.10

* Significance reached

SC = Structural Complexity; LD = Lexical Density; A = Accuracy; F = Fluency; HR = Holistic Rating; PSSR = Problem-Solution Structure Rating

Research Question Five:

Table 5 *Results for the Unstructured Group's T1 – T3 Performance*

Measure	N	T1		T3		Sig.
		Mean	Std.	Mean	Std.	
SC	15	1.48	.31	1.43	.23	.58
LD	15	46.46	5.29	48.05	4.58	.39
A	15	.17	.14	.18	.19	.77
F	15	11.71	2.80	10.52	1.79	.11
HR	15	2.47	1.12	2.47	1.40	.88
PSSR	15	2.63	.61	3.06	.59	.03*

* Significance reached

SC = Structural Complexity; LD = Lexical Density; A = Accuracy; F = Fluency;
HR = Holistic Rating; PSSR = Problem-Solution Structure Rating

Research Question Six:

Table 6 *Results for the Unstructured Group's T1 – T4 Performance*

Measure	N	T1		T4		Sig.
		Mean	Std.	Mean	Std.	
SC	15	1.48	.31	1.43	.24	.56
LD	15	46.46	5.29	47.46	5.09	.77
A	15	.17	.14	.08	.14	.07
F	15	11.71	2.80	13.18	3.69	.19
HR	15	2.47	1.12	1.93	.96	.10
PSSR	15	2.63	.61	2.26	.92	.19

SC = Structural Complexity; LD = Lexical Density; A = Accuracy; F = Fluency;
HR = Holistic Rating; PSSR = Problem-Solution Structure Rating

Research Questions Seven, Eight, and Nine:

Structural Complexity:

Table 7 *Pre-writing Structural Complexity ANOVA Results*

Group	N	Mean	St. Deviation	f	Sig.
Str.	8	1.54	.42	3.40	.03*
Str. + ref.	13	1.61	.19		
Unstr.	15	1.77	.29		
Control	11	1.45	.11		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

* Significance reached

Table 8 *Post-writing Structural Complexity ANCOVA Results*

Group	N	Mean	St. Deviation	f	Sig.
Str.	8	1.5600	.33624	1.22	.31
Str. + ref.	13	1.4631	.29341		
Unstr.	15	1.5400	.27633		
Control	11	1.3355	.16525		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Lexical Density

Table 9 *Pre-writing Lexical Density Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	43.86	3.17	3.97	.26
Str. + ref.	13	42.11	4.92		
Unstr.	15	44.02	3.89		
Control	11	46.52	4.74		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 10 *Post-writing Lexical Density Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	45.4913	4.08700	1.79	.61
Str. + ref.	13	45.6485	9.70758		
Unstr.	15	40.4473	14.38540		
Control	11	45.5073	3.06416		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Accuracy

Table 11 *Pre-writing Accuracy Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	.53	.07	6.90	.07
Str. + ref.	13	.18	.15		
Unstr.	15	.10	.15		
Control	11	.16	.09		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 12 *Post-writing Accuracy ANOVA Results*

Group	N	Mean	St. Deviation	f	Sig.
Str.	8	.0900	.12467	.72	.54
Str. + ref.	13	.1623	.09697		
Unstr.	15	.1547	.14798		
Control	11	.1164	.13231		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Fluency

Table 13 *Pre-writing Fluency Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	13.44	5.34	5.01	.17
Str. + ref.	13	13.30	2.18		
Unstr.	15	14.61	2.70		
Control	11	12.91	2.66		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 14 *Post-writing Fluency Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	13.7450	4.83405	3.07	.38
Str. + ref.	13	12.5877	2.74682		
Unstr.	15	12.9093	2.24236		
Control	11	11.7264	2.41706		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Holistic Rating:

Table 15 *Pre-writing Holistic Rating Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	2.13	1.24	1.48	.68
Str. + ref.	13	2.31	1.10		
Unstr.	15	2.60	1.29		
Control	11	2.27	.78		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 16 *Post-writing Holistic Rating Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	2.00	1.309	.25	.96
Str. + ref.	13	2.23	1.301		
Unstr.	15	2.20	1.082		
Control	11	2.18	1.079		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Problem-solution Structure Rating:

Table 17 *Pre-writing Problem-solution Structure Rating Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	1.75	.65	1.92	.58
Str. + ref.	13	1.69	.43		
Unstr.	15	1.90	.89		
Control	11	2.00	.45		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 18 *Post-writing Problem-solution Structure Rating Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	8	1.875	.7440	.08	.99
Str. + ref.	13	1.769	.6330		
Unstr.	15	1.867	.7432		
Control	11	1.773	.5179		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Appendix E: Texts to Teach Reformulation

Original Text

All of the people know drugs but I don't know why they sometimes accepted them. In my opinion some the people don't have information about drugs or they are curiosity about them and they want to try them but that is so stupidly while they are know about harmful of drugs, use it again.

Drugs are harmful for families and sociaty. In depends on people who addicted of drugs, they insulting their family and others because they influnced other people sometimes. These people are symbol for their children and their children continue father's way. People who are addicted to drugs don't have good communication withe their families and others and always they isolated from others.

In my opinion, polices are very important to stop these people but important that polices should catch drugs leaders because they are like (Farsi word for root) and everyday they are growing and expanding every where. Some of people needs money and they have financial problems They solve their problem by drugs so government should try correct Financial problem that people don't have reason for their task.

Public education is very important, too. We can use commercail and programs to show how drugs are influnced our body and learn that it is very useless things in our life. Awarness about drugs is important problem Drugs are dangrous all of people who live world. I wish we could solve these problem as soon as possiable.

Reformulated Text:

Everyone knows what drugs are, but I cannot understand why people sometimes use them. In my opinion, some people try drugs because they either do not have information about them or because they are curious. However, I think it is stupid when they use drugs again, having discovered their harmful effects. Other people use drugs because they need money and have financial problems, and they try to escape these issues by using drugs.

Drugs are harmful for families and societies. This is because addicts insult their families and others. Sometimes, these people are examples for their children and their children follow them. People who are addicted to drugs do not have a good relationship with their families and others, and thus, are always socially isolated.

There are some ways to solve this problem. In my opinion, the police play an important role in stopping addicts, but it is more important to stop drug dealers because they are the roots of the problem and they are growing in numbers as time passes. They exist everywhere.

Another way to stop this problem is that the government should try to solve poor people's financial problems, so that they do not have an excuse to use drugs.

Public education is very important, too. People should be informed, through TV commercials and programs, about how drugs affect human bodies and know that they are totally useless. Awareness about drugs is very important. Drugs are dangerous for everyone in the world. I hope these problems are solved as soon as possible.

Appendix F: Example of a Reformulated Text

Original Text:

Drugs are one of the big problem in whole country. There are lots of information around drugs and their effects and we can see or hear from everywhere. But people especialy young people don't believe that drugs have harmful effects on them and their life. On top of that youth use drugs more than older people for some reason such as curiosity and lack of awareness. Becouse young people think drugs are just for fun and they can prevent themselves from using them.

Anyway, drugs are used in everywhere and it makes problems for family and society that we can't hid it. Drug addiction also cause bad effect and one of them that society face with it, is finacial problem because government must spend lots of money to treat drug addicted. On the other hand family is important group to think about it. Because drug addition destroy family too.

Also we face with problem, but we can prevent it to not spred. Police-manpower is big group to stop drug dealers and drug user too. In addition, in my idea it's not useful to arrest them because it just has financial burden for governement to destroy drug and treat them.

One big efficient solution is public education. It is important to teach people when they are students. They must are teached around drugs and how danger it is. Awarness from everythings can prevent from big problem.

Reformulated Text:

Drug are one of the biggest problems in the country. There is lots of information about drugs and their effects, which we can see and hear everywhere. However, people, especially young people, do not believe that drugs have harmful effects on their bodies and their lives. On top of that, youths use drugs more than older people because of curiosity and a lack of awareness. They think drugs are just for fun and that they will not become addicted.

Drugs are used everywhere and this has very bad effects on families and societies, effects that cannot be ignored. One of those effects that societies are faced with is the financial burden drug use imposes on governments because governments must spend lots of money to treat drug addicts. Family is also an important community to be considered because drug addiction destroys families, too.

Although we are faced with this problem, we can prevent it from spreading. The police have the power to stop drug dealers as well as users. However, in my opinion, it is not useful to arrest them because destroying drugs and treating users imposes a financial burden on governments.

One efficient solution is public education. It is important to teach students. They should be taught about drugs and their dangers. Awareness can prevent big problems.

Appendix G: Main Study Group Differences

Table 1 *Pre-writing Structural Complexity Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	27	1.54	.27	6.20	.10
Str. + ref.	28	1.71	.36		
Unstr.	26	1.61	.53		
Control	23	1.72	.43		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 2 *Pre-writing Lexical Complexity ANOVA Results*

Group	N	Mean	St. Deviation	f	Sig.
Str.	27	7.75	.88	.46	.70
Str. + ref.	28	7.48	.84		
Unstr.	27	7.57	.96		
Control	24	7.53	.84		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 3 *Pre-writing Grammatical Accuracy ANOVA Results*

Group	N	Mean	St. Deviation	f	Sig.
Str.	27	.34	.14	.47	.70
Str. + ref.	28	.32	.11		
Unstr.	27	.35	.14		
Control	24	.29	.16		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 4 *Pre-writing Mechanical Accuracy Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	27	.26	.17	3.20	.36
Str. + ref.	28	.32	.17		
Unstr.	27	.29	.15		
Control	24	.31	.13		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 5 *Pre-writing Fluency1 ANOVA Results*

Group	N	Mean	St. Deviation	f	Sig.
Str.	27	213.19	51.63	1.13	.34
Str. + ref.	28	213.82	64.71		
Unstr.	27	242.19	71.21		
Control	24	224.50	76.25		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 6 *Pre-writing Fluency2 Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	27	12.89	3.12	2.38	.49
Str. + ref.	27	13.99	4.29		
Unstr.	26	13.04	3.84		
Control	24	13.07	2.38		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 7 *Pre-writing Idea Expression Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	26	3.55	.71	5.98	.11
Str. + ref.	28	3.71	.51		
Unstr.	24	3.35	.72		
Control	23	3.78	.42		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 8 *Pre-writing Paragraph Division and Focus Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	27	3.12	.75	2.38	.49
Str. + ref.	28	2.87	.10		
Unstr.	27	2.74	.91		
Control	24	2.83	1.10		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 9 *Pre-writing Presence and Order of Elements Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	27	2.50	.76	8.50	.04*
Str. + ref.	28	2.89	.90		
Unstr.	27	2.39	.90		
Control	24	2.20	.76		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

*Significance reached

Table 10 *Pre-writing Development of Elements Kruskal-Wallis Test Results*

Group	N	Mean	St. Deviation	χ^2	Sig.
Str.	27	1.61	.67	4.47	.21
Str. + ref.	28	1.58	.67		
Unstr.	27	1.38	.58		
Control	24	1.29	.44		

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Appendix H: Main Study T1, T2, and T3 Descriptive Statistical Results

Table 1 *Descriptive Statistics for T1, T2, and T3 Structural Complexity*

Group	N	T1		T2		T3	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	1.46	.19	1.54	.29	1.59	.24
Str. + ref.	28	1.51	.17	1.52	.27	1.54	.17
Unstr.	27	1.53	.18	1.61	.21	1.54	.21

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 2 *Descriptive Statistics for T1, T2, and T3 Lexical Complexity*

Group	N	T1		T2		T3	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	7.88	.77	7.81	.97	8.07	.90
Str. + ref.	28	8.05	.75	7.78	.71	8.14	.77
Unstr.	27	7.96	.86	7.93	.69	8.28	.82

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 3 *Descriptive Statistics for T1, T2, and T3 Grammatical Accuracy*

Group	N	T1		T2		T3	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	.26	.10	.26	.09	.34	.11
Str. + ref.	28	.30	.11	.30	.08	.31	.10
Unstr.	27	.31	.14	.28	.09	.30	.14

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 4 *Descriptive Statistics for T1, T2, and T3 Mechanical Accuracy*

Group	N	T1		T2		T3	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	.24	.15	.28	.15	.29	.14
Str. + ref.	28	.27	.16	.33	.17	.30	.17
Unstr.	27	.30	.17	.30	.19	.25	.17

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 5 *Descriptive Statistics for T1, T2, and T3 Fluency1*

Group	N	T1		T2		T3	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	228.00	47.14	225.58	40.90	250.85	74.47
Str. + ref.	28	258.88	63.91	194.58	54.70	287.18	94.74
Unstr.	27	258.72	57.25	263.40	64.40	295.32	66.30

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 6 *Descriptive Statistics for T1, T2, and T3 Fluency2*

Group	N	T1		T2		T3	
		Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	12.19	2.71	12.47	2.33	12.13	8.07
Str. + ref.	28	12.51	2.17	12.31	2.15	11.94	5.89
Unstr.	27	12.78	2.21	13.45	2.24	11.86	5.60

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Appendix I: Mixed ANOVA Results for Research

Question 3

Structural Complexity

Table 1. *Descriptive Statistics for Pre-W – Post-W Structural Complexity*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	26	1.51	.22	1.56	.31
Str. + ref.	28	1.71	.36	1.56	.23
Unstr.	26	1.53	.30	1.58	.23
Control	22	1.61	.23	1.61	.26

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 2. *Mixed ANOVA for Structural Complexity*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	.01	1	.01	.13	.71
Group	.35	3	.11	1.24	.29
Time × Group	.34	3	.11	1.88	.13

Lexical Complexity

Table 3. *Descriptive Statistics for Pre-W – Post-W Lexical Complexity*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	7.75	.88	7.79	.83
Str. + ref.	27	7.50	.85	7.73	.79
Unstr.	25	7.65	.88	7.65	.79
Control	24	7.53	.84	7.50	.80

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 4. *Mixed ANOVA for Lexical Complexity*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	.17	1	.17	.65	.41
Group	1.68	3	.56	.49	.69
Time × Group	.51	3	.17	.67	.57

Grammatical Accuracy

Table 5. *Descriptive Statistics for Pre-W – Post-W Grammatical Accuracy*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	26	.33	.12	.28	.11
Str. + ref.	28	.32	.11	.31	.12
Unstr.	27	.35	.14	.29	.11
Control	23	.36	.17	.33	.09

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 6. *Mixed ANOVA for Grammatical Accuracy*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	.07	1	.07	6.82	.01*
Group	.05	3	.02	.89	.44
Time × Group	.01	3	.00	.53	.66

* Significance reached

Mechanical Accuracy

Table 7. *Descriptive Statistics for Pre-W – Post-W Mechanical Accuracy*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	.26	.17	.26	.14
Str. + ref.	26	.30	.15	.30	.15
Unstr.	27	.29	.15	.28	.16
Control	24	.31	.12	.30	.13

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 8. *Mixed ANOVA for Mechanical Accuracy*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	.00	1	.00	.01	.92
Group	.06	3	.02	.68	.56
Time \times Group	.00	3	.00	.02	.99

Fluency 1 (number of words per text)**Table 9.** *Descriptive Statistics for Pre-W – Post-W Fluency 1*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	26	211.81	52.15	217.00	49.62
Str. + ref.	26	205.23	54.01	243.69	64.12
Unstr.	24	223.92	49.86	231.12	55.39
Control	23	215.35	63.06	203.86	49.37

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 10. *Mixed ANOVA for Fluency 1*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	4785.31	1	4785.31	3.64	.06
Group	10189.17	3	3396.39	.71	.54
Time \times Group	16215.17	3	5405.05	4.12	.01*

* Significance reached

Fluency 2 (number of words per T-unit)**Table 11.** *Descriptive Statistics for Pre-W – Post-W Fluency 2*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	25	12.39	2.67	12.07	1.62
Str. + ref.	26	13.46	3.33	12.86	2.11
Unstr.	26	12.46	2.46	12.15	2.18
Control	22	12.69	1.95	12.53	2.47

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 12. *Mixed ANOVA for Fluency 2*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	5.97	1	5.97	1.40	.23
Group	27.35	3	9.12	1.23	.30
Time \times Group	1.25	3	.41	.10	.96

Idea Expression

Table 13. *Descriptive Statistics for Pre-W – Post-W Idea Expression*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	3.46	.85	3.92	.26
Str. + ref.	28	3.71	.51	3.96	.13
Unstr.	27	3.09	1.01	3.77	.57
Control	24	3.66	.70	3.95	.20

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 14. *Mixed ANOVA for Idea Expression*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	9.42	1	9.42	28.81	.00*
Group	5.47	3	1.82	4.31	.00*
Time \times Group	1.57	3	.52	1.60	.19

* Significance reached

Paragraph Division and Focus

Table 15. *Descriptive Statistics for Pre-W – Post-W Paragraph Division and Focus*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	3.12	.75	2.11	1.05
Str. + ref.	28	2.87	.99	2.35	1.01
Unstr.	27	2.74	.91	2.57	1.19
Control	24	2.83	1.10	3.14	.99

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 16. *Mixed ANOVA for Paragraph Division and Focus*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	6.38	1	6.38	6.41	.01*
Group	4.82	3	1.60	1.54	.20
Time \times Group	12.13	3	4.04	4.06	.00*

* Significance reached

Presence of Problem-solution Structure Elements

Table 17. *Descriptive Statistics for Pre-W – Post-W Presence of Elements*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	2.50	.75	2.87	.81
Str. + ref.	28	2.89	.89	3.01	.61
Unstr.	27	2.38	.90	3.24	.57
Control	24	2.20	.76	2.33	.76

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 18. *Mixed ANOVA for Paragraph Division and Focus*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	7.15	1	7.15	13.25	.00*
Group	13.21	3	4.40	6.82	.00*
Time \times Group	4.69	3	1.56	2.89	.03*

* Significance reached

Development of Problem-solution Structure Elements

Table 19. *Descriptive Statistics for Pre-W – Post-W Development of Elements*

Group	N	Pre-writing		Post-writing	
		Mean	Std. Deviation	Mean	Std. Deviation
Str.	27	1.61	.67	1.53	.60
Str. + ref.	28	1.58	.66	1.87	.71
Unstr.	27	1.38	.57	1.64	.70
Control	24	1.29	.44	1.62	.55

Note. Str. = Structured, Str. + ref. = Structured + reformulation, Unstr. = Unstructured

Table 20. *Mixed ANOVA for Paragraph Division and Focus*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	2.13	1	2.13	6.07	.01*
Group	2.20	3	.73	1.69	.17
Time \times Group	1.38	3	.46	1.31	.27

* Significance reached