



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 EFFECTS OF WRITTEN FEEDBACK AND
REVISION-FOCUS DIRECTION ON STUDENT
REVISION AND WRITING IMPROVEMENT**

HUA GENG

A THESIS SUBMITTED IN FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY (PH.D.) IN EDUCATION
THE UNIVERSITY OF AUCKLAND, AUCKLAND
NEW ZEALAND, 2017

ABSTRACT

Despite the copious studies on written corrective feedback (CF), a number of issues in the field of written feedback and L2 writing remain under-researched, including L2 student writing development in content and rhetorical aspects (Goldstein, 2004), the effect of feedback on student autonomous revision skills (Bitchener & Ferris, 2012), and the moderation of contextual factors (Ellis, 2010). This thesis reports on a study that addresses the aforementioned research gaps by drawing on L1 revision theories. It investigates the effects of written feedback type and revision-focus manipulation on the revision and writing development of upper-intermediate English-as-a-foreign-language (EFL) learners in China. The mediation of textual level (i.e., discourse and linguistic levels) is also examined.

The study was conducted in a Chinese university. Seventy-seven first-year non-English-major undergraduates participated in this five-week project which involved a treatment stage as well as pre- and post-treatment stages. In each of the pre- and post-treatment stages, the participants wrote an argumentative essay and revised it autonomously. In the treatment stage, they were first assigned to four treatment groups formed by the manipulation of two factors: written feedback type (identification + diagnosis vs. identification + solution) and revision-focus manipulation (\pm revision-focus direction), and one control group who received no feedback or revision-focus direction. The participants then completed three writing-revising tasks on three controversial topics. The design allowed for the analysis of the immediate treatment effect on revision, the treatment effect on autonomous revision and new writing.

Results revealed a positive immediate effect of treatment on revision. The effect was more evident for content, organisation, and grammatical accuracy and less evident for lexical accuracy. Written feedback type did not make a difference. Discourse-level text quality benefited more from the treatment with revision-focus direction, while grammatical accuracy benefited more from the treatment without revision-focus direction.

Results concerning autonomous revision demonstrated an overall ineffectiveness of written feedback, irrespective of feedback type or textual level. However, treatment with revision-focus direction enhanced student autonomous revision on discourse levels.

Results concerning new writing indicated the effectiveness of treatment for organisation quality improvement and its overall ineffectiveness for accuracy increase, the advantage of diagnostic feedback in improving content quality and grammatical accuracy, and the advantage of + revision-focus direction in developing content and organisation quality.

To Sam and Sherry

ACKNOWLEDGEMENTS

I am most indebted to my main supervisor Professor Lawrence Jun Zhang, who has offered me continuous help and support throughout this PhD journey. His open-mindedness and clear guidance has allowed me to proceed quite smoothly with much free space. His feedback on my work has always been enlightening and encouraging. More importantly, Professor Zhang has encouraged my confidence in academic pursuit. His work ethic has been a constant role model.

I am deeply grateful to my cosupervisor Dr Rosemary Erlam. Her critical feedback and constructive suggestions to my thesis proposal helped to reshape the research design into a feasible project. Her meticulous comments on my thesis draft enlightened me about clear writing.

Dr Shintani Natsuko deserves my gratitude for her valuable statistical advice on data processing.

I would like to thank the staff and students of the Chinese university for their participation in this study and co-operation during the data collection.

I would also like to thank my research friends, the fellow PhD candidates, who gave me useful feedback on my study. The shared journey was less painful because of the mutual support.

On a more personal note, special thanks go to my parents-in-law for their taking good care of my daughter when my husband and I were not available.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	xvii
LIST OF TABLES	xix
LIST OF ACRONYMS	xxiv
CHAPTER ONE INTRODUCTION	1
1.1 Overview	1
1.2 The Research Context	1
1.2.1 English learning and teaching in Chinese education system	1
1.2.2 EFL writing instruction in Chinese tertiary institutes	3
1.2.2.1 Feedback practices in writing instruction	4
1.3 Literature Related to the Area of Teacher Written Feedback on Student Writing	5
1.3.1 Definition of key terms	5
1.3.2 Key issues and gaps in research on teacher written feedback	6
1.3.3 Theoretical frameworks applied to the research area	8
1.4 Scope of This Study and Major Research Questions	9
1.5 Significance of the Study	9
1.6 An Outline of the Thesis	11
CHAPTER TWO LITERATURE REVIEW	12
2.1 Overview	12
2.2 Argument Overview	12

2.3 Establishment of Revision as a Legitimate Research Area.....	13
2.4 Theoretical Frameworks.....	14
2.4.1 Conceptual components.....	14
2.4.1.1 Twofold constructs of revision	14
2.4.1.2 Conceptualisation and operationalisation of revision for this research .	15
2.4.2 Major models of revision.....	15
2.4.2.1 The Scardamalia and Bereiter (1983) CDO procedure.....	16
2.4.2.2 The Flower et al. (1986) cognitive process model of revision	17
2.4.2.3 The Hayes (1996) cognitive process model of revision.....	18
2.4.2.4 The Butterfield et al. (1996) cognitive-metacognitive model of revision.....	19
2.4.2.5 Summary and critique of the revision models	20
2.5 Research Approaches of This Study	21
2.6 L1 and L2 Studies on Feedback and Revision	22
2.6.1 Specification of terms used in the following review of empirical studies....	22
2.6.2 Research on teacher written feedback in relation to L1 student revision and writing improvement	23
2.6.2.1 The effect of teacher written feedback on student revision and writing improvement	23
2.6.2.2 Summary of the review of teacher feedback and L1 student revision...27	
2.6.3 Research on teacher written feedback in relation to L2 student revision and writing improvement	28
2.6.3.1 The effect of written CF	28
2.6.3.2 The effect of content feedback	37
2.7 L1 and L2 Studies on Revision Performance in Relation to Allocation of Cognitive Resources	39
2.7.1 L1 revision studies from the perspective of cognitive resources.....	40

2.7.2 L2 revision studies from the perspective of cognitive resources.....	42
2.7.3 The approach of this study to studying revision in relation to cognitive resources	42
2.7.3.1 Theories about task complexity	43
2.7.3.2 Wickens’s (2002, 2007) multiple resource model	43
2.7.3.3 Studies related to information processing on meaning versus form	46
2.8 Summary of Chapter Two	47
2.8.1 Major findings	47
2.8.2 Major gaps in the literature	48
2.8.3 Research perspectives of this study	49
CHAPTER THREE METHODOLOGY	50
3.1 Overview	50
3.2 The Postpositivist Worldview	50
3.3 Research Questions	50
3.3.1 RQ1 Does treatment have any immediate effect on student revision?.....	51
3.3.2 RQ2 Does treatment have any effect on student autonomous revision?	51
3.3.3 RQ3 Does treatment have any effect on student new writing?	52
3.4 Factors of Treatment and Generation of Treatment/Control Groups.....	52
3.4.1 WF type	52
3.4.2 Direction of student revision-focus	57
3.4.3 Treatment/Control groups.....	58
3.5 Dependent Variables	58
3.5.1 Between-draft text quality change	59
3.5.2 Text quality of new writing	59
3.6 The Relationships between Independent and Dependent Variables	59

3.7 Design of the Study	60
3.8 Validating the Instruments and Procedures for Data Collection and Data Analysis	61
3.8.1 Participants in the pilot session	62
3.8.2 Instruments in the pilot session	62
3.8.3 The data collection procedure in the pilot session.....	64
3.8.4 Data analysis in the pilot session.....	65
3.8.4.1 Assessing quality of content	65
3.8.4.2 Assessing quality of organisation	66
3.8.4.3 Assessing grammatical accuracy and lexical accuracy.....	66
3.8.5 Changes to the main study.....	67
3.9 The Main Study	68
3.9.1 Population and sample.....	68
3.9.2 Instruments	70
3.9.2.1 Questionnaire	70
3.9.2.2 Writing-revising tasks.....	70
3.9.3 The procedure for data collection.....	73
3.9.3.1 Grouping the participants.....	75
3.9.3.2 The writing session for the five tasks	76
3.9.3.3 Giving feedback for Tasks 2–4.....	76
3.9.3.4 The revising session for the five tasks	80
3.9.4 Data analysis.....	81
3.9.4.1 Textual analysis of drafts	81
3.9.4.2 Statistical analysis.....	84
3.9.4.3 Post hoc textual analysis of drafts to address RQ1	87

CHAPTER FOUR RESULTS (1): THE IMMEDIATE EFFECT OF TREATMENT ON REVISION	89
4.1 Overview	89
4.2. Results for Content Quality.....	91
4.2.1 The immediate effect of group conditions on content quality across drafts.	91
4.2.2 The immediate effect of WF type and \pm RFD on content quality across drafts	95
4.2.3 Summary of results for content quality	99
4.3 Results for Organisation Quality.....	102
4.3.1 The immediate effect of group conditions on organisation quality across drafts	102
4.3.2 The immediate effect of WF type and \pm RFD on organisation quality across drafts	106
4.3.3 Summary of results for organisation quality	109
4.4 Results for Grammatical Accuracy	112
4.4.1 The immediate effect of group conditions on grammatical accuracy across drafts	112
4.4.2 The immediate effect of WF type and \pm RFD on grammatical accuracy across drafts	116
4.4.3 Summary of results for grammatical accuracy	120
4.5 Results for Lexical Accuracy	122
4.5.1 The immediate effect of group conditions on lexical accuracy across drafts	122
4.5.2 The immediate effect of WF type and \pm RFD on lexical accuracy across drafts	124
4.5.3 Summary of results for lexical accuracy	127
4.6 Summary of Results in Relation to the Immediate Effects of Group Conditions on Text Quality across Drafts.....	129

4.6.1 The effects of treatment conditions compared with the control condition .	129
4.6.2 The mediation of textual level in the effects of group conditions	130
4.7 Summary of Results in Relation to the Immediate Effects of the Factors of WF Type and \pm RFD on Text Quality across Drafts.....	133
4.7.1 The effect of WF type and the mediation of textual level	133
4.7.2 The effect of \pm RFD and the mediation of textual level	134

CHAPTER FIVE RESULTS (2): THE EFFECT OF TREATMENT ON AUTONOMOUS REVISION136

5.1 Overview	136
5.2 Results for Content Quality	138
5.2.1 The effect of group conditions on autonomous revision: Content gain score across stages	138
5.2.2 The effect of WF type and \pm RFD on autonomous revision: Content gain score across stages	140
5.2.3 Summary of results for content quality	143
5.3 Results for Organisation Quality	143
5.3.1 The effect of group conditions on autonomous revision: Organisation gain score across stages	143
5.3.2 The effect of WF type and \pm RFD on autonomous revision: Organisation gain score across stages	145
5.3.3 Summary of results for organisation quality	147
5.4 Results for Grammatical Accuracy	148
5.4.1 The effect of group conditions on autonomous revision: Grammatical accuracy gain score across stages	148
5.4.2 The effect of WF type and \pm RFD on autonomous revision: grammatical accuracy gain score across stages	150
5.4.3 Summary of results for grammatical accuracy	152

5.5 Results for Lexical Accuracy	153
5.5.1 The effect of treatment conditions on autonomous revision: Lexical accuracy gain score across stages	153
5.5.2 The effect of WF type and \pm RFD on autonomous revision: Lexical accuracy gain score across stages	155
5.5.3 Summary of results for lexical accuracy	156
5.6 Summary of Results Regarding the Effect of Treatment on Autonomous Revision in Terms of Text Quality across Drafts	157
5.6.1 The effect of group conditions and the mediation of textual level	157
5.6.2 The effect of WF type and \pm RFD and the mediation of textual level	157
CHAPTER SIX RESULTS (3): THE EFFECT OF TREATMENT ON NEW WRITING.....	158
6.1 Overview	158
6.2 Results for Content Quality in New Writing.....	159
6.2.1 The effect of group conditions on content quality in new writing	159
6.2.2 The effects of WF type and \pm RFD on content quality in new writing	161
6.2.3 Summary of results for content quality	164
6.3 Results for Organisation Quality in New Writing.....	166
6.3.1 The effect of group conditions on organisation quality in new writing	166
6.3.2 The effects of WF type and \pm RFD on organisation quality in new writing.....	168
6.3.3 Summary of results for organisation quality	172
6.4 Results for Grammatical Accuracy in New Writing	175
6.4.1 The effect of group conditions on grammatical accuracy in new writing ..	175
6.4.2 The effect of WF type and \pm RFD on grammatical accuracy in new writing	177
6.4.3 Summary of results for grammatical accuracy	180

6.5 Results for Lexical Accuracy in New Writing	181
6.5.1 The effect of group conditions on lexical accuracy in new writing	181
6.5.2 The effects of WF type and \pm RFD on lexical accuracy in new writing	182
6.5.3 Summary of results for lexical accuracy	184
6.6 Summary of Results Regarding the Effect of Treatment on Text Quality in New Writing	185
6.6.1 The effect of group conditions and the mediation of textual level	185
6.6.2 The effects of WF type and \pm RFD and the mediation of textual level	186
CHAPTER SEVEN DISCUSSION	189
7.1 Overview	189
7.2 The Immediate Effect of Treatment on Student Revision	189
7.2.1 Overall effects of treatment methods in comparison to the control condition	189
7.2.2 The effect of treatment for text quality on discourse levels	190
7.2.2.1 Overall effectiveness of treatment methods for both content and organisation	190
7.2.2.2 Treatment being more effective for organisation	192
7.2.2.3 The relative effects of the treatment methods	193
7.2.2.4 The effect of treatment in relation to WF type	194
7.2.2.5 The effect of treatment in relation to \pm RFD	195
7.2.3 The effects of treatment methods for text quality on linguistic levels	197
7.2.3.1 Overall effectiveness of treatment for grammatical accuracy and overall ineffectiveness of treatment for lexical accuracy	197
7.2.3.2 The different effects of the treatment methods	198
7.2.3.3 The effect of treatment in relation to WF type	199
7.2.3.4 The effect of treatment in relation to \pm RFD	201

7.3 The Effect of Treatment on Student Autonomous Revision	203
7.3.1 The effect on autonomous revision concerning text quality on discourse levels	204
7.3.1.1 Overall effects of treatment methods	204
7.3.1.2 The effect of treatment in relation to WF type	204
7.3.1.3 The effect of treatment in relation to \pm RFD	205
7.3.2 The effect on autonomous revision concerning linguistic accuracy.....	205
7.3.2.1 The effect of treatment in relation to WF type and \pm RFD.....	206
7.3.3 Ineffectiveness of feedback for autonomous revision	206
7.4 The Effect of Treatment on Text Quality in Student New Writing	207
7.4.1 The effect for content quality in new writing	208
7.4.1.1 The effect of WF type	208
7.4.1.2 The effect of \pm RFD	209
7.4.2 The effect for organisation quality in new writing	209
7.4.2.1 The effect of WF type	211
7.4.2.2 The effect of \pm RFD	211
7.4.3 The effect on grammatical accuracy in new writing	211
7.4.3.1 The effect of WF type	213
7.4.3.2 The effect of \pm RFD	214
7.4.4 The effect for lexical accuracy in new writing	214
7.4.4.1 The effects of WF type and \pm RFD.....	215

CHAPTER EIGHT CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH 216

8.1 Overview	216
8.2 Major Findings	216
8.2.1 The immediate effect of treatment on revision.....	216

8.2.2 The effect of treatment on autonomous revision	216
8.2.3 The effect of treatment on new writing	217
8.2.3.1 Content quality in new writing	217
8.2.3.2 Organisation quality in new writing	217
8.2.3.3 Grammatical accuracy in new writing	218
8.2.3.4 Lexical accuracy in new writing	218
8.3 Theoretical Implications.....	218
8.3.1 The relevance of L1 revision theories to understanding the issue of teacher feedback and L2 revision.....	218
8.3.2 Corroborating L1 revision theories with L2 evidence.....	220
8.4 Empirical Insights	221
8.4.1 The immediate effects on revision.....	222
8.4.2 The effects on autonomous revision.....	223
8.4.3 The effects on new writing	224
8.5 Methodological Contribution	225
8.6 Practical Implications	225
8.6.1 Informing teachers of effective feedback practice	226
8.6.2 Fostering self-sufficient revisers	227
8.7 Limitations	229
8.8 Recommendations for Further Research.....	230
8.9 Conclusions	231
REFERENCES.....	232
APPENDIX A: ETHICS DOCUMENTS.....	250
APPENDIX B: INSTRUMENTS FOR COLLECTING DATA.....	259
APPENDIX C: INSTRUMENTS FOR ANALYSING DATA.....	260

APPENDIX D: SAMPLE DRAFTS AND FEEDBACK	270
APPENDIX E: SCORES IN TASK 1 AND TASK 5	273

LIST OF FIGURES

Figure 3.1 A schematic diagram of the treatment/control group conditions	58
Figure 4.1 Mean content scores for Drafts 1 and 2 over Tasks 2–4 by group condition	93
Figure 4.2 Mean content scores for Drafts 1 and 2 over Tasks 2–4 by WF type	97
Figure 4.3 Mean content scores for Drafts 1 and 2 over Tasks 2–4 by \pm RFD	97
Figure 4.4 Mean organisation scores for Drafts 1 and 2 over Tasks 2–4 by group condition	104
Figure 4.5 Mean organisation scores for Drafts 1 and 2 over Tasks 2–4 by WF type	107
Figure 4.6 Mean organisation scores for Drafts 1 and 2 over Tasks 2–4 by \pm RFD ...	107
Figure 4.7 Mean grammatical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by group condition	114
Figure 4.8 Mean grammatical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by WF type.....	117
Figure 4.9 Mean grammatical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by \pm RFD.....	117
Figure 4.10 Mean lexical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by group condition	123
Figure 4.11 Mean lexical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by WF type	125
Figure 4.12 Mean lexical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by \pm RFD	126
Figure 5.1 Mean content gain scores in S1 and S3 by group condition.....	139
Figure 5.2 Mean content gain scores in S1 and S3 by WF type and \pm RFD	141
Figure 5.3 Mean organisation gain scores in S1 and S3 by group condition.....	144
Figure 5.4 Mean organisation gain scores in S1 and S3 by WF type and \pm RFD	146
Figure 5.5 Mean grammatical accuracy gain scores in S1 and S3 by group condition	149

Figure 5.6 Mean grammatical accuracy gain scores in S1 and S3 by WF type and \pm RFD	151
Figure 5.7 Mean lexical accuracy gain scores in S1 and S3 by group condition	154
Figure 5.8 Mean lexical accuracy gain scores in S1 and S3 by WF type and \pm RFD .	155
Figure 6.1 Mean first-draft content scores over Tasks 1/2–5 by group condition.....	160
Figure 6.2 Mean content scores of the first draft over Tasks 1/2–5 by WF type.....	162
Figure 6.3 Mean content scores of the first draft over Tasks 1/2–5 by \pm RFD.....	163
Figure 6.4 Mean first-draft organisation scores over Tasks 1/2–5 by group condition	167
Figure 6.5 Mean first-draft organisation scores over Tasks 1/2–5 by WF type	169
Figure 6.6 Mean first-draft organisation scores over Tasks 1/2–5 by \pm RFD	170
Figure 6.7 Mean first-draft grammatical accuracy scores over Tasks 1/2–5 by group condition	176
Figure 6.8 Mean first-draft grammatical accuracy scores over Tasks 1/2–5 by WF type	178
Figure 6.9 Mean first-draft grammatical accuracy scores over Tasks 1/2–5 by \pm RFD	179
Figure 6.10 Mean first-draft lexical accuracy scores over Tasks 1/2–5 by group condition	181
Figure 6.11 Mean first-draft lexical accuracy scores over Tasks 1/2–5 by WF type .	183
Figure 6.12 Mean first-draft lexical accuracy scores over Tasks 1/2–5 by \pm RFD	183

LIST OF TABLES

Table 1.1 Chinese education system	2
Table 3.1 Aspects of discourse- and linguistic-level problems to be targeted by feedback	54
Table 3.2 Relationships between independent and dependent variables	60
Table 3.3 Global design of the study	61
Table 3.4 Correspondence of the writing-revising tasks to the research questions	61
Table 3.5 Data collection procedure for the pilot session.....	65
Table 3.6 Number of participants in each group condition	70
Table 3.7 Data collection procedure for the main study.....	74
Table 3.8 Linguistic error categories targeted for feedback	77
Table 3.9 Criteria and examples for providing ID and IS feedback on content	78
Table 3.10 Criteria and examples for providing ID and IS feedback on organisation	79
Table 4.1 Means and standard deviations for content quality in Tasks 2–4 by draft and group condition	92
Table 4.2 Effects of draft and group condition on content quality in Tasks 2–4.....	93
Table 4.3 Means and standard deviations for content quality in Tasks 2–4 by draft, WF type, and \pm RFD	96
Table 4.4 Effects of draft, WF type, and \pm RFD on content quality in Tasks 2–4	98
Table 4.5 Significant within-subjects contrasts of the group conditions regarding content quality across drafts in Tasks 2–4.....	100
Table 4.6 Significant between-subjects contrasts across the group conditions regarding content quality in Tasks 2–4	101
Table 4.7 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding content quality in Tasks 2–4	102
Table 4.8 Mean and standard deviations for organisation quality in Tasks 2–4 by draft and group condition	103

Table 4.9 Effects of draft and group condition on organisation quality in Tasks 2–4	104
Table 4.10 Means and standard deviations for organisation quality in Tasks 2–4 by draft, WF type, and \pm RFD	106
Table 4.11 Effects of draft, WF type, and \pm RFD on organisation quality in Tasks 2–4	108
Table 4.12 Significant within-subjects contrasts of the group conditions regarding organisation quality across drafts in Tasks 2–4	110
Table 4.13 Significant between-subjects contrasts across the group conditions regarding organisation quality in Tasks 2–4	111
Table 4.14 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding organisation quality in Tasks 2–4	112
Table 4.15 Means and standard deviations for grammatical accuracy ^a in Tasks 2–4 by draft and group condition	113
Table 4.16 Effects of draft and group condition on grammatical accuracy in S2	114
Table 4.17 Means and standard deviations for grammatical accuracy ^a in Tasks 2–4 by draft, WF type, and \pm RFD	116
Table 4.18 Effects of draft, WF type, and \pm RFD on grammatical accuracy in Tasks 2–4	118
Table 4.19 Significant within-subjects contrasts of the group conditions regarding grammatical accuracy across drafts in Tasks 2–4	120
Table 4.20 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding grammatical accuracy in Tasks 2–4	121
Table 4.21 Means and standard deviations for lexical accuracy ^a in Tasks 2–4 by draft and group condition	122
Table 4.22 Effects of draft and treatment condition on lexical accuracy in Tasks 2–4	124
Table 4.23 Means and standard deviations for lexical accuracy ^a in Tasks 2–4 by draft, WF type, and \pm RFD	125
Table 4.24 Effects of draft, WF type, and \pm RFD on lexical accuracy in Tasks 2–4	127

Table 4.25 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding lexical accuracy across drafts in Tasks 2–4	128
Table 4.26 Significant within-subjects contrasts of the group conditions regarding text quality across drafts in Tasks 2–4	131
Table 4.27: Significant between-subjects contrasts across the group conditions regarding text quality in Tasks 2–4	132
Table 4.28 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding text quality across drafts in Tasks 2–4	134
Table 4.29 Significant between-subjects contrasts across WF types and \pm RFD conditions regarding text quality in Tasks 2–4	135
Table 5.1 Means and standard deviations for the content gain score by stage and group condition	138
Table 5.2 Effects of stage and group condition on the content gain score	139
Table 5.3 Means and standard deviations for the content gain score by stage, WF type, and \pm RFD	141
Table 5.4 Effect of stage, WF type, and \pm RFD on the content gain score	142
Table 5.5 Means and standard deviations for the organisation gain score by stage and group condition	144
Table 5.6 Effects of stage and group condition on the organisation gain score	145
Table 5.7 Means and standard deviations for the organisation gain score by stage, WF type, and \pm RFD	146
Table 5.8 Effects of stage, WF type, and \pm RFD on the organisation gain score	147
Table 5.9 Means and standard deviations for the grammatical accuracy gain score by stage and group condition	149
Table 5.10: Effects of stage and group condition on the grammatical accuracy gain score	150
Table 5.11 Means and standard deviations for the grammatical accuracy gain score by stage, WF type, and \pm RFD	151

Table 5.12 Effects of stage, WF type, and \pm RFD on the grammatical accuracy gain score	152
Table 5.13 Means and standard deviations for the lexical accuracy gain score by stage and group condition	153
Table 5.14 Effects of stage and group condition on lexical accuracy gain score	154
Table 5.15 Means and standard deviations for the lexical accuracy gain score by stage, WF type, and \pm RFD	155
Table 5.16 Effects of stage, WF type, and \pm RFD on the lexical accuracy gain score	156
Table 6.1 Means and standard deviations for first-draft content quality by task and group condition	160
Table 6.2 Effects of task and group condition on first-draft content quality in Tasks 1/2–5.....	161
Table 6.3 Means and standard deviations for first-draft content quality by task, WF type, and \pm RFD.....	162
Table 6.4 Effects of task, WF type, and \pm RFD on first-draft content quality in Tasks 1/2–5	164
Table 6.5 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding first-draft content quality across Tasks 1/2–5.....	165
Table 6.6 Means and standard deviations for first-draft organisation quality by task and group condition	166
Table 6.7 Effects of task and group condition on first-draft organisation quality in Tasks 1/2–5	168
Table 6.8 Means and standard deviations for first-draft organisation quality by task, WF type, and \pm RFD	169
Table 6.9 Effects of task, WF type, and \pm RFD on first-draft organisation quality in Tasks 1/2–5	171
Table 6.10 Significant within-subjects contrasts of the group conditions regarding first-draft organisation quality across Tasks 1/2–5.....	172

Table 6.11 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding organisation quality across Tasks 1/2–5	174
Table 6.12 Means and standard deviations for first-draft grammatical accuracy in task and group condition	175
Table 6.13 Effects of task and group condition on first-draft grammatical accuracy in Tasks 1/2–5	177
Table 6.14 Means and standard deviations for first-draft grammatical accuracy by task, WF type, and \pm RFD	178
Table 6.15 Effects of task, WF type, and \pm RFD on first-draft grammatical accuracy in Tasks 1/2–5	179
Table 6.16 Means and standard deviations for first-draft lexical accuracy by task and group condition	181
Table 6.17 Effects of task and group condition on first-draft lexical accuracy in Tasks 1/2–5	182
Table 6.18 Means and standard deviations for first-draft lexical accuracy by task, WF type, and \pm RFD	183
Table 6.19 Effects of task, WF type, and \pm RFD on first-draft lexical accuracy in Tasks 1/2–5	184
Table 6.20 Significant within-subjects contrasts of the group conditions regarding first-draft text quality across Tasks 1/2–5	186
Table 6.21 Significant within-subjects contrasts concerning WF types regarding first-draft text quality across tasks 1/2–5	187
Table 6.22 Significant within-subjects contrasts concerning \pm RFD conditions regarding first-draft text quality across Tasks 1/2–5	187
Table 7.1 Revisions on discourse levels for Task 3 by WF type	195
Table 7.2 Revisions on discourse levels for Task 3 by \pm RFD	196
Table 7.3 Revisions on linguistic levels for Task 3 by WF type	199
Table 7.4 Revisions on linguistic levels for Task 3 by \pm RFD	203

LIST OF ACRONYMS

CF	corrective feedback
EFL	English-as-a-foreign-language
ESL	English-as-a-second-language
ID	identification + diagnosis
IS	identification + solution
L1	first language
L2	second language
RFD	revision-focus direction
SLA	second-language-acquisition
WF	written feedback

CHAPTER ONE

INTRODUCTION

1.1 Overview

This research on teacher intervention in student writing was undertaken in a Chinese university context. This chapter will first give an overview of English teaching in the Chinese education system. It will then elaborate on English writing instruction in Chinese tertiary institutes and explain how I generated an interest in the field of teacher feedback on student writing. Next, I will present key terms in this thesis and highlight gaps in related literature and the theoretical framework that led to the research questions for this study. After stating the significance of the study, I will provide an outline of the thesis.

1.2 The Research Context

1.2.1 English learning and teaching in Chinese education system

The Chinese education system comprises two broad stages: compulsory education and noncompulsory education (see Table 1.1). The former comprises six-year primary schooling and three-year junior secondary schooling. English is established as a core subject in junior secondary education, and some primary schools in major cities have started to incorporate English curriculum in the past decade (Ministry of Education of the People's Republic of China [MoE], 2001).

The education following junior secondary school is no longer compulsory and falls into academic and vocational directions. Students opting for an academic direction may attend three-year senior secondary schooling, where English is one of the three core subjects with the other two being Chinese and mathematics. The score of the English test constitutes a crucial component of the matriculation result obtained at the end of the senior secondary education. Academically oriented tertiary education comprises undergraduate, graduate, and doctoral programmes. For any of these programmes, English is normally a required course for one to two years. In these tertiary programmes non-English majors, who choose to major in a discipline different from English when applying to a university, take one or two English courses which focus on basic English language skills. Some institutes offer optional English courses, and some provide

tailored courses for certain majors, such as Medical English for medical students and Business English for business students. Non-English majors are required to sit College English Test (known as CET), a national standard test for undergraduates and postgraduates (College English Test Committee, n.d.), to obtain a testimony to their English proficiency. The condition is different for English majors, who learn English in an intensive manner throughout a certain programme by taking a variety of English courses, encompassing English literature, Western culture, as well as basic and advanced language skills. English majors take Test for English Majors (known as TEM) to obtain an acknowledged certificate of competency (Board of Foreign Language Teaching in Higher Education Institutions, 2000).

With respect to students heading in a vocational direction after junior secondary schooling, they may choose secondary vocational education which lasts two to three years and proceed to a three-year vocational college; or they may attend a five-year vocational college upon graduation from a junior secondary school. For these students, English is also important but the courses are of a lower difficulty level. There is also a national test targeting vocational college students. In general, English learning and teaching is a national undertaking in China (Zhang, 2013a).

Table 1.1 Chinese education system

Noncompulsory	Doctoral programme (4 years) Graduate programme (2–3 years) Undergraduate programme (4 years)	Higher vocational education (3 years)
	Senior secondary education (3 years)	Secondary vocational education (2–3 years)
Compulsory	Junior secondary education (3 years)	
	Primary education (6 years)	

As an English teacher for non-English-major undergraduates, I am interested in investigating English teaching for this group of students. A brief introduction to the English learning history of a current non-English-major undergraduate is provided as

follows. He/she is likely to start English learning in Grade Three in the primary school, have around four English lessons per week, and experience a communicative teaching style during this stage. During the following six-year secondary schooling, he/she takes on average five hours of English classes per week and receives more explicit instruction on linguistic forms although the communicative skills are still emphasised (Wen, 2009). Upon graduation from a senior secondary school, he/she should have received around 1,500 hours of English instruction. After entering a tertiary institute, the student enrolls in a compulsory English curriculum which comprises on average four-hour weekly lessons and lasts throughout the first or the first two academic years; he/she may also enrol in some elective English courses if available (Wang & Wang, 2011). In the meantime, he/she takes College English Test. It is notable that the above description is a rough picture and that English curricula across Chinese districts, schools, and institutes vary more or less due to the inequality in education resources.

1.2.2 EFL writing instruction in Chinese tertiary institutes

The English curriculum for non-English majors in Chinese tertiary institutes normally covers the teaching of listening, reading, speaking, and writing (Wang & Wang, 2011). Some universities offer a specific EFL writing course while some others embed writing instruction in a comprehensive course or a reading-and-writing course. Considered as an advanced skill, English writing is relatively less emphasised compared to the more basic skills, namely, listening, reading, and speaking (Wang, 2014; Wang & Wang, 2011).

In recent years, Chinese tertiary institutes have invested growing attention and energy in English writing instruction. This move, for one thing, responds to the promotion of second language writing as a construct and a tool of second language learning by the academic and pedagogical community (Zhang, 2013b), and for another, responds to Chinese students' relatively low proficiency in this aspect (Guo, 2015; Tang & Wu, 2012; Wang, 2014). In Chinese tertiary institutes new teaching approaches including the process-based approach, the content- or task-based approach, and the genre-based approach have been added to the traditional product-focused approach (Deng, Chen & Zhang, 2014; Jiang, 2014). The flipped-classroom teaching mode has been introduced to complement the traditional teacher-centred teaching mode (Wang, 2014). In addition, new media such as the Internet and corpora have been integrated into writing instruction,

serving purposes such as providing background readings and portfolio management (Yu, Qi & Guo, 2012). Furthermore, some institutes have started to offer academic English writing courses to prepare students for further academic studies (Zeng & Li, 2014). With respect to feedback and assessment, new endeavours have converged in the application of Automatic Essay Scoring, i.e., computer systems that generate scores and feedback on student essays (Yang & Dai, 2015).

1.2.2.1 Feedback practices in writing instruction

The commonly used feedback methods in writing instruction include teacher feedback, peer response, and computer-generated feedback. The last method is gaining increasing popularity mainly to respond to two challenges facing Chinese EFL teachers. One relates to the workload of teaching large classes or several small parallel classes. A teacher collecting over one hundred essays for one assignment would hardly do more than giving a simple mark to each student; or the teacher might ask students to do peer view. Another challenge for teachers is the lack of knowledge and strategies of providing feedback. Many teachers are inexperienced or ill-informed in giving feedback. Some only address language problems in student writing and overwhelm students with red marking all over the paper, and some give minimal comments by writing a couple of words at the end of an essay.

While the automated scoring systems meet the above two challenges to some extent, they bear their respective drawbacks, such as a failure to equally address grammar and discourse, occasional misunderstandings of student intentions, and a tendency to reward long text with high scores (Ge & Chen, 2007; Yang & Dai, 2015; Zhang, 2013). Therefore students still expect written feedback from teachers, on their language, content, and organisation, and regard it as the most reliable and authoritative (Zhang, 2006). This distance between student expectation and teacher inadequacy has aroused my interest in the area of teacher feedback on EFL student writing. In addition, as the most frequently used intervention method in L2 student writing, teacher written feedback deserves efforts to maximise its efficacy. It is worth investigating how teachers can provide helpful written feedback in an efficient manner. When equipped with relevant knowledge, teachers can better benefit student writing development. If their institutes invest in automatic scoring systems, they can guide students in the use of them and provide supplementary information instead of totally relying on automatic scoring.

With such an interest in teacher written feedback on student writing, I navigated related literature in both L1 and L2 contexts and identified some gaps worth further investigation.

1.3 Literature Related to the Area of Teacher Written Feedback on Student Writing

This section first provides the definition of key terms. It then presents important research gaps in the empirical literature. It finally displays the theoretical frameworks and explains their relevance to the focus of this research.

1.3.1 Definition of key terms

To facilitate the understanding of the following thesis, it is necessary to provide definitions of key terms used in this thesis in this introduction chapter. This subsection introduces “teacher written feedback”, “revision”, “autonomous revision”, “writing improvement”, “cognitive resources”, “revision-focus direction”, as well as “textual level”.

It needs to be clarified that in this research the term “teacher written feedback” includes feedback on linguistic accuracy, content, as well as organisation in student writing. Written CF referred to feedback on linguistic errors while global feedback meant feedback on content and organisation. The term “revision” has two referents (Fitzgerald, 1987; for further reference, see section 2.4.1). It may refer to the mental process of making changes to a piece of writing or a discrete-point change made to a piece of writing. It is used as an uncountable noun for the first referent and countable for the second referent. I use “autonomous revision” to refer to student revision of a new text without receiving external feedback. This concept has been termed as “ability to revise” (Adams, Simmons, Willis & Pawling, 2010; Yin, 1995) and “revision skills” (Butler & Britt, 2011; Carifio et al., 2001). It is opposed to student revision of a text with the help of teacher feedback on it. I use “writing improvement” or “writing development” to refer to student writing quality in a new piece of writing rather than in a revised text.

Manipulation of student cognitive resources, as part of the research design of this study, is termed “revision-focus direction”. The concept “cognitive resources” (or termed “resources”) means the mental energy available at a certain moment to carry out cognitive processes (Baddeley, 1986; Halford, Wilson & Phillips, 1998). The term

“textual level” refers to an aspect of a text (e.g., grammar, content, organisation). Some studies used terms such as “sentence-level”, “discourse-level”, “category”, and “scope” to classify the aspects of revision (e.g., Sommers, 1980; Wallace, Hayes, Hatch, Miller, Moser & Silk, 1996). I use “textual level” as the umbrella term for the convenience and relative clarity of reference. “Revision-focus” is meant to refer to the textual level on which students focus their cognitive resources during a revising task. To put it simply, it means what students focus on improving while revising. It seems that this construct has not been examined in the empirical literature although focus of feedback has been studied (e.g., Ashwell, 2000). “Revision-focus direction”, accordingly, refers to direction of student cognitive resources toward particular textual aspects during revision; the direction is given by me. The operationalisation of this term is to direct learners to focus their attention, among other resources, on global levels first and on linguistic levels later when revising their written text. The specific manipulation method is explained in detail in section 3.4.2.

1.3.2 Key issues and gaps in research on teacher written feedback

A review of the research on teacher written feedback on student writing has found that studies on teacher written feedback in the L1 context have tended to observe teacher written feedback in general. In other words, most of the studies have investigated both feedback on linguistic accuracy and feedback on content and organisation (e.g., Carifio, Jackson & Dagostino, 2001; Cho & MacArthur, 2010; Frey & Fisher, 2013; McCutchen, Francis & Kerr, 1997; Parr & Timperley, 2010; Peterson & Portier, 2014). However, most L2 studies on teacher written feedback on student writing have adopted a single focus.

Unsurprisingly, existent L2 studies on teacher written feedback seem to focus on either the effect of written CF on linguistic accuracy (e.g., Diab, 2015; Shintani & Aubrey, 2016) or the effect of global feedback on content and rhetoric development (e.g., Conrad & Goldstein, 1999; Goldstein, 2004). Few studies have included both perspectives in their investigation, and prominence has been given to written CF revealing an SLA-centred orientation (for a review, see Hyland & Hyland, 2006a). A single focus on either CF or global feedback in the domain of teacher written feedback research is unrealistic with what happens in reality, where students usually receive both CF and global feedback on their writings and make corresponding responses. Therefore,

findings based on a research design involving only CF or only global feedback might not reflect the reality of the classroom. It is significant to investigate both written CF and global feedback on L2 student writing. In response to a design incorporating both CF and global feedback, this study will examine the effect on student writing quality on separate textual levels that incorporate grammatical and lexical accuracy, content quality, and organisation quality.

The neglect of the effect of teacher written feedback on student autonomous revision has been identified as another research gap. Autonomous revision ability is essential for student writers and a primary goal of writing instruction (Bitchener & Ferris, 2012; Song & Ferretti, 2013). Therefore there have been abundant investigations into revision skills (e.g., Chanquoy, 2001; Yan, 2010). However, this issue seems to be long neglected in the area of teacher feedback research. The majority of research on teacher written feedback, be it in the L1 or the L2 context, has been concerned with the impact of feedback on student immediate revision of the annotated draft and/or the impact on student writing on a new prompt. Few studies have examined whether teacher feedback on student previous writing helps students revise their new writing with no feedback. In other words, little is known about whether teacher feedback fosters student autonomous revision skills. It is necessary to investigate their possible relationship.

A third major research gap in the teacher written feedback area lies in the limited investigation into the mediation of task conditions in the effect of teacher feedback on student writing. The existent studies on teacher written feedback, in the L2 context in particular, have generally focused on the differential types of feedback. Only a few studies touched upon other task conditions, such as the type of writing assignments (Riazantseva, 2012) and the timing of feedback (Shintani & Aubrey, 2016). The condition under which students conduct a writing task has been reported to exert a strong effect on student cognitive processes and the writing outcome (Ong, 2014; Robinson, 2011). It is therefore worthwhile to explore various conditions that may influence student cognitive processing during a writing/revising task and mediate student engagement of teacher feedback. In this research the designed task conditions involve directing versus not directing students to focus on certain aspects of the text while they were revising the text.

1.3.3 Theoretical frameworks applied to the research area

Previous studies on teacher CF on L2 student writing have mainly drawn on theories in second language acquisition (SLA) due to their important contribution to the SLA/L2 writing interface (Bitchener & Ferris, 2012; Ellis, 2010; Polio, 2012). SLA theories, however, may have limited application to the issues of global feedback on content and organisation. As there seems to be no comprehensive writing theories specifically developed for L2 writing, I reviewed writing theories (Hayes, 1996; Hayes & Flower, 1980, 1983) that were developed in the L1 context, finding alternative theoretical frameworks applicable to the research focus in question. These frameworks centre around theoretical models produced to illustrate the revising process in writing (e.g., Butterfield, Hacker & Albertson, 1996; Flower, Hayes, Carey, Schriver & Stratman, 1986; Hayes, 1996; Scardamalia & Bereiter, 1983). Therefore I refer to them as revision theories in this thesis.

Within these theoretical models, four essential components can be extracted: (1) the subprocesses fundamental to successful revision (detection, diagnosis, and solution of problems); (2) task definition or task schema that controls how the reviser will approach the revision task and activate relevant knowledge to complete the basic subprocesses; (3) the reviser's cognitive resources that play a role in the revising process; (4) the task environment which comprises a reviser's physical and social environmental factors (e.g., topic of text to be revised and target audience) and is external to the reviser's cognition. The four components interact with one another throughout the revising process. Those propositions have the potential of helping to understand the efficacy of external feedback for feedback-assisted revision and autonomous revision.

According to the theories, the efficacy of teacher written feedback relates to whether the information it provides can facilitate the proposed essential revising subprocesses. Helpful feedback is supposed to contain information useful for conducting the revising subprocesses. The usefulness of teacher feedback can also be examined in terms of its influence on student task schema, which in turn accounts for their revision based on feedback and on their autonomous revision. On the other hand, the propositions regarding cognitive resources can account for the effect of task conditions on revisers' cognitive processing and their use of teacher feedback during the revising process.

Hence, the revision theories have great relevance to the effect of teacher written feedback on student revision and autonomous revision.

1.4 Scope of This Study and Major Research Questions

In general, this study intends to investigate the effect of teacher written feedback on student revision and writing development. To address the research gaps mentioned above, it looks at both feedback on language problems and feedback on content and organisation. Furthermore, it examines the effect of feedback on student revision of writing based on teacher feedback, student autonomous revision, and student new writing as well. Drawing on the revision theories, this research compares types of feedback according to the type of information (identification, diagnosis, and solution of problems) feedback offers to assist the revising process. It also examines whether task conditions regarding manipulation of student cognitive resources during revision mediates the effect of teacher written feedback. Specifically, this study addresses the following two major questions:

- (1) Does written feedback type have any effect on student revision and writing development?
- (2) Does revision-focus direction have any effect on student revision and writing development?

1.5 Significance of the Study

This study is needed for the following reasons: a) promoting an integrated approach to studying feedback on student writing that considers the reviser's cognitive factors and task environmental factors, b) widening the perspectives adopted in research on feedback and revision, c) enhancing the efficiency of teacher feedback to L2 student writing, and d) raising teacher and learner awareness of the limited cognitive resources available during revision. The former two have implications for theory and research, while the latter two are of pedagogic value.

As stated in the above, previous studies on teacher feedback to L2 student writing have mainly drawn on SLA theories. As far as I know, no study has exploited L1 revision theories to approach this issue. A close examination of revision theories reveals that the

propositions regarding the essential revising processes, the reviser's cognition, and the task environmental factors help to explain the efficacy of external feedback for revision and writing improvement. From another perspective, in L2 studies comparing the effects of different teacher written feedback types, the typology of feedback methods has been inconsistent across studies, and the same term has seen different ways of conceptualisation and operationalization (Ellis, 2009; Liu & Brown, 2015). The research findings are consequently not comparable in some cases. Hence this research develops a more rigorous typology of written feedback based on revision theories. This approach has implications for introducing new theories into the L2 research on teacher feedback and revision in writing.

Another significance of the study is that it broadens the perspectives in research on feedback and revision. Existent L2 studies on this issue tend to focus on either the effect of CF on accuracy or the effect of global feedback on content and rhetoric development (Hyland & Hyland, 2006a) but not bring the two perspectives together. Furthermore, few studies have examined the effect of teacher feedback on student revision of new writing with no feedback (i.e., autonomous revision), although many have delved into its effect on student immediate revision of the annotated draft and the written accuracy when students write on a new topic. It is worth knowing how teacher feedback can raise student writers' awareness of their problems and become self-sufficient revisers. This study will expand the existent research perspectives by overcoming these two gaps in research on feedback and student revision.

This study has practical value in addition to theory and research-related significance. The academic and pedagogical community has promoted second language writing as a construct of and a tool of second language learning (Zhang, 2013b). Teacher written feedback, as the most frequently used intervention in L2 student writing, deserves efforts to maximise its efficacy. This study expects to find out the effects of different feedback types in relation to discourse and linguistic problems. Such findings may provide implications for pedagogy as to how to tailor feedback to different dimensions of student written text. This investigation also intends to compare the effects of teacher feedback on the revised text and on new writing, and the findings may offer insights into how to prolong the efficacy of teacher feedback to L2 student writing.

Finally, this study may raise teacher and learner awareness of student limited cognitive resources during revision and the impact of task conditions on student cognitive processing. Revision theories propose that revision is a process subject to the cognitive demands of the multiple sub-processes and the reviser's cognitive resources (Butterfield et al., 1996; Hayes, 1996). L1 empirical evidence indicates that revisers' limited cognitive resources play a role in the revising process and result. Whether it is the same with L2 student writers is under researched. Anticipating that the revision task may be more cognitively taxing for L2 students who are not linguistically competent as their L1 counterparts, I plan to approach L2 student revision from this cognitive perspective by designing two differential task conditions regarding the manipulation of revision-focus. The findings are expected to raise teacher and learner awareness in this respect. Teachers may in turn consciously direct student attention in a favourable way during task design, and students may modify their revision strategies to achieve better results.

1.6 An Outline of the Thesis

The thesis consists of eight chapters. Chapter Two overviews L1 revision theories against which some of the perspectives of this study were developed, and it then reviews empirical studies that investigated revision in relation to teacher written feedback or student allocation of cognitive resources in response to revision task conditions. Chapter Three describes the methodology for the main study as well as the method and outcome of a pilot session. The next three chapters (Chapters Four, Five, and Six) respectively report results in response to the three sets of research questions. Chapter Seven discusses the results reported in the preceding chapters. The final chapter, Chapter Eight, presents theoretical and practical implications derived from the research findings and offers recommendations for future research.

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

This chapter will start by explaining what I will argue through the literature and why. It will then outline the background of research on revision in the L1 context and explicate the theoretical frameworks; specific revision models will be presented following an account of conceptual issues in revision research. Next, it will review the empirical research in L1 and L2 contexts with regard to the effect of written feedback on revision. Following that, it will survey the studies on the role of the task condition in the allocation of cognitive resources during revision. Finally, a summary of the chapter is provided outlining the major findings and research gaps in the literature.

2.2 Argument Overview

This argument overview serves to explain the function of the following sections in this chapter, mainly how they relate to one another and how they shed light on my research focus and methodology. Section 2.3 is aimed to provide a brief background to the L1 research on revision that has in turn promoted the development of revision theories which are used as theoretical frameworks for this study. Section 2.4 attempts to present the theoretical frameworks including the conceptual components related to revision and major models of revision. An account of the conceptual elements (section 2.4.1) facilitates the understanding of the term “revision” in the empirical literature as well as in theoretical models. In addition, those constructs informed me with respect to the conceptualisation and operationalization of revision in the methodology of my research. While presenting the models of revision, section 2.4.2 explains and compares the key elements of them, and furthermore illustrates how these theoretical propositions relate to the research approach and research design of this study. Section 2.5 explains how my interest in teacher written feedback and revision theories have contributed to research approaches of this research. Sections 2.6 and 2.7 review empirical studies relevant to the research questions of this research. The final section (2.8) summarises major findings of related empirical literature as well as major gaps in it, and supplies a detailed account of the research perspectives of this research.

2.3 Establishment of Revision as a Legitimate Research Area

As introduced in Chapter One, this study aims to address the effects of teacher written feedback and manipulation of student cognitive resources on the English writing of Chinese university students. It mainly draws on L1 revision theories to build the theoretical framework for the research project. Before the theories are presented, development in L1 research on revision is briefly provided below. It can be seen that it was the development in the theoretical discussions about writing that fostered research on revision.

The establishment of revision as a legitimate research area was shaped by Hayes and Flower's (1980) cognitive process model of writing and Murray's (1978) new thinking on revision. In the then new trend toward process-oriented writing instruction, Hayes and Flower (1980) proposed a cognitive process model to theorise the writing process. Their framework identifies three basic processes: planning, translating, and reviewing within the writing process, which operate under the control of a monitor. In marked contrast to previous models, it depicts planning, translating, and reviewing as hierarchically organized, with certain processes embedded within others. In other words, planning, translating, and reviewing do not proceed step by step, but operate in a flexible and recursive manner. With regard to reviewing, it is not explained as a unique stage, but as a thinking process that can occur whenever writers examine their text or plans, and this thinking process can constantly lead to new planning and translating. Hayes and Flower's work has brought new insights into the writing process and greatly contributed to later research on writing and revising (Alamargot & Chanquoy, 2001).

The thinking on revision evolved along with the historical theoretical discussions about writing. Murray (1978) proposes that writing is a discovery process; but unlike some researchers, who emphasise the importance of prewriting for idea formation (e.g., Britton, Burgess, Martin & Rosen, 1975), he believes that continuous revision is the key to moving toward meaning. A host of researchers hence started to explore the territory of revision both theoretically and empirically.

2.4 Theoretical Frameworks

Theoretical frameworks for this research are discussed in two aspects: conceptual components and major models of revision. The discussion of the former attempts to foreshadow the twofold constructs of revision, which have both been studied in the literature and were both examined in this research. The display of the theoretical models explains and compares the key elements of them, and furthermore illustrates how these theoretical propositions relate to the research approach and research design of this study.

2.4.1 Conceptual components

2.4.1.1 Twofold constructs of revision

The term “revision” can refer to both the mental revising process and the execution resulting from this process (Allal, Chanquoy & Largy, 2004). In respect of the mental aspect of revision, researchers primarily think of it as a three-step problem solving process. This conception has been reflected in Bartlett’s (1982) detection-identification-modification outline, Scardamalia and Bereiter’s (1983) compare-diagnose-operate plan, and Flower et al.’s (1986) detection-diagnosis-strategy selection scheme. The shared essence of these procedures is that, while revising, writers begin with evaluating their written (or mental) text or a writing plan and detecting problems either in produced text or original goals, next diagnose problems determining what can be or need be improved or modified, and finally operate and carry out actual changes.

Besides the three basic elements involved in the revision process, researchers have perceived other aspects concerning the nature of the revising process. Sommers (1980) argued that recognition of “dissonance”, i.e., discrepancies between writers’ intention and execution, is the key to revision. Later researchers have added that revision is also a process of evaluating and clarifying goals and thoughts (Chanquoy, 2001; McCutchen et al., 1997). Further, revision can be merely activated by discovery through the process of writing while having nothing to do with problem detection (Galbraith, 1992; Hayes & Flower, 1986). Moreover, researchers have increasingly recognised that revision is a cognitive process subject to writers’ short- and long-term memory as well as a task’s cognitive demands, and influenced by revisers’ cognitive and metacognitive factors (Beal, 1996; Butterfield et al., 1996; McCutchen, 1996, 2000; Olive, Kellogg & Piolat, 2001; Ong, 2014; Scarmalia & Bereiter, 1983).

Another dimension to examine revision concerns the physical changes made to the text, and researchers have designed different criteria to categorise those changes. These taxonomies encompass types of revision operation (e.g., deletion, substitution, addition), textual levels (e.g., syntactic, discourse) on which changes are made, revision purposes (e.g., transitional, informational, cosmetic, stylistic), revision effect (e.g., correct, erroneous, neutral; or meaning-preserving versus meaning-changing), amount of attention engaged (e.g., automatically released, controlled), and occasions (e.g., in-draft, between-draft) (Allal, 2000; Bridwell, 1980; Faigley & Witte, 1981; Monahan, 1984; Sommers, 1980).

2.4.1.2 Conceptualisation and operationalisation of revision for this research

In the exploration of teacher written feedback and revision-focus direction on student revision in writing, this research mainly examined the concept of revision in terms of its physical product aspect. With respect to its operationalisation, this research approached revision in terms of the overall change in text quality caused by the discrete changes students made to their original text and the textual level on which such change in quality occurred (e.g., the overall change in grammatical accuracy between the original and the revised drafts). When analysing a segment of the data, this research operationalised revision with respect to the discrete physical changes, examining the number of the changes and the textual level each change involves.

This research, on the other hand, probed the mental process aspect of revision. Although it did not record or measure such cognitive process, it attempted to interpret the effects of feedback and revision-focus direction on the physical revision results by linking to students' revising process. The role of students' cognitive factors in the revising process is a focus of discussion.

2.4.2 Major models of revision

With the increasing attention to revision from the 1980s, several specific models of revision have been constructed. These models are the Scardamalia and Bereiter (1983) compare/diagnose/operate (CDO) procedure, the Flower et al. (1986) model of cognitive processes in revision, the Hayes (1996) model of cognitive processes in revision, and the Butterfield et al. (1996) modernised cognitive-metacognitive version of the revision model by Flower et al. (1986). These models dynamically illustrate how the

subprocesses of revision function and interact with each other and with other factors. In the following review, the key elements of the models are identified. Furthermore, how these theoretical propositions relate to the focus and methodology of this research is explicated.

2.4.2.1 The Scardamalia and Bereiter (1983) CDO procedure

Scardamalia and Bereiter (1983) designed a compare/diagnose/operate (CDO) procedure to help children with revision, and it has been viewed as a tool of understanding the revising process. In this framework, writers keep two types of representation in mind while writing: representation of intended text and representation of actual text produced. First, writers compare and evaluate the two representations, searching for discrepancies between them. When a mismatch is recognised, they diagnose the nature of the problem and consider possible solutions. Next, they choose to change the original plan or reserve the plan and operate on the written text. Finally, “operate” is realised through two steps: selecting a strategy (e.g., deleting, rewriting) to solve the problem followed by generating new text. When a CDO cycle is completed, the resulting mental and textual changes enter the mental representation, leading to a new cycle of CDO. The CDO procedure may stop halfway, for instance, when no mismatch is sensed, when diagnosis fails, when a plan is changed, and when writers do not know how to generate a better text.

The value of this framework lies in that it embodies three basic subprocesses of the revising process, i.e., evaluating and searching for the problem, diagnosing the problem and considering possible solutions, and operating on the problem, and details the mechanism of each of them. The proposed basic subprocesses can be related to the effect of teacher feedback on student revision. According to the framework, feedback facilitating the three basic subprocesses of revising is supposed to be helpful for successful revision.

Arguably, the extent of the effect of feedback also relies on revisers’ own processing of the text to be revised, their cognitive resources (e.g., what knowledge and skills they already have and what they lack) and their processing of feedback (e.g., deep processing versus surface processing). This framework is confined to the central revising process

without incorporating relevant factors of and beyond the reviser. Such a limitation was overcome by the subsequent models of revision, as introduced below.

2.4.2.2 The Flower et al. (1986) cognitive process model of revision

Flower et al. (1986) developed an elaborate model of revision which has two major advancements. First, it describes the evaluation procedure in a more complicated way though the core process still follows a detection-diagnosis-strategy selection pattern like the CDO scheme (Scardamalia & Bereiter, 1983). Second, it integrates factors from outside of the revision process itself, i.e., knowledge and task definition, and places the central evaluation procedure in interaction with these factors.

As to the subprocesses of evaluation, three types of reading purposes are distinguished. Reading to comprehend is the least demanding; some revisions may occur reactively if the writer feels something that hinders comprehension. Reading to evaluate involves more intentional assessment and deploys more goals or criteria; problem detection falls into this type of reading. Finally, reading to define problems goes beyond assessment to diagnosis. Flower et al. (1986) further point out that “Evaluation itself can lead to the discovery of new possibilities, not just errors ... [it] operates not only on written text, but on mental text and even plans” (p. 25). The detected problems may range from being ill- to well-defined based on the degree to which the reviser understands the problems. Next, five strategies are identified in the reviser’s actions for the represented problems: ignoring the problem if it is of little worth or too much difficulty, delaying action, searching for a better representation of the problem by further evaluation, rewriting, and revising.

This model of revision further proposes that the whole revision processes depend on the factor of the reviser’s knowledge. The “knowledge” component includes not only declarative knowledge about the feature of a problem but procedural knowledge as to how to address the problem. Therefore, a successful revision necessitates active usable knowledge. The proposition in this aspect can shed light on the effect of teacher feedback on student revision. The differential effects of feedback type relates to type of knowledge students need or lack most. Compared with subsequent models, the “knowledge” component roughly corresponds to the “long-term memory” in both Hayes’s (1996) and Butterfield et al.’s (1996) revision models.

This model highlights the importance of task definition, i.e., the writer's plans for revision, e.g., to work on word level, syntactic or discourse level, to revise immediately or after reading the whole text, what goals to achieve and what criteria to apply. It is posited that the revision activity starts with task definition, which shapes the way of evaluation. This idea concerning task definition may help to interpret student revisers' engagement of feedback during their revising process. For instance, a student with little intention to improve the organisation of the written text might overlook the relevant comment offered by the teacher. This task definition component of the framework may also provide explanations for the influence of feedback on student autonomous revision. Since task definition largely determines the revision outcome with other individual factors controlled, feedback on appropriate task definition is supposed to guide students' autonomous revision. The concept of task definition is equivalent to task schema in Hayes's (1996) model of revision.

2.4.2.3 The Hayes (1996) cognitive process model of revision

Based on related research findings since 1980, Hayes developed an updated model of writing with a submodel for revision. This new revision model comprises three parts: the control structure, fundamental processes, and resources in working memory and long-term memory. The fundamental processes are similar to the "Processes" component in Flower et al.'s (1986) framework and include text processing, reflection, and text production. What controls the fundamental processes is a "task schema", defined as "a package of knowledge, acquired through practice, that is useful for performing the task and is retrieved as a unit when cues indicating the relevance of the schema are perceived" (Hayes, 1996, p. 16). It might include a goal, expected activities, attentional subgoals, criteria for quality, and strategies for fixing text problems, and it determines how the basic processes are triggered and sequenced. Therefore, "task scheme" echoes the concept of task definition in the Flower et al. (1986) model. They both refer to a plan or a monitor controlling how revisers approach a revision task.

In addition, this model recognises that revision is a cognitively demanding process influenced by available resources in working memory and long-term memory. This is a major new development in research on revision. Drawing on Baddeley's (1986) work, this model conceptualises working memory as a limited resource that is exploited both for storing information and for performing cognitive processes which are not automated.

In addition, working memory is postulated to comprise a central executive, a phonological loop, and a visual-spatial sketchpad, each of them having distinctive functions in storing and processing information. Long-term memory is where information can be stored for a long period. Knowledge (e.g., knowledge of audience, vocabulary, and strategies of problem solving; see Flower et al., 1986) stored in long-term memory can be activated and retrieved into working memory to participate in the revising process.

The proposed limited resource in working memory stimulated me to investigate the role of revision-focus direction which I designed as a specific task condition. As revision is cognitively demanding process, it is worthwhile to pursue if different conditions would demand different amounts of resources and in turn result in different cognitive processing and revision results.

2.4.2.4 The Butterfield et al. (1996) cognitive-metacognitive model of revision

Butterfield et al.'s revision model comprises two major parts: the "Environment" and the "Cognitive/Metacognitive System". The former includes the "rhetorical problem" (i.e., the topic, the audience and the importance of the text to be revised) and the "actual text being revised" (characterised by its format, genre, etc.). These factors are all external to the reviser.

The latter falls into two subsystems: working memory and long-term memory. Globally, the elements in the "environment" and those in the cognitive/metacognitive system interact constantly during the revising process. The merit of this model is its elaboration on how these systems of memory function in relation to the revising process (Chanquoy, 2009). It is notable that it specifies the roles of working memory and long-term memory, and distinguishes metacognition from cognition during the revising activity (cf. Zhang, 2010).

Butterfield et al. posit that within working memory occur the fundamental revising processes: (1) representing the rhetorical problem, planning, and deciding on standards of evaluation (similar to Flower et al.'s "task definition" and Hayes's "task schema"); (2) reading to represent the text; (3) detection and diagnosis; (4) strategy selection for revising text; (5) translation from mental revision to actual text.

Long-term memory is described as a cognition-metacognition dual system. The cognition component keeps knowledge (about the text topic, language and writing, and standards of evaluation), strategies (related to thinking, reading, and writing), and a representation of the text being revised. The metacognition component contains the same knowledge and strategies but in a more complete, systematic and synthesised form (termed as “models of knowledge” and “understanding of strategies”) and therefore determines “when, where, how, and why it is necessary to use, evaluate, and control cognitive strategies and cognitive knowledge” (Chanquoy, 2009, p. 85). In other words, metacognition can guide the procedural deployment of cognition. Further, elements within long-term memory function in a relatively automatic manner with little mental energy required to perform certain cognitive processes (Baddeley, 1986, 1996), and as a result the knowledge and strategies retrieved from long-term memory can facilitate the revising processes and free the resources in working memory.

2.4.2.5 Summary and critique of the revision models

The frameworks presented above, either subsumed under more general models of writing or specifically conceived for revision, have demonstrated an evolution in order:

firstly, to describe as precisely as possible the different processes involved during revision, secondly, to consider the great difficulty of this activity, necessitating a writer’s and a reviser’s careful reading of the text in order to detect some errors or problems, and thirdly, to introduce long-term memory, working memory, and a metacognitive system in order to dynamically explore the revision process. Globally, revision is unanimously considered as a very complex process, weighing heavily on writers’ attention and limited capacities in working memory. (Chanquoy, 2009, p. 86)

The gradual evolution of the revision models has two major merits. While the CDO procedure only captures the fundamental processes of revision, the later models (Butterfield et al., 1996; Flower et al., 1996; Hayes, 1996) have adopted a holistic scheme incorporating factors both inside and outside of the reviser that influence the fundamental processes of revision in an interactive way. Another strength of the recent models lies in their sophistication in the modelling of cognitive factors.

These models have a few weaknesses, though. While highlighting factors on the cognitive level, they neglect the reviser’s affective elements, such as motivation and anxiety. In addition, they do not elaborate contextual factors that may impact the

revising process (cf. Ellis, 2010). The macrocontext that may be brought into the models involves the educational background of the reviser (e.g., age, educational level, and cultural background). The microcontext to be considered may include, in addition to those listed in the Butterfield et al. model, revising with or without external resources from peers, teachers, or readings, revising on the computer or on paper, revising a text in one's native language or a text in one's second language.

Having surveyed both conceptual knowledge and theoretical models in the domain of revision, I will outline the research approaches of this study and then review the related empirical literature.

2.5 Research Approaches of This Study

Research approaches of this study are, on one hand, motivated by my interest in the area of teacher written feedback on student writing, and on the other hand, grounded in propositions of revision theories. The above review of revision theories indicates the importance of revision, as revision leads to rediscovery of meaning, reorganisation of ideas, and refinement of language and style. Therefore, teachers need to help students better revise their writing. Various studies have been conducted on instructional intervention methods (e.g., Reynolds & Bonk, 1996; Song & Ferretti, 2013; Wigglesworth & Storch, 2012). Since I pursue the area of teacher written feedback in view of students' expectations for it and the incompetency in offering feedback among Chinese EFL teachers (see section 1.2.2), this study attempts to investigate how to help students revise their text with written feedback, the common intervention mode in writing instruction.

According to the aforementioned revision theories, detection, diagnosis, and solution are the three basic elements in the revising process. Availability of all the three elements is normally the prerequisite to successful revision. Lack of or failure to activate knowledge concerning one or all of these elements may hinder revision. If these speculations are linked to written feedback—the common instructional intervention mode—a question arises that provision of what type of knowledge would most greatly benefit EFL students' processing of different textual levels during revision: knowledge about location of a problem, nature of the problem, or solution to the problem.

A follow-on question is whether the effect on revision of written feedback providing different types of information will transfer to students' new pieces of writing and unassisted revision of their new writing. In other words, this study attempts to explore whether different types of input through feedback will alter students' long-term memory in relation to writing and revising. This enquiry is worth pursuing since the ultimate goal of EFL writing instruction is to foster self-sufficient writers and revisers (Bitchener & Ferris, 2012).

Another theory-motivated query of this study pertains to the issue of cognitive demand during revision. Since revision is proposed as a cognitively demanding process with multiple subprocesses such as evaluative reading, diagnosis, and transcribing, a question is whether manipulating allocation of cognitive resources will affect EFL students' revising process and results. Investigations along this dimension look into the distribution of cognitive resources during revision, enquiring whether students would produce different performance when revising under a manipulated condition. The related terms and considerations for validity will be explained in section 2.7.

The following two sections (2.6 and 2.7), therefore, will respectively survey studies touching upon the effect of written feedback in facilitating detection, diagnosis, and solution, and those that have approached revision from a viewpoint of allocation of resources in response to cognitive demands.

2.6 L1 and L2 Studies on Feedback and Revision

This section will first outline L1 writing studies, specifically, those involving teacher feedback providing information about detection, diagnosis, solution to assist student revision. Then with the same foci this section will review studies in L2 settings.

2.6.1 Specification of terms used in the following review of empirical studies

In the following review of empirical studies, "revision" may refer to physical changes made to a text, the whole process of performing a revising task, the consequent quality change in a revised version of the original text. Although it may also refer to a mental process (see section 2.4.1.1), that referent is not included in the following studies to be reviewed. The term "autonomous revision" refers to student revision of a new text (either written by students themselves or preset by researchers) without receiving

feedback. “New writing” and “write anew” both mean writing on a new topic as opposed to rewriting on the same topic. “Writing improvement” or “writing development” refers to text quality of a piece of new writing rather than of a revised text.

2.6.2 Research on teacher written feedback in relation to L1 student revision and writing improvement

Most L1 studies on teacher feedback have surveyed the features of authentic teacher comments or student response to teacher comments (e.g., Beach & Friedrich, 2006; Connor & Lunsford, 1993; Lunsford & Lunsford, 2008; Stern & Solomon, 2006), while limited research has delved into the effect of teacher feedback on student revision and writing improvement (Chanquoy, 2009). The following review will centre on the studies that examined the impact of teacher feedback on revision or writing improvement.

The following studies carry various perspectives and purposes, and some of them do not specifically look at the effect of teacher feedback on revision, but they all offer insight into the questions intended to be addressed in my research. The studies are roughly grouped according to their relevance to two questions: (a) Does teacher written feedback facilitate student revision? and (b) What type of teacher written feedback facilitates student revision? A summary will follow the survey of the individual studies, commenting on their findings and methodological limitations.

2.6.2.1 The effect of teacher written feedback on student revision and writing improvement

Haswell (1983) found that marginal check-marks indicating presence of mechanic and grammatical errors could prompt college students to successfully self-edit 60% of the errors noted. Beach (1979) found that high school students receiving teacher evaluation made significantly higher “degree of change” than the other self-evaluation groups and produced higher quality of argument. It is suggested that teacher feedback helps students detect the problems in their writing, who otherwise may have difficulty in critically assessing their own writing.

In Ziv’s (1982) study, however, the college students made few between-draft changes as a result of teacher feedback. Qualitative data revealed that the students were unable to make effective revisions although they sensed the dissonance between their

representation of their texts and that of the teacher's. Ziv argued that, to make comments beneficial, teachers should not only indicate the problems in students' texts but suggest strategies to solve them.

In their study on the effect of written comments on autonomous revision of preset texts, Carifio et al. (2001) compared the revision skills of the experimental group (who revised the texts based on diagnostic and prescriptive comments) with that of the control group (who revised the same texts with no external comments), finding that over a three-week treatment both groups significantly improved their revising skills, but the experimental group improved to a larger extent on effective major changes. They did not investigate the efficacy of diagnostic and prescriptive comments separately, though.

2.6.2.1.1 Type of teacher written feedback, mediation of textual level and learner educational level

Some researchers compared the effects of different types of teacher feedback on student revision and writing improvement, and some others further touched upon the mediation of textual level and learner educational level. Hillocks (1982) studied the effects of treatment on text quality (with respect to writing specifically with a precise focus) in new writing produced by 7th- and 8th-graders. The treatment involved two types of feedback: brief comments (less than 10 words, one or more suggestions for improvement) and long comments (well over 10 words, one or more very specific suggestions for improvement). Findings showed that the effect of the two types of comment interacted with two other factors, i.e., the presence or absence of prewriting activities and revision. Sperling and Freedman's (1987) case study discovered that comments which referred to previous classroom information were easy to be understood and correctly responded to while those with no reference to previous instruction were of little help. It suggested that young students needed sufficient information to understand comments and revise effectively.

In another study looking at feedback type, Parr and Timperley (2010) linked pupils' long-term writing achievement to their teachers' written feedback practice. The results indicate that teacher feedback most conducive to writing development supplies students with features of the desired performance, students' distance to the desired performance,

and how to achieve the performance. This study suggested that constructive feedback needed to provide information about diagnosis and solution.

Ziv (1984) looked at both the type of teacher comment and textual levels addressed by teacher comment when she examined four college freshmen's reactions to her comments on their writing. Ziv first developed a three-category taxonomy to categorise her comments on both macro- (conceptual and structural) and micro- (sentential, lexical, grammar, and mechanics) levels of student writing: explicit cues, implicit cues, and direct corrections. Then she coded her comments on participants' drafts based on this taxonomy. She found that on the macro textual level the participants made better revisions in response to her explicit comments and benefited from her implicit comments under certain conditions, whereas on the micro textual level the participants did not react favourably to her implicit comments since they often did not understand the nature of the indicated problems. The majority of her direct corrections were made on the levels of grammar, spelling, and punctuation, and they seemed unhelpful in the long run because the students' think-aloud responses indicated that the students tended to accept the corrections without recognising the reason. Explicit cues, which offered diagnosis, enabled the students to self-correct their errors in grammar and spelling.

Based on these results, Ziv posits that it is necessary for inexperienced revisers to receive explicit cues about how to revise on both the macro and micro levels. She adds that teachers may move toward the implicit end of the commenting continuum when students are more experienced revisers. The research results, however, may be partial because they were only based on Ziv's own comments while Ziv seemed to be likely to heavily use one type of comment to address a certain textual level instead of balancing the three types of comment on each textual level concerned. Another limitation is that the study did not assess the between-draft text quality change (i.e., change in the global quality of a certain textual level, such as content, organisation, and word choice) in relation to the participants' revisions (i.e., individual physical changes, such as addition, deletion, substitution, on different textual levels).

Plumb, Butterfield, Hacker and Dunlosky (1994) also investigated the effects of different types of external help on student revision and the mediation of textual level, but, different from the preceding studies reviewed in this section, their study asked learners to correct implanted errors other than their own errors. The participants revised

implanted errors under three conditions: errors being unmarked, errors marked, and errors marked plus correction suggested. They made significantly more revisions, particularly for surface errors, when errors were located for them. The researchers concluded that revisers owned the knowledge but could not activate it, and once the errors were pointed out, they could revise a large part of them successfully. This conclusion is open to question, however, considering that only 10 surface errors and 10 meaning errors were embedded in each of the two texts used, and so they could not cover many subcategories of error type.

Similarly, McCutchen et al. (1997) examined the effects of error identification (i.e., locating the errors for students) on students' revision of implanted surface errors (misspellings) and meaning errors (illogical sequence). Findings showed that while error identification helped college students to revise both the surface and the meaning errors, it restricted middle school students to processing surface problems. It seemed that information about error location played different roles for students of different ages and levels.

Cho and MacArthur (2010) examined undergraduate students' revisions in relation to the type of feedback from peer and expert reviewers and their writing quality improvement in relation to revision type. The researchers found that the expert provided more directive feedback whereas the peers more nondirective feedback. Analysis showed that directive feedback was positively related to simple repairs of mechanics, and that nondirective feedback significantly predicted complex repair revisions and new content revisions. Further analysis showed that only complex repairs positively predicted writing quality improvement (measured in terms of prose flow, argument, and insight) across drafts. Cho and MacArthur pointed out that the subject-matter expert's directive comments on macro textual levels might be difficult to understand for the participants although direct, specific comments had been reported to be helpful for revisers. They also found that the participants' macro-level changes such as new content revisions and structural changes were often unsuccessful, which explained why they did not predict better text quality. A merit of this study is that it not only looked at how feedback affected revisions but assessed how revisions affected writing quality in the revised draft.

2.6.2.2 Summary of the review of teacher feedback and L1 student revision

The above survey of L1 studies reaches some conclusion and reveals some questions for further investigation. In general, teacher feedback plays an important role in providing information concerning detection, diagnosis, and solution of problems in student writing. Teacher feedback conducive to better revisions tends to be explicit, easy to understand, and include specific suggestions or directions (Hillocks, 1982; Parr & Timperley, 2010; Sperling & Freedman, 2010; Ziv, 1984).

What remains unclear is that for revision of different textual aspects students seem to lack different types of knowledge. In other words, information about problem detection, diagnosis, and solution seems to assist revision of different aspects in different ways. Research findings tend to show that for surface errors, misspellings in particular, error identification would be sufficient help (Haswell, 1983; McCutchen et al., 1997; Plumb et al., 1994), whereas for content, organisation, lexical choice, and relatively complex syntactic features, students need more help in diagnosis and even solution (Plumb et al., 1994; Ziv, 1982, 1984). In some cases, mere provision of solutions to grammatical errors will not help, and students will repeat the problem without understanding the nature of the problem in new writing (Ziv, 1984). The question arises about what kind of help teachers should provide for different aspects of problems, merely identification, or more information about diagnosis and solution.

The survey also exhibits methodological limitations in this small body of research. One problem lies in its diffuse foci and diverse methods, which made the research findings incomparable to some extent. In addition, many studies have looked at the effect of feedback on revision in terms of discrete changes instead of on revision with respect to between-draft quality improvement. Moreover, few studies have addressed the long-term effect of teacher feedback, i.e., its effect on student writing development indicated in a new writing task (for an exception, see Hillocks, 1982) and on student revision of new writing without external feedback (for an exception, see Carifio et al., 2001). In addition, some of the studies (e.g., Cho & MacArthur, 2010; Ziv, 1984) based their analysis on authentic teacher feedback when discussing the effect of feedback type on revision, so the result might be invalid as the amount of feedback in different types may vary greatly.

2.6.3 Research on teacher written feedback in relation to L2 student revision and writing improvement

Compared with the study of it in the L1 context, the effect of teacher written feedback on student writing has drawn much more attention in the research area of L2 writing and SLA. Since the 1990s, L2 research on teacher written feedback has been conducted on three broad themes: (a) ways in which writing instructors and language teachers¹ respond to L2 student writing (e.g., Ferris, Pezone, Tade & Tinti, 1997; Lee, 2008), (b) effects of written feedback on student writing (e.g., Conrad & Goldstein, 1999; Sampson, 2012), and (c) student and teacher views of teacher response (e.g., Hedgcock & Lefkowitz, 1994; Montgomery & Baker, 2007). Studies on the second theme will be selectively surveyed for one of the purposes of this study, which is to explore the effect of teacher written feedback on student revision and writing improvement.

Studies on the effect of teacher written feedback can be divided into two subfields: those oriented at the effect of CF on linguistic accuracy and those more concerned with the effect of content-related feedback on improvement in content. The following two subsections will thus review studies in these two subfields in turn, and each subsection will close with a summary.

2.6.3.1 The effect of written CF

Among the burgeoning research on written CF, a number of studies have explored its effect on student revision (e.g., Ashwell, 2000; Ferris & Roberts, 2001), which usually found that compared with content-focused feedback or no feedback, CF substantially helped students to reduce linguistic errors in revision. After Truscott (1999, 2004, 2007) argued against CF's efficacy for long-term language learning, more studies examined student accuracy gains from CF in new writing or error correction tasks in order to test the long-term effect of CF (e.g., Bitchener & Knoch, 2010a; Diab, 2015; Frear, 2012; Riazantseva, 2012; Sheen, 2007). This research has yielded accumulative evidence in support of the long-term learning effect of CF (Kang & Han, 2015). However, it is notable that the evidence is not conclusive because most of these studies only targeted

¹ Writing instructors and language teachers in the L2 context have somewhat different foci in instruction (Ferris, 2010). The former are concerned with learner development in both writing skills and language learning, whereas the latter are mainly concerned with students' language learning.

particular linguistic features, use of articles or verb tense (for a review, see Bitchener, 2012). The long-term effect of feedback which addresses a broader range of error types has been scant.

Only very few studies have investigated the effect of CF on both revision and new writing (for the difference between the two concepts, see section 2.6.1). Truscott and Hsu (2008) found that the CF group was significantly more successful than the control group in reducing errors during revision, but that when writing anew one week later, the two groups did not show distinguishable difference in grammatical accuracy. The researchers therefore argued that CF's effectiveness for error reduction in a revised version of a text might not transfer to improvement over time in new writing. By contrast, Chandler (2003) and Van Beuningen, De Jong & Kuiken (2012) observed that CF treatment groups not only benefited revision but also new writing in respect to linguistic accuracy. The following subsections will look closely at the findings and methods of previous research, surveying the reported efficacy of CF in relation to CF type, error type, and individual learner factors.

2.6.3.1.1 The effect of written CF type

This subsection will start with a brief account of the common written CF types. Ellis (2009; see also Ellis, 2010) proposed a typology of written CF based on the diverse CF methods used in instruction and research designs. Among the options in this typology, the most commonly researched feedback types are “direct”, “indirect”, “metalinguistic”, “focused”, and “unfocused”. “Direct” feedback refers to locating errors plus providing correction; “indirect” feedback refers to locating errors only (often by underlining and marginal indication). “Metalinguistic” feedback takes the form of “coded” feedback that indicates error nature by codes, or supplying brief grammatical explanations. Moreover, “focused” feedback targets one to three type of errors, while “unfocused” or “comprehensive” feedback attempts to address all or most of error types.

If related to the three elements (problem detection, diagnosis, and solution) necessary for revision, “indirect” feedback roughly corresponds to information about detection, “direct” feedback corresponds to information about detection plus solution; “coded” and “grammatical explanations” feedback both correspond to information about diagnosis, with “grammatical explanations” being more explicit.

Previous studies have generated complicated findings about the effect of CF type on revision. Strong evidence has shown that identifying error location suffices to help students to self-edit a fairly large portion (around 40% to 80%) of errors (e.g., Ferris, 1997, 2006; Truscott & Hsu, 2008). As to the relative efficacy of different CF methods, it is usually found that direct correction yields the best revision result (Chandler, 2003; Frear, 2012; Van Beuningen et al., 2012). Chandler (2003), for example, reported significant differences between the effects of four treatments, with direct correction being the most beneficial, followed by underlining with codes (coded feedback), underlining (indirect feedback), and marginal codes in order.

A few studies found that CF groups significantly outperformed control groups but CF types of varying explicitness only resulted in slight differences. For instance, Van Beuningen et al. (2012) found that direct feedback was only slightly superior to indirect feedback (informing error location and error category, actually coded feedback according to Ellis's typology), indicating that information about error location and category might be sufficient for student self-editing. In Ferris and Roberts's (2001) study, the group receiving underlining without codes (indirect feedback) did not differ much from the one receiving underlining with codes (coded feedback) in editing their texts, suggesting that less explicit CF might be adequate for some contexts.

In sum, the research concerning the assistance of CF in student editing is still limited in scope, especially systematic investigations where findings can be compared about relative effects of different CF types. Moreover, the frequently tapped feedback types are direct, indirect, and coded, while evidence regarding grammatical explanations is largely lacking.

With regard to the efficacy of different CF types on long-term accuracy development, a large number of studies focused on student acquisition of particular formal features (e.g., articles, verb tense, copular 'be', preposition) shown in new writing tasks, and they have yielded mixed findings. They generally found that focused, direct, and metalinguistic feedback was more effective than unfocused, indirect, and nonmetalinguistic feedback (e.g., Bitchener, 2008; Bitchener & Knoch, 2010b; Diab, 2015; Frear, 2012; Sheen, Wright & Moldawa, 2009; Shintani & Ellis, 2013). On the other hand, a couple of studies reported no different results from treatment types (Bitchener & Knoch, 2010a; Ellis, Sheen, Murakami & Takashima, 2008; Stefanou & Révész, 2015).

A smaller number of studies looked at the overall written accuracy, i.e., the accuracy of various linguistic features rather than that of particular linguistic features. They found that the effects of treatment methods in the long run varied from those shown in revision. Chandler (2003) reported that over time direct correction and underlining worked significantly better than the other coded feedback types. Van Beuningen, De Jong & Kuiken (2008, cited in Van Beuningen et al., 2012) observed improved accuracy in revision for both direct and coded feedback conditions but long-term improvement only existed in the direct feedback condition. However, their later study (Van Beuningen et al., 2012) found that while direct and coded methods generated equal effects in a new writing task one week after the treatment, coded feedback yielded larger accuracy gains in another writing task four weeks later. Sampson (2012) compared the effects of direct correction and coded feedback not only on EFL students' accuracy in new pieces of writing but on their ability to revise their drafts which they had revised with teacher feedback. His data showed that while both feedback methods helped learners to correct errors in their drafts and to improve accuracy over time, coded feedback appeared to be more effective.

2.6.3.1.2 The mediation of error type

A few researchers examined the mediation of error type when looking at the effect of CF. Before the related studies are surveyed, two dichotomies will be introduced regarding linguistic error types: the grammatical vs. nongrammatical division by Truscott (1996, 2001) and the treatable vs. untreatable division by Ferris (1999, 2006). Truscott (1996) posits that “syntactic, morphological, and lexical knowledge are acquired in different manners. If it is the case, then probably no single form of correction can be effective for all three” (p. 343). He further claims that CF could only be effective for “errors that involve simple problems in relatively discrete items” (Truscott, 2001, p. 94), such as misspellings, and might not be for grammatical errors that are complex and need to be treated by instructions on rules. Truscott therefore argues against grammar correction. However, it should be noted that the complexity of grammatical errors is not absolute but relative. For one thing, grammatical forms have a range of complexity levels. For another, the complexity level of a grammatical feature may be changeable for a learner during the process he/she learns the second language, as indicated in earlier SLA research (Ellis, 2015). In addition, CF can provide metalinguistic explanations

which have a similar function to explicit instructions on grammatical rules. Therefore, Truscott's reasoning against the efficacy of CF in treating grammatical errors is subject to question.

Ferris (1999, 2006) agrees with Truscott on the different amenability of error types to CF, but her further analysis is at odds with him. According to Ferris, many grammatical errors (e.g., verb tense) are "treatable" because they can be explained by certain patterns or rules; conversely, some problems with lexical choice and sentence structure are "untreatable" because they are idiosyncratic for which there are no rules to consult. Ferris proposes that for "treatable" errors indirect feedback (identification of errors, sometimes with diagnostic information) is suitable while for "untreatable" errors direct correction is preferable.

A few studies differentiated across error types in their data analysis. Ferris and Roberts (2001) compared the effects of CF on student revision, finding that both treatment groups (underlining, underlining with codes) reduced more errors across drafts in the treatable categories (verb, noun, article) than in the untreatable types (word choice, sentence structure). Their finding seemed to confirm Ferris's proposition while refuting Truscott's.

More studies looked at linguistic accuracy improvement in relation to error type in new pieces of writing other than in a revised draft. In Bitchener, Young and Cameron's (2005) study, the adult migrants were assigned into three groups (written plus conference feedback, written feedback, no feedback) and wrote four letters over 12 weeks. Both experimental groups received focused feedback on three types of errors in their writing: preposition, simple past tense, and the definite article. The group receiving the first type of focused feedback produced in new writing a significantly higher accuracy level in verb tense and articles but not in prepositions. This finding lent some support to Ferris's (1999) dichotomy of error types because verb tense and articles, which belong to treatable type, responded to CF differently to prepositions, which are untreatable according to Ferris.

Ferris (2006) reported a study where university students' essays were marked in 15 error categories over a semester. The comparison between the first and the last essays demonstrated that the students significantly reduced verb errors, improved accuracy in

noun form and lexis moderately, but regressed slightly in the accuracy of articles and sentence structure. This finding did indicate that different error types respond to CF differently.

Sampson (2012) also investigated the effects of comprehensive CF on participants' accuracy improvement in new pieces of writing, finding that compared with other linguistic errors, spelling, verb tense, and word choice appeared to be more persistent across different narrative writing tasks and less amenable to either direct correction or coded feedback with other types of assistance (e.g., peer/teacher explanation).

Riazantseva (2012) compared graduate students' written texts on three writing tasks (in-class essays, in-class summaries, at-home summaries) completed 12 weeks apart. She discovered that CF over a semester helped to significantly reduce lexical errors in all the three types of writing tasks and a significant effect on grammatical errors appeared in two of the three writing tasks. Further analysis revealed that in different writing tasks particular error categories responded to the treatment differently.

Van Beuningen et al.'s (2012) study, examining the effects of CF on both revision and new writing of Dutch learners, yielded more complicated evidence. They looked at the effects on grammatical errors (a wide range rather than focused) and nongrammatical errors (including lexical choice and mechanics). In the revision session, direct and coded feedback had very similar positive effects on both grammatical and nongrammatical errors, and nongrammatical errors were correctly edited to a slightly larger extent. The results of new writing tasks showed that in the long term, only direct correction helped to develop grammatical accuracy while coded CF showed its advantage in reducing nongrammatical errors. These findings seemed to be at odds with Ferris's (1999) contention that direct correction is more effective for nongrammatical problems. Van Beuningen et al. reasoned that the revision effect of coded feedback did not transfer to long-term accuracy development probably because the pupils were not sure about the accuracy of their own hypothesised solutions and so failed to internalise the self-corrected structures. They speculated that only when direct corrections were offered for the grammatical errors would pupils adopt the correct forms for learning in the long term.

Several other factors may nevertheless account for the complex findings of Van Beuningen et al.'s (2012). First, the participants, unlike those university learners or adult learners in previous studies (e.g., Bitchener et al., 2005; Ferris & Roberts, 2001; Sampson, 2012), were 14-year-olds who probably had only received limited instructions on grammar and might lack analytical abilities to solve grammatical problems based on information about error category. In other words, the distinctive demographic characteristics of these young learners intervened in the research results, indicating that the "treatability" of error types could vary with learners' backgrounds and educational contexts. In addition, a problem might exist concerning the division of error categories, which may have clouded the findings. The researchers subsumed word choice and mechanics under the same (nongrammatical) category and analysed them as a whole, but related research has suggested that lexical choice and mechanics would respond to different feedback methods differently (Ferris, 2006; Plumb et al., 1994; Ziv, 1984). To reveal clearer details, it is better to differentiate between lexis and mechanics in data analysis.

A more recent study by Diab (2015) compared the effects of direct, coded feedback and coded feedback on two types of errors (pronoun agreement errors and lexical errors) in new writing. The participants were Lebanese EFL learners whose native language was Arabic. They wrote an argumentative essay (pretest) and received form-focused instructions on the use of pronoun agreement and lexical choices. Next they completed three rounds of treatment, i.e., writing-feedback-editing assignments. Finally, the participants completed an immediate posttest and a delayed posttest, both of which took the form of argumentative essay writing. Diab found that the group receiving direct, coded feedback significantly increased the accuracy in pronoun agreement from the pretest to the immediate posttest and that that group made the least lexical errors in the delayed posttest.

Diab's results were, however, open to question, due to her research design and data analysis. The form-focused instructions provided for the participants after the pretest might have contributed to the decrease in errors. The instructions were supposed to be a mediating variable although they were also given to the control group. With regard to the statistical analysis, it seemed that Diab only performed repeated-measures ANOVAs but did not conduct pairwise comparisons. According to her report, the results were

partly based on the figures that graphically represented the descriptive statistics, which was not a valid method to gain results.

Given the scarcity of the CF research addressing error-type factors as well as the mixed findings, more evidence is definitely in need to better inform effective feedback intervention since one size does not fit all (Ferris, 1999). Finally, just as error type is not negligible for a solid analysis of the effect of CF, the mediation of learner factors has also attracted attention from this research area.

2.6.3.1.3 The mediation of learner differences: cognitive and affective factors

There has been limited evidence concerning the mediation of learner differences in the field of written CF. The convergent investigations into language analytical ability and metalinguistic knowledge have yielded mixed findings. Sheen (2007) discovered a significant positive correlation between language analytical ability and student gains from direct CF, and the correlation was stronger when direct CF was offered with metalinguistic information. Furthermore, Stefanou and Révész (2015) found that students with greater grammatical sensitivity and knowledge of metalanguage benefited more from direct CF than from the direct CF plus metalinguistic information treatment. Shintani and Ellis (2015) discovered that the effect of language analytical ability interacted with factors including opportunity to revise, the target structure, and the lapse of time. With respect to affective factors, Sheen (2011) reported that learners' attitudes intervened in the effect of written CF but that anxiety did not, and Hyland (2003) found that learners tended to closely follow CF if they considered written accuracy to be important.

2.6.3.1.4 Summary of the review of written CF and L2 student writing

The above review suggests that teacher written CF positively influences L2 accuracy in student writing. There seems to be greater effects for direct feedback over indirect or coded feedback on student revision and greater effects for focused, coded feedback over unfocused, uncoded feedback on long-term accuracy development. The existent evidence also shows the influence of error type (e.g., grammatical, lexical) and learner factors (e.g., language analytical ability, attitude) on CF effect.

The research findings are still not consistent or conclusive. This partly results from limitations in methodology (for a discussion, see Gu nette, 2007; Liu & Brown, 2015; Riazantseva, 2012) and the variations in research methodology (e.g., different participants, writing tasks, measures of accuracy, distinction between error types; see Kang & Han, 2015; Liu & Brown, 2015).

As an example to show limitations in methodology, Chandler's (2003) findings are somewhat blurred due to her study design. All her participants received the four types of feedback actually, only in different orders, so possibly their experience with one type of feedback influenced their processing of another feedback method. The distinctive effects of a particular CF type could not be ascertained. In Sampson's (2012) study, the coded feedback group received extra oral feedback so the results might not be valid to account for the different effects between coded and uncoded feedback. In another example, Ferris and Roberts (2001) tested their treatment groups on their knowledge about the error categories targeted for feedback before they received feedback and revised their text. Such a test might have increased both treatment groups' metalinguistic awareness of the targeted forms and partly led to the insignificant difference between the revision performances of the two feedback groups.

The variations in research methodology mainly exist in the choice of participants, writing tasks, and measurement of accuracy increase. The participants recruited in the CF studies varied from university learners (e.g., Sampson, 2012) to junior high school students (Van Beuningen et al., 2012), from English-as-a-second-language (ESL) learners (e.g., Chandler, 2003) to EFL learners (e.g., Sampson, 2012), and from graduate students (Riazantseva, 2012) to adult migrants (Bitchener et al., 2005). Next is an example to illustrate the variation in the writing tasks used. Ferris and Roberts (2001) designed a timed in-class argumentative writing session and a timed in-class revision session, whereas Chandler (2003) used a narrative writing task and assigned the writing and the revision tasks as homework, although both studies compared the effect of coded feedback and underlining on revision. Finally, a range of methods were employed to assess accuracy change across drafts or across tasks, such as subtracting the error rate of one draft from that of another (e.g., Frear, 2012) and dividing the number of errors corrected by the number of errors marked in the feedback (Ferris & Roberts, 2001; Sampson, 2012).

The next section will turn to L2 research on written feedback targeting issues beyond linguistic accuracy. For the sake of convenience, “content feedback” will be used to refer to teacher written feedback on both content and rhetorical aspects.

2.6.3.2 The effect of content feedback

Studies on the effect of content feedback are much fewer compared to research on written CF. A large part of research on content feedback has looked at teacher commentary through other media instead of written comments, such as audio-taped feedback (e.g., Morra & Asis, 2009), online annotations (e.g., Yeh & Lo, 2009), and more widely, writing conferences where teachers communicate orally with students about how to improve their writings (e.g., Goldstein & Conrad, 1990; Mota de Cabrera, 2003; Weissberg, 2006; Williams, 2004).

Among the small scope of research on content feedback, several studies examined the effects of written feedback on content versus feedback on form. The findings tended to be that content feedback had a weaker effect on students’ content improvement than the effect CF had on accuracy improvement (Ashwell, 2000; Fathman & Whalley, 1990; Lee, 1997; Sampson, 2012). The moderate effect of content feedback might relate to the features of the content feedback provided in some studies. For example, Ashwell (2000) offered many positive comments (e.g., good description), general suggestions (e.g., develop paragraphs), and questioning, which might be inapplicable or too implicit for students to make revisions (Ferris, 1997; Hattie & Timperley, 2007). For another thing, correcting meaning problems demands “text processing skills of a higher order” (Lee, 1997, p. 471), so mere identification of problems or vague cues could hardly help. On the other hand, Morra and Asis (2009) reported that their EFL students significantly reduced both linguistic errors and problems in content and organisation after receiving coded feedback with finely developed codes targeting specific problems.

2.6.3.2.1 The mediation of the characteristics of content feedback, problem type, and learner factors

A couple of studies explored the efficacy of content feedback in relation to its characteristics. A notable instance is the work by Ferris and colleagues (Ferris, 1997; Ferris et al., 1997) based on a teacher’s comments on over 100 student drafts. The researchers analysed the pragmatic and syntactic features of authentic teacher comments,

reporting that direct, declarative comments appeared to be more effective than indirect ones; requests for specific information and provision of a revision strategy usually led to successful changes, and that use of hedges made little difference.

Building on Ferris's (1997) approach, Conrad and Goldstein (1999) explored not only the relationship between comments' characteristics and student revisions but also the factor concerning the type of the problem to be revised. They discovered that successful revisions were associated with certain features of feedback, e.g., declaratives instead of questions, direct instead of indirect language. Further analysis revealed that the type of the problem addressed by the feedback determined the revision results much more significantly than the features of the feedback. Compared with problems regarding coherence, cohesion, purpose, and lexical choice, problems regarding explanation, analysis, and explicitness of argument were revised least effectively irrespective of feedback type, and the teacher's comments on them often lacked features which seemed to yield successful revisions (i.e., declaratives, direct comment). These findings are enlightening. Both the type of comments and the type of problems influence the result of student revision.

Very few studies have investigated learner factors in the effect of teacher comments on student revision. As an exception, Goldstein's (2006) qualitative data revealed that students' content knowledge and viewpoints about the topic, self-motivation, and other schoolwork might account for students' revision decisions (see also Conrad & Goldstein, 1999).

2.6.3.2.2 Summary of the review of teacher content feedback and L2 student writing

The research on L2 teacher content feedback suggests that content feedback does not work so effectively on student writing in content and rhetorical aspects as CF does on written accuracy. Direct, specific, and declarative comments seem to be more effective; provision of solutions facilitates effective revisions. Furthermore, revision results appear to be related to problem types (e.g., coherence, explanation of argument) and learner factors (e.g., content knowledge, motivation).

The meagre studies touching upon the features of content feedback shared a few limitations (Conrad & Goldstein, 1999; Ferris, 1997). First, the data for analysis were restricted in scope. In Conrad and Goldstein's case, only drafts of three students and

comments by one teacher were analysed. Second, since the studies were based on authentic teacher feedback, the data concerning the effect of feedback types were not comparable in some cases because of their uneven occurrences with different types of problems. Therefore, more studies are needed which exert some control to overcome the drawback of authentic teacher comments.

2.7 L1 and L2 Studies on Revision Performance in Relation to Allocation of Cognitive Resources

The previous section surveyed pedagogy-oriented revision research, i.e., research on the effect of teacher written feedback on revision, centring on the assistance of teacher feedback in the three essential subprocesses of revision (detection, diagnosis, solution). This section, instead, will review research that approached revision from the perspective of cognitive resources, including resources in working memory and long-term memory and cognitive effort affected by revision task conditions. Feedback was not involved in this body of research. As introduced in section 2.4, revision has been conceived as a cognitively demanding process because of the multiple subprocesses engaged within working memory during the revising process, such as reading to represent the text, reading to detect and diagnose problems, searching for revising strategy, and translating from mental revision to actual text. In addition, this cognitively demanding process is affected by available resources in working and long-term memory (Butterfield et al., 1996; Hayes, 1996).

Before the review continues, it is necessary to give a brief account of related technical terms in cognitive science, which centre on the concept of “resources”. The key concept “resources” means the mental energy available at a certain moment which cognitive processes require to operate (Baddeley, 1986; Halford, Wilson & Phillips, 1998). “Working memory” refers to the overall resource limitations during the processing and retention of information which is attended to (Baddeley, 1986; Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005), while “long-term memory” denotes a system where information is stored for a long time and can be managed and retrieved for later use. Transferring information between long-term memory and working memory is regarded as a way to lessen the load on working memory (Butterfield et al., 1996; McCutchen, 2000). Finally, “capacity” refers to the maximal amount of resources available to an individual (Gathercole & Baddeley, 1993).

Another important concept is “attention”. Perceived as a limited capacity or resource, it comprises three interrelated mechanisms: attention as alertness, attention as orientation, and attention as detection (Posner & Petersen, 1990; Tomlin & Villa, 1994). Attention as alertness denotes overall readiness to deal with stimuli, attention as orientation represents actual deployment of resources to stimuli, and attention as detection refers to mental registration of some features of stimuli.

Shifting the focus on the individual to the task, the amount of resources demanded by a given task is termed “cognitive effort” (Kellogg, 1987, 1988). The cognitive effort is subject to the task environment or task conditions, which are external to an individual’s cognition (Butterfield et al., 1996; Hayes, 1996; see also Ong, 2014).

The above concepts, critical for perception and learning in general, are also crucial elements involved in writing and revising processes. In the past two decades numerous studies, mainly in the L1 context, have explored the issue of revision from the perspective of cognitive resources. The following sections will survey research along this vein and present the specific perspective developed for this study.

2.7.1 L1 revision studies from the perspective of cognitive resources

L1 revision studies from a cognitive approach have mainly explored three types of factors deemed to influence revision performance: individual working memory capacity, topic knowledge, and the cognitive effort demanded by task conditions.

According to Hayes’ (1996) and Butterfield et al.’s (1996) revision models, various data are stored and processed in the system of working memory; therefore, individual differences in working memory capacity would result in different revision results. The empirical evidence is mixed. It is found that high capacity students corrected more spelling and syntactic errors but not more coherence problems (Piolat, Roussey, Olive & Amada, 2004; Roussey & Piolat, 2008). Adams et al. (2010) found working memory capacity to be significantly positively correlated with error identification, but not with error diagnosis or the quality of error correction.

From another perspective, McCutchen (1986, 2000) postulates that sound knowledge of a text’s topic will facilitate the detection and correction of meaning-related problems in the text and will reduce the burden of working memory. McCutchen et al. (1997) found

that when revising a text the topic of which was familiar to them, students revised more meaning-level problems but not spelling errors. Adams et al. (2010) also reported significant correlation between topic knowledge and revision, but the correlation appeared to be stronger with language errors rather than structure or argument, somewhat contrary to McCutchen et al.'s findings.

Some other studies examined students' revision performance under different revision task conditions, and they can be divided into two groups. One group, inspired by the hypotheses about working memory components involved in writing (Hayes, 1996; Kellogg, 1999), attempted to assess the engagement of particular cognitive resources in the various processes in writing, and they usually designed conditions impeding the operation of the very resources to be assessed. For example, Levy and Marek (1999) asked participants to detect spelling and grammatical errors in a given text under two conditions—while being exposed to irrelevant speech (supposed to impair the functioning of phonological loop in working memory) and while not. They found no different performance across the conditions. For another example, Chenoweth and Hayes (2003) found that articulatory suppression (another way to affect phonological loop by requiring a person to repeat a syllable again and again) led to more uncorrected surface errors. There is also evidence that the effect of revision conditions varies with error type, suggesting that revision of different error types requires different amounts of processing resources (Larigauderie, Gaonac'h & Lacroix, 1998).

The other group of studies involving revision conditions did not intend to differentiate the resources in working memory in relation to their engagement in the subprocesses of writing; they approached working memory as a whole which contains limited resources to store and process information. They designed task conditions expected to influence overall cognitive effort for the revising process. For example, Chanquoy (1997b, cited in Chanquoy, 2009) asked schoolchildren to revise either during or after the writing process, expecting that the delayed revision would generate better results because the cognitive resources would not be shared with the writing task. Surprisingly, few between-condition differences were found. A later study by Chanquoy (2001) did find that delayed revision resulted in more frequent and more meaning-level revisions, suggesting that simultaneously writing and revising exert a heavier cognitive load on the pupils. Galbraith and Torrance (2004) described a study that compared revision

results with and without the initial draft being available. They reported that the effect of the two conditions on final draft quality varied with the initial drafting formats.

2.7.2 L2 revision studies from the perspective of cognitive resources

L2 studies on revision have scarcely approached ESL/EFL learners' revision from the perspective of cognitive resources. The few exceptions have followed the methods used in L1 studies. Looking at the effect of working memory capacity on Chinese EFL students, Yan (2010) found that capacity significantly positively correlated with meaning-level revision, but not with revision on linguistic levels. Ong and Zhang (2013) replicated Galbraith and Torrance's (2004) manipulation (i.e., \pm availability of the initial draft), finding the same salient interaction between revision conditions and the initial drafting formats.

2.7.3 The approach of this study to studying revision in relation to cognitive resources

This will start with a brief explanation of the approach of this study to studying revision in relation to cognitive resources and then present the rationale of the proposed approach. This study plans to assess revision performance in relation to task conditions through manipulation of learners' revision-focus. The term "revision-focus" refers to what students focus on while revising a text (for more information, see section 3.4.2). The specific manipulation method was to direct learners to focus their attention, among other resources, on global levels first and on surface levels later when revising their written text. I suppose that revision of global features (content and organisation) and revision of language engage two different directions of cognitive efforts. The former entails consideration on a relatively macro level beyond the unit of a single sentence, involving the global conceptual and structural construction; on the other hand, the latter has a narrower focus on within-sentence elements such as sentence structure and word choice (Faigley & Witte, 1981; Lee, 1997). It is unknown whether task conditions in terms of revision-focus manipulation would affect cognitive effort and in turn influence revision performance.

This approach involving revision-focus direction is, on one hand, motivated by the propositions about limited resources in contrast to the demanding nature of a revision activity (Butterfield et al., 1996; Hayes, 1996), as introduced in section 1.3.2. On the other hand, it is based on some theoretical discussions about the impact of task demands

on learning and performance that centre on cognitive resources and attention. In the meanwhile, this approach is inspired by a few studies which examined the writing/revising outcome in relation to cognitive resources devoted to meaning versus form.

2.7.3.1 Theories about task complexity

Task complexity theories in the sphere of L2 learning can relate to the revision-focus manipulation proposed for this study. Skehan (1998; see also Skehan & Foster, 2001) postulates that non-native speakers have limited attentional resources in task execution, which will result in a trade-off effect in their allocation of attention to different aspects of performance due to the task's cognitive demands. Further to that, Robinson (2005, 2007) claims in his Cognition Hypothesis that when non-native language learners solve two or more tasks simultaneously their cognitive resources will disperse, which means that their attentional and working memory resources will be diverted away from focusing on any particular aspects of solving the tasks. He also posits that the amount of "computation" affects the complexity of cognitive processing: the larger the amount of information is processed, the more resources a task will demand.

Though these hypotheses are aimed to gauge linguistic performance in relation to task complexity, their insights are applicable for assessing the influence of revision-focus direction on revision performance. Students working on global and surface levels separately are supposed to work with a lower level of task complexity with less amount of information to compute during a specific time period. By contrast, students without receiving revision-focus direction would probably process text on global and surface levels simultaneously, and their cognitive resources consequently disperse.

2.7.3.2 Wickens's (2002, 2007) multiple resource model

Wickens developed a multiple resource model to address the issue of divided attention and selective attention in concurrently performed tasks. He claims that two concurrent tasks will interfere with each other because the performer usually needs to divide attention between them. Although initially aimed at fields other than L2, the model can serve to predict the cognitive demand of a revision task influenced by manipulation of student revision-focus.

In a revising task, revising global features of the text and revising surface features can be perceived as concurrently performed subtasks. The question is whether and to what extent they may interfere with each other. The answer to this question predicts the potential validity of the above-mentioned revision-focus manipulation (directing a reviser to work on global features and surface features separately). If, based on Wickens's theory, revising global features and surface features only minimally compete for the reviser's attention and other resources, the devised revision-focus direction is unlikely to play an evident role in revision performance. Then it would be unwarranted to study it as a factor.

According to the multiple resource model, revising global features and revising surface features are plausible to interfere with each other. The model postulates that three elements influence the extent to which one or both of the tasks suffer from the performer's divided attention. They include the amount of cognitive effort demanded by each task, the nature of the cognitive effort demanded, and the performer's allocation of cognitive resources to each task.

(1) The amount of cognitive effort demanded by each task

With respect to the first element, it is hypothesised that the objective difficulty level of each task affects the degree of competition between the two tasks for the performer's available resources. For Chinese university EFL students, revising meaning and structure in argumentative writing is supposed to be difficult because of the need for reflective reading and generation of ideas, compared to English reading comprehension (Flower et al., 1986). This difficulty is aggravated by the accompanying transcribing process, which is particularly challenging for EFL learners (Kormos, 2011; Zhang, 2006). Revising surface features such as grammar and wording is also difficult for many Chinese EFL learners because they have not mastered written English as well as their mother tongue. In this sense, revising global features and revising surface features of a piece of English writing will interfere with each other if performed concurrently.

(2) The nature of the cognitive effort demanded by each task

With regard to the second element pertaining to the nature of cognitive resources required by tasks, Wickens proposes three dichotomous dimensions to classify resources according to the mechanism of information processing: stage, perceptual modality, and processing code. Stage includes “perceptual and cognitive” vs. “selection and execution of actions”, representing two different stages of information processing; this dichotomy corresponds to the distinct functions of different brain regions. Perceptual modality comprises “visual” vs. “auditory” depending on the sensory organ deployed in information processing. Processing code comprises “verbal material” vs. “spatial nonverbal material” based on the target to be processed. Wickens further posits that interference of two tasks with each other will be greater if they share common levels along more of the above dimensions. In other words, the more similar the required resources are in nature, the more interference exists between two tasks.

Relating these dimensions to the dual revising task, it is found that revising on global levels and revising on surface levels demand similar resources in all the three dimensions. First, both of them require more “perceptual and cognitive” resources than motor-related resources. Second, they both need the visual sense rather than the auditory sense. Finally, they both involve work on verbal materials, i.e., transcription and text editing, although revising globally involves more nonverbal effort such as creation and organisation of ideas. Therefore, it is plausible to believe that revising on global and surface levels simultaneously will incur a competition of cognitive resources from the reviser.

(3) The allocation of cognitive resources to each task

The performer’s allocation of cognitive resources to each task will mediate the extent to which the first two elements impact task interference. It is an individual choice as to which task “should be ‘protected’, and which may be ‘sacrificed’ when there is resource competition” (Wickens, 2007, p. 183), and this choice is “heavily driven by performer perceptions of task importance” (p. 187). Therefore, this element is subjective, different from the preceding two elements that derive from objective task conditions. Since the multiple resource model is consulted in order to predict whether revising on global and surface levels compete for resources, the purpose is objective computation based on task conditions regardless of performer personal choice.

The above discussion of the multiple resource model in relation to the dual revising task indicates that revising globally and revising locally at the same time will compete for resources from an EFL student. Therefore, the proposed revision-focus manipulation (i.e., directing learners to focus their resources on global levels first and on surface levels later during revision) is plausible to be an influential factor for revision performance because by asking students to revise on the two levels separately it reduces the degree of task interference and effort competition.

2.7.3.3 Studies related to information processing on meaning versus form

Another source of literature that has prompted me to inspect revision-focus direction is a few studies which touched upon writing/revising performance in relation to cognitive resources devoted to meaning versus form. Glynn, Britton, Muth & Dogan (1982) reported that L1 writers attempting to juggle goals of idea formulation and idea translation produced fewer arguments than those free of translation concerns. Galbraith (1980) discovered that an L1 reviser's priority given to language hindered the revision of content. Two L2 studies (Ashwell, 2000; Fathman & Whalley, 1990) compared students' revision after receiving feedback of different foci: feedback on content vs. feedback on language vs. feedback on content and language. It was predicted that the foci of feedback would direct students' attention to that focused aspect and induce larger gains. The findings did not support the trade-off effect between attentional focus on meaning and form. The students receiving feedback on both meaning and form improved both content and accuracy, and overall they outperformed the students with feedback of a single focus.

To sum up, the relationship between revision-focus direction and revision performance has been under researched and deserves investigation. If a reviser generally needs to cope with demands from the discourse level (developing and organising content) and the surface level (translating mental content into correct language), it is important to explore how to allocate limited resources to satisfy the dual demands. Teachers need to know whether it is necessary to direct student revisers' attention to content and form in turn and to give feedback on the two foci separately as recommended by Sommers (1982) and Zamel (1985). For these reasons, this study intended to examine the role of revision-focus manipulation in student revision and writing improvement by designing a way of directing student revision-focus (as to be elaborated in Chapter Three).

2.8 Summary of Chapter Two

This chapter started with the background and theoretical development of research on revision in the L1 writing context. It presented conceptions and theoretical models concerning revision in writing. Next, it briefly surveyed the existent L1 and L2 research approaches to revision and presented the approach of this study. After that, this chapter reviewed studies pertaining to the foci of this study. The following sections summarise major findings of the literature review, identify research gaps, and finally outline the research perspectives for this study.

2.8.1 Major findings

Early L1 theoretical research on writing revision has led to the proposition that revision is a crucial component in the composing act. Revision can be approached as either a thinking process or products of the thinking process, i.e., physical changes to the original written text. As a thinking process, it usually entails the reviser's detection of a problem, diagnosis of it, and strategy selection for handling the problem. Such a thinking process is subject to the reviser's cognitive/metacognitive resources in long-term memory as well as his/her resources in working memory. On the other hand, as physical changes, revisions may occur on different textual levels, on different occasions, and with different effects.

Based on the L1 revision theories, I developed two major foci for further review of studies: (a) the effect of teacher feedback on student revision in relation to the information it provides, and (b) the effect of cognitive demand on student revision.

The review of L1 and L2 studies on teacher written feedback and revision suggests a generally positive role which teacher feedback plays in providing information concerning problem detection, diagnosis, and solution for performing revisions. Furthermore, studies in both L1 and L2 settings have found that the effect of teacher feedback on revision and writing improvement overtime may be mediated by the characteristic of feedback, the type of discourse problems and linguistic errors targeted for feedback, and learner factors.

The revision studies on cognitive factors suggest that revision performance is related to the reviser's working memory capacity, topic knowledge, and the cognitive effort

demanded by task conditions. In addition, the specific effects of the above factors seem to be mediated by linguistic error type and textual level targeted for revision.

2.8.2 Major gaps in the literature

The above-mentioned review of literature identified six major research gaps.

1. The whole body of L2 research on teacher written feedback and student writing has demonstrated unbalanced concerns in favour of CF and linguistic accuracy development. English writing, however, is not a grammatical or lexical exercise; it is also a tool of thought (Bruton, 2009). It is therefore equally valuable for teachers to respond to “the content of their students’ work as well as to rhetorical concerns such as purpose, audience, organisation, and development” (Goldstein, 2004, p. 65). Whether and how teacher written feedback can help L2 writers to improve content and rhetorical aspects is a worthwhile but neglected topic.
2. Student autonomous revision (i.e., revision of new writing with no external feedback) has been scarcely addressed in written feedback research. It is worth exploring whether WF helps student writers to raise awareness of their problems and develop their strategies for improving their written text given that one of the eventual aims of EFL writing instruction is to foster self-sufficient revisers (Bitchener & Ferris, 2012).
3. Longitudinal investigation has been scant among the research on written feedback and student writing. Findings based on a one-time treatment are questionable (see Liu & Brown, 2015).
4. Few studies on content feedback have systematically examined revision results in relation to the type of information offered through feedback although most have focused on the pragmatic or syntactic characteristics of comments. However, it is significant to examine the effect of feedback in relation to its information regarding problem detection, diagnosis, and solution, since such information is supposed to impact revision more directly than the pragmatic or syntactic features of comments. This approach to categorising feedback is not only theoretically grounded but helps to reveal what type of information student writers already have and what they still lack to conduct revisions.

5. Mediation of discourse-related problem type and linguistic error type in the effect of written feedback needs further investigation, since one size does not fit all as captured in the existent evidence (e.g., Conrad & Goldstein, 1999; Riazantseva, 2012).
6. In the area of written feedback on student writing, the role of cognitive factors related to resources, such as working memory and cognitive effort, as well as the effect of task conditions on allocation of resources, have been under researched.

2.8.3 Research perspectives of this study

In order to fill some of the gaps mentioned above, this study was designed for the following purposes. First, it aimed to address the effects of both content feedback and CF on students writing. Second, it planned to examine the effect of written feedback in three aspects: student revision in response to the feedback, student autonomous revision (or revision skills), and student writing development (i.e., new writing). To increase the validity of research results this study would employ an experimental design with multiple rounds of treatment. In addition, feedback would be classified based on the information provided within, and the effect of feedback would be inspected with respect to the type of discourse problems and the type of linguistic errors targeted. Another initiative would be to examine the role of revision task conditions in student revision result and writing development over time.

CHAPTER THREE

METHODOLOGY

3.1 Overview

This chapter will first outline the research questions of this study. It will then introduce independent and dependent variables and the design for the study. It will next report the pilot session and discuss how this trial session informed the main study. The chapter will continue to present the main study with respect to population and sample, instruments, and procedures for data collection and data analysis.

3.2 The Postpositivist Worldview

I adopted the postpositivist worldview as the philosophical justification for the methodological approach for this study. The postpositivist worldview holds the following basic principles (Creswell, 2009). First, there exist laws or theories that govern the world. Second, causal relationships exist among subjects, and that causes probably determine outcomes. In addition, observations of the world can be presented in the form of numeric measures. Therefore quantitative strategies of inquiry are appropriate to be employed to investigate the world. Based on these principles, I drew on revision theories to formulate my research questions which aimed to identify relationships between variables in the area of EFL writing instruction. Furthermore, I employed an experimental design involving treatment and control groups to collect data. The collected data were transformed into numeric measures and submitted to statistical analyses.

3.3 Research Questions

This study intended to address research questions concerning the effect of treatment on student revision and new pieces of writing. The design of treatment manipulated two factors: written feedback (WF) type and direction of student revision-focus (termed as “revision-focus direction”), the conceptual and operational definitions of which are given in section 3.4. Here, *treatment*, used as an uncountable noun, is regarded as an umbrella term, referring to both specific treatment methods (i.e., treatment conditions; see section 3.4) and the factors of WF type and \pm revision-focus direction (\pm RFD) that constructed the treatment methods.

With respect to dependent variables, revision was conceptualised as between-draft text quality change. As reviewed in Chapter Two, revision can be approached from both the “process” and the “product” perspectives. It was examined from the latter perspective in this study and in terms of the overall effect it rendered on the text quality of a piece of writing rather than in terms of the effect of discrete changes. The conceptualisation of text quality is explained in section 3.5. New pieces of writing, referring to student original writing on new topics, were also examined in terms of text quality that was conceptualised in the same way as it was for analysing revision. Specifically, three areas were investigated by addressing the following research questions.

3.3.1 RQ1 Does treatment have any immediate effect on student revision?

RQ1.1 What are the relative effects of different treatment methods on student revision?

RQ1.2 Do the factors of WF type and \pm revision-focus direction play a role in the immediate effect of treatment methods on student revision?

RQ1.3 Does textual level mediate the immediate effect of treatment?

The term “immediate effect on revision” refers to the effect on student revision of their original draft with the help of WF on it. This term is used in order to differentiate from the concept of “autonomous revision” which is a second area investigated by this study.

3.3.2 RQ2 Does treatment have any effect on student autonomous revision?

RQ2.1 What are the relative effects of different treatment methods on student autonomous revision?

RQ2.2 Do the factors of WF type and \pm revision-focus direction play a role in the effect of treatment methods on student autonomous revision?

RQ2.3 Does textual level mediate the effect of treatment?

This study was concerned about not only student revision with WF but also student revision of their original draft on a new topic without reference to external feedback, hence termed as “autonomous revision” here. This concept has been referred to as “ability to revise” (Adams et al., 2010; Yin, 1995) and “revision skills” (Butler & Britt, 2011; Carifio et al., 2001) in previous literature.

3.3.3 RQ3 Does treatment have any effect on student new writing?

RQ3.1 What are the relative effects of different treatment methods on the text quality of student new writing?

RQ3.2 Do the factors of WF type and \pm revision-focus direction play a role in the effect of treatment methods on the text quality of student new writing?

RQ3.3 Does textual level mediate the effect of treatment?

3.4 Factors of Treatment and Generation of Treatment/Control Groups

The two factors of treatment included WF type and \pm revision-focus direction, each having two levels. Manipulation of the two factors generated four treatment groups.

3.4.1 WF type

WF type was manipulated in terms of types of information WF offered regarding the proposed essential elements for revision, i.e., detection of problems, diagnosis of problems, and solution to problems (Hayes, 1996; Scardamalia & Bereiter, 1983). This manipulation was aimed at discovering what kind of assistance was more effective for Chinese EFL learners' revision in writing and writing development. Two types of WF were compared in this study: WF offering information regarding detection and diagnosis (identification + diagnosis), and WF offering information regarding detection and solution (identification + solution). They were applied to both discourse levels and linguistic levels¹ of student writing.

The rationale for selecting targets for providing feedback mainly derived from the evaluative criteria in the literature that involved evaluations of at least one of the dimensions of content, organisation, grammar, lexis, and mechanics. They include

¹ While evaluating texts, teachers and researchers usually look at the dimensions of content, organisation, grammar, lexis, and mechanics (East, 2009; Jacobs, Zingraf, Wormuth, Hartfiel & Hughey, 1981; Nimehchisalem & Mukundan, 2011). Further, they tend to group content and organisation under one level while the other three dimensions under another level. They have used "discourse level" and "linguistic level" (Bridwell, 1980; Watson Todd, Khongput & Darasawang, 2007), "global" and "local/surface" (Wallace & Hayes, 1991; Butler & Britt, 2011), or "macro-level" and "micro-level" (Ziv, 1984) to distinguish these two categories of dimensions. This study adopted the terms "discourse-level" and "linguistic-level" because they were considered to plainly reflect the nature of the corresponding aspects.

general analytical scoring criteria (East, 2009; Jacobs et al., 1981; Reid, 1993), rationale for assessing argumentative essays (Ferretti, MacArthur & Dowdy, 2000; Midgette, Haria & MacArthur, 2008; Nimehchisalem & Mukundan, 2011; Qin & Karabacak, 2010; Wolfe, Britt & Butler, 2009), and certain specific criteria for rating content (Ong, 2013), organisation (Ferretti, Lewis & Andrews-Weckerly, 2009; Lee, 2002; Watson Todd et al., 2007), or language (Ferris & Roberts, 2001; Fritz & Ruegg, 2013; Housen, Kuiken & Vedder, 2012).

To further select targets for written feedback, essays of the potential participants'², who were selected for the pilot session and the main study, were examined, and their lecturer was consulted about their writing proficiency. It was found that most participants had problems with content and organisation and made various errors in grammar and lexical choice, but that they made few errors in mechanics, i.e., spelling, punctuation, and capitalisation. Based on this preliminary enquiry and the prospective workload involved in giving feedback, the following aspects of discourse- and linguistic-level problems were to be targeted by feedback (see Table 3.1).

² Information of the participants in the pilot session and information of those in the main study are given in sections 3.8.1 and 3.9.1, respectively.

Table 3.1 Aspects of discourse- and linguistic-level problems to be targeted by feedback

Aspects to be addressed by feedback	Specific aspects to be addressed and notes for reference
Content	relevance to the topic; effective argumentation (including a claim, support, counterargument, and rebuttal) <i>Claim</i> means the statement of the main idea; <i>support</i> means the evidence for the claim; <i>counterargument</i> refers to the possible ideas opposing the writer's claim; <i>rebuttal</i> refers to the writer's response to counterargument.
Organisation	macrostructure (presence of introduction, body, and conclusion; paragraphing); connectedness in writing (cohesion, coherence) <i>Paragraphing</i> means dividing the body part into paragraphs so that each paragraph develops only one idea. <i>Cohesion</i> refers to the linguistic devices that connect the sentences in a text, and <i>coherence</i> refers to the implicit links between the concepts and propositions in a text.
Grammatical accuracy	part of speech (including word formation), syntax (agreement, tense, articles, pronouns, word order, sentence structure) Some hold that errors related to word formation belong to lexical errors. This study classified them into grammatical errors given that word formation relates to the part of speech, which is subject to grammatical rules according to the function each part of speech performs in a sentence.
Lexical accuracy	word/phrase choice (including connotation, collocation, redundant or incomplete information)

(1) Identification + diagnosis (ID)

This type of feedback identified the location in student writing where a problem existed or an improvement could be made and provided relevant diagnosis. Identification was made at discourse and linguistic levels, using underlines for within-sentence elements, parenthesis for single sentences and marginal vertical lines for multiple sentences.

Diagnosis at discourse levels pointed out the inadequacy of content and organisation in relation to the specific aspects that had been selected for identification (see above); the diagnosis was made with reference to the evaluative criteria in the literature on assessing writing. Diagnosis at linguistic-levels pointed out the nature of an error; metalinguistic

explanations were employed rather than codes because some students might be unfamiliar with or confused by codes (Ferris & Roberts, 2001; Lee, 1997). Diagnoses were written in the marginal or interline space on copies of participants' drafting sheets, and they were written mainly in Chinese to avoid obscurity to participants. The following are some examples of diagnosis (see Tables 3.9 and 3.10 for more examples of feedback on discourse levels). The examples addressing linguistic problems include the relevant sentences in student writing, which are put in italics.

Diagnosis of problems in content

Ex. 有力的论证要考虑反方观点以及如何反驳之。(Effective argumentation includes counterarguments and rebuttals.)

Diagnosis of problems in organisation

Ex. 英语议论文需要一个结尾段落，结尾段经常重申主要观点或发出呼吁。(English argumentation includes a conclusion, which often reiterates the main arguments or makes an appeal.)

Diagnosis of problems in grammar and lexis

Ex. 1

动词将来时用 will do (The verb form should be "will do" for the future tense.

I will lost all my enthusiasm.

Ex. 2

有歧义; 词汇选择不当 (ambiguous; wrong word choice)

Many medicines have been tested by animals.

It is admitted that diagnosis of lexical problems could hardly contain such metalinguistic rules as applicable to grammatical errors. For many cases of lexical errors in this study,

the diagnostic feedback was merely “wrong word choice”, “improper phrase”, “ambiguous”, or “meaning unclear”.

(2) Identification + solution (IS)

This type of feedback added solution to identification but provided no diagnosis. Identification was operationalised in the same way as in the ID feedback type. Solution refers to suggested execution on discourse and linguistic levels of student writing. In general, suggested executions at discourse levels were text-specific instead of being vague or interchangeable across texts while those at linguistic levels were more directly prescribed as corrections or better expressions. Solutions were also written in the marginal or interline space, and discourse-level solutions were mainly written in Chinese. The following are some examples of solution (see Tables 3.9 and 3.10 for more examples of feedback on discourse levels). The original sentences in student writing are put in italics.

Solution to problems in content

Ex. 有人会提出驳论，比如“我们可以通过阅读译著了解名著要表达的东西”。对此，你如何回应呢？(Some people may propose counterarguments, e.g., “We can learn the message in classic works by reading translated versions.” How do you argue against such counterarguments?)

Solution of problems in organisation

Ex. 增加结尾段落，可以概述你的主要论点。(Add a conclusion paragraph. You may summarise your major arguments.)

Solution of problems in grammar and lexis

Ex. 1

lose

I will lost all my enthusiasm.

Ex. 2

on

Many medicines have been tested by animals.

3.4.2 Direction of student revision-focus

Whether to direct student revision-focus was a second factor manipulated in the treatment. In this study revision-focus refers to the particular textual levels to which learners allocate focused resources. Revision task conditions influence the revising process and results, according to the proposition that the revising process is subject to the cognitive effort demanded by a given task as well as the reviser's cognitive capacity (Butterfield et al., 1996; Hayes, 1996). Revision task conditions were manipulated with respect to the way of directing students' attentional focus: \pm directing participants to revise discourse and linguistic levels in sequence.

(1) + Revision-focus direction (+RFD)

This manipulation was conceptualised as an intervention that directed participants' attention and energy to discourse and linguistic levels separately during their revising process. This direction was operationalised by two means: (a) explicit instruction, and (b) separate presentation of the two levels of feedback. First, participants received verbal instruction to focus on revising content and organisation of their drafts in the first half of the revising process and then focus on revising grammar and lexis in the second half of the time frame. In addition, discourse-level feedback and linguistic-level feedback were given respectively on two separate copies of participants' first drafts and then distributed to them in sequence during the revising session, which was meant to direct participants' attention to different levels of feedback separately. Such direction of focus was not absolute but relative, though, as it was unlikely for students to make a complete separation between working on discourse and on language. It was expected that

participants with revision-focus direction could and would still attend to both levels simultaneously from time to time, but to a less extent compared with those without such an intervention.

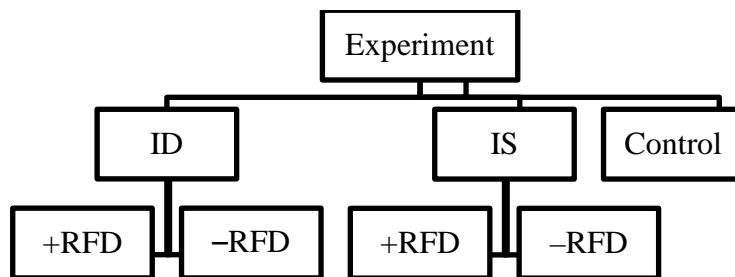
(2) – Revision-focus direction (–RFD)

Participants under this condition did not receive direction to focus on discourse and linguistic levels separately during their revising process, and the two levels of WF was written on a single copy of their first drafts and given to them together at the beginning of the revising session.

3.4.3 Treatment/Control groups

Manipulation of the above two factors (WF type and \pm revision-focus direction) generated four treatment groups: ID+RFD, ID–RFD, IS+RFD, and IS–RFD. In addition, a control group was included that received neither WF nor revision-focus direction. Hence five group conditions were produced for investigation (see Figure 3.1).

Figure 3.1 A schematic diagram of the treatment/control group conditions



Note. ID = identification + diagnosis, IS = identification + solution, RFD = revision-focus direction

3.5 Dependent Variables

Dependent variables related to student revision (i.e., between-draft text quality change) and text quality of new writing were studied to test the effect of treatment.

3.5.1 Between-draft text quality change

Many revision studies measured between-draft text quality change with regard to one or more of these aspects: global quality (Butler & Britt, 2011; Graham, MacArthur & Schwartz, 1995; Mirzaii & Aliabadi, 2013; Wallace et al., 1996), quality of content (Ashwell, 2000; Cho & MacArthur, 2010; Midgette et al., 2008; Song & Ferretti, 2013), quality of organisation (Song & Ferretti, 2013), and linguistic accuracy (Chandler, 2003; Hanaoka & Izumi, 2012; Sachs & Polio, 2007). In this study, quality in terms of the discourse and linguistic levels targeted by the WF (i.e., content, organisation, grammar, and lexis) was measured as constructs of text quality. As discussed in section 2.6.3.1.2, with respect to the effect of WF on linguistic accuracy both Truscott's (1996, 2001) grammatical/nongrammatical and Ferris' (1999, 2006) treatable/untreatable dichotomies have been investigated with mixed findings. Given that mechanics were not targeted in this study and that the treatability of error types might not be unanimous for various populations of L2 learners (Chan, 2010; Van Beuningen et al., 2012), neither Truscott's nor Ferris' approach was strictly followed. The data analysis concerning linguistic quality simply differentiated between change of quality in grammar and lexis. The conceptualisation of quality at each textual level corresponded to the aspects addressed by the WF (see section 3.4.1). Specific criteria for measuring text quality are elaborated in section 3.9.4.1.

3.5.2 Text quality of new writing

Text quality of new writing was approached in the same way as it was measured for assessing between-draft text quality change. Therefore, quality regarding discourse and linguistic levels were analysed. Detailed criteria for measurement are discussed in section 3.9.4.1.

3.6 The Relationships between Independent and Dependent Variables

Table 3.2 below illustrates the relationships between independent and dependent variables as well as how these relationships correspond to the research questions.

Table 3.2 Relationships between independent and dependent variables

Dependent variables	Independent variables		
	Group condition	WF type; \pm RFD	Textual level
Student revision with the help of WF (Between-draft text quality change in content, organisation, grammar, and lexis)	RQ1.1	RQ1.2	RQ1.3
Student autonomous revision (Between-draft text quality change in content, organisation, grammar, and lexis)	RQ2.1	RQ2.2	RQ2.3
Text quality of new writing (Quality of content, organisation, grammar, and lexis)	RQ3.1	RQ3.2	RQ3.3

3.7 Design of the Study

Table 3.3 summarises the study design (for a time frame, see Table 3.7). This study comprised three stages: a pretreatment stage (Stage 1), a treatment stage (Stage 2), and a posttreatment stage (Stage 3). The three stages encompassed five two-draft writing-revising tasks, with one task (Task 1) for Stage 1, three (Tasks 2–4) for Stage 2, and one (Task 5) for Stage 3. The original plan was to include four writing-revising tasks for Stage 2. However, as the study proceeded halfway, some participants informed me that they could not attend Stage 3 of the study because they would be engaged in the preparation for the end-of-year celebrations of their respective departments. As a result, a task which had been planned for Stage 2 was dropped, and the posttreatment task was implemented one week in advance. The writing-revising tasks in Stages 1 and 3 did not involve intervention; all the participants revised their original writing without WF or revision-focus direction. During Stage 2 participants completed writing-revising tasks according to the group conditions they were assigned to (see Figure 3.1).

Table 3.3 Global design of the study

Pretreatment stage (Task 1)	All participants wrote and revised without WF or RFD
Treatment stage (Tasks 2–4)	Participants wrote and revised based on group conditions
Posttreatment stage (Task 5)	All participants wrote and revised without WF or RFD

Table 3.4 illustrates how the design served to address the research questions. Task 1 was intended to obtain participants’ written drafts to measure their writing proficiency and their autonomous revision before the treatment as a baseline measure. Tasks 2–4 were used to test the immediate effect of treatment on student revision. Task 5 was planned to examine the treatment effect on autonomous revision. In addition, the writing session of each task was used to obtain student first drafts on a new topic in order to examine the development of text quality in new writing.

Table 3.4 Correspondence of the writing-revising tasks to the research questions

	Task 1		Tasks 2–4		Task 5	
	writing	revising	writing	revising	writing	revising
RQ1 (Immediate effect of treatment on revision)			√	√		
RQ2 (Effect of treatment on autonomous revision)	√	√			√	√
RQ3 (Effect of treatment on new writing)	√		√		√	

3.8 Validating the Instruments and Procedures for Data Collection and Data Analysis

Before moving on to report the instruments and procedures for data collection and data analysis in the main study, this section will briefly describe how a pilot session tested their feasibility, reliability, and validity, and how this piloting informed the methodology for the main study.

For this trial 48 participants were recruited, assigned into the four treatment conditions (see Figure 3.1), and completed one writing-revising task. Initial analysis was performed on the collected written drafts.

3.8.1 Participants in the pilot session

The population for the main study was Chinese upper-intermediate university EFL students. The participants for the pilot session and the main study were recruited from the same university in Nanjing, China (hereafter University X, for convenience of reference). A detailed account of the English curriculum in University X is given in section 3.9.1.

The participants in the pilot session were two intact Level 3 English classes of University X, with 25 Physics students and 23 Journalism students respectively. The two classes shared the same English Reading & Writing teacher. This pilot was conducted during their regular lessons as part of their writing practice. The Physics students were randomly assigned to the treatment conditions with revision-focus direction (ID+RFD, IS+RFD) and the Journalism students to the other treatment conditions without revision-focus direction (ID-RFD, IS-RFD). All the 48 students participated throughout the writing-revising task. No control group was included for the pilot session.

3.8.2 Instruments in the pilot session

Instruments to be employed for collecting data for the main study included a questionnaire to collect participants' background information and writing-revising tasks. A detailed description is provided in section 3.8.2. Given the significance of the writing-revising tasks and my uncertainty of the validity and reliability, one of the writing-revising tasks was piloted. This writing-revising task required the participants to write an argumentative essay and revise it one week later. The writing prompt and the revising prompt were as follows.

The writing prompt:

You are allowed 30 minutes on this task.
Write at least 150 words. Do NOT consult any dictionaries, books, or other references.

Write about the following statement:

When college graduates seek jobs, interest should be their top concern because one will not devote himself to his work unless he is doing what he really loves.

To what extent do you agree or disagree with this statement?

Audience/Readers: Your teacher and fellow students

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

You may take notes on the blank sheet. Please write your essay on the lined sheet every second line.

The revising prompt for the –RFD group conditions (ID–RFD, IS–RFD):

You will have 25 minutes for revising your first draft.
Do NOT consult any dictionaries, books, or other references.

Here is a copy of your first draft with teacher feedback on the content, organisation, and language in your writing.

First, please read through your original draft and the feedback on it. Then, revise your draft and pay attention to such aspects as content, organisation, and language (grammar and word choice).

You may take notes on the blank sheet. You may also write on the copy of your first draft. When you finish, please transcribe your revised version on the lined sheet provided to you. Please write every second line.

The revising prompt for the +RFD conditions (ID+RFD, IS+RFD):

You will have 25 minutes for revising your first draft.
Do NOT consult any dictionaries, books, or other references.

You will first receive a copy of your first draft with teacher feedback on content and organisation in your writing. Please read through your original draft and the feedback. Then, you have around 10 minutes to revise the content and organisation of your draft.

Next, you will receive another copy of your first draft with teacher feedback on the language in your writing. Again please read through your original draft and the feedback on it. Then, you have around 10 minutes to improve your language (grammar and word choice).

You may take notes on the blank sheet. You may also write on the copies of your first draft. When you finish, please transcribe your revised version on the lined sheet provided to you. Please write every second line.

3.8.3 The data collection procedure in the pilot session

The data were collected during the participants' normal lessons with the assistance of their teacher. Table 3.5 summarises the procedure. In one of their lessons the participants signed a consent form, as required by the University of Auckland Human Participants Ethics Committee, and they wrote the first draft in response to the writing prompt.

Their first drafts were collected and copied. Written feedback was given on the copies according to treatment conditions and the self-developed linguistic error categories as well as standards for distinguishing between ID and IS feedback on discourse levels (for a detailed account of the linguistic error categories and categories of discourse-level feedback, see section 3.9.3.3). A maximum of 5 feedback points were provided for discourse and a maximum of 10 feedback points for linguistic errors. It turned out that on average 15 minutes were spent writing feedback on a single draft and in total around 12 hours for the entire feedback work on the 48 first drafts.

In the following week the participants received the copies of their first drafts with WF and revised them based on the revising prompt in accordance with their group condition.

The procedure for implementing writing-revising tasks for the main study is presented in detail in section 3.8.3.

Table 3.5 Data collection procedure for the pilot session

Week 1	Interval	Week 2
Ethics documents Writing Draft 1 (30 minutes)	Giving written feedback according to participants' treatment condition	Participants revised Draft 1 and completed Draft 2 (25 minutes)

3.8.4 Data analysis in the pilot session

The complete data analysis should have consisted of textual analysis and follow-on statistical analysis, as conducted for the main study (see section 3.9.4 for a detailed description). Given that this trial session was mainly intended to assess the feasibility and validity of the instruments and procedures for data collection and analysis, textual analysis was only conducted on a sample of collected draft pairs (i.e., first drafts and their revised versions). The sample contained 10 draft pairs randomly selected from the four treatment groups, constituting 21% of the total 48 draft pairs collected. The sample drafts were mixed and then analysed in terms of their quality on discourse and linguistic levels. An assistant helped me with the analysis. The assistant was an associate professor of University X, not the teacher of the participants. She had over 20 years' EFL teaching experience and had earned a PhD degree in applied linguistics. She assisted in data analysis for the main study as well as for this pilot session.

3.8.4.1 Assessing quality of content

Content was assessed using Ong's (2013) scheme (see Appendix C) designed to assess ideas in EFL students' argumentative writing. The assistant and I each independently rated a batch of 10 drafts after we examined and discussed the rating scheme. The interrater reliability computed using intraclass correlation coefficient turned out inadequate (.604).³ We discussed the differences, finding the weakness of the rating

³ The interrater reliability is considered acceptable if the intraclass correlation is larger than .7, and the reliability is excellent if the correlation value reaches .9 (Fleiss & Cohen, 1973).

scheme. It did not provide the scale for separate constituents of the overall quality of content, which contributed to much of our disagreement. In addition, it placed much weight on counterargument, but the trial rating revealed that a majority of the drafts lacked this element. Therefore, I revised the scheme accordingly. The assistant and I rated another 10 drafts by the revised scale and reached a higher agreement (.710).

3.8.4.2 Assessing quality of organisation

Organisation was measured against a scale targeted at the macrostructure and connectedness in writing, which was developed based on research related to writing assessment (e.g., Nimehchisalem & Mukundan, 2011; Watson Todd, 2007).

The assistant and I rated the drafts in two batches. The rating of the first batch of 10 drafts resulted in a low (.672) interrater reliability. The drawback in the rating scheme was found to be a lack of separate criteria for different components of organisation, so the scheme was modified. The subsequent corating round based on the new scale gained an interrater reliability of .832.

3.8.4.3 Assessing grammatical accuracy and lexical accuracy

Both grammatical accuracy and lexical accuracy were measured by a ratio: the total number of errors per 100 words. I started with drawing up guidelines for error-coding and trained the assistant how to apply the guidelines. Next we performed error-coding with a batch of 10 drafts independently, and the intraclass correlation was .611 for grammatical accuracy and .549 for lexical accuracy. When discussing the disagreements, we found that we tended to disagree with the counting of grammatical errors and neglect some grammatical errors, and that we sometimes differed in the judgment concerning prepositions and pronouns. I amended the coding guidelines accordingly. Then we continued coding with the other 10 drafts with increased sensitivity to grammatical errors. This time the intraclass correlation increased to .878 and .653. Again we resolved the differences by discussion.

3.8.5 Changes to the main study

The trial procedure throughout this pilot session validated some of the methodological propositions and meanwhile revealed several problems and limitations of the original methods, which informed the following changes to the main study.

1. Instruments: prompts and time limit for the writing-revising tasks
 - a. In the writing prompt a sentence would be added intending to remind participants to consider the given statement more critically. The collected essays indicated that many participants assumed the supporting element in the given statement to be truth and therefore failed to argue about that element. The added reminder would be: You can respond to both the claim and the supporting reason in the statement.
 - b. In the writing prompt and the revising prompt, participants would be told that a title was unnecessary for their essay. It was found that a few of the collected essays included a title while many others did not. It was hard to decide whether to count the titles into the text length, so it was decided to control this variance for the convenience of judging word length.
 - c. The time limit for both writing and revising sessions would be extended to 35 minutes. It was noted that the original time limit was pressing for around half of the participants, who were found writing in a hurry when the time ran out.
2. Data analysis: rating scales and guidelines for error-coding
 - a. The revised versions of the rating scales for content and organisation would be used for the main study (see Appendix C), because they had manifested an advantage in reliability over the former ones.
 - b. To achieve a higher level of agreement, guidelines for error-coding were amended based on our coding experiences and discussions.

3. The setting of the study

- a. The main study would still recruit a sample from the first-year undergraduates of University X but would be conducted out of class as practice supplementary to participants' regular English lessons. The English teacher involved in the pilot indicated that although a single writing-revising task could be integrated in her normal teaching, a planned six-week main study would interfere in her syllabus which was institutionally formulated and followed. Therefore, the main study would take place as participants' extracurricular writing practice. It was expected that a number of students would volunteer to participate in the main study given their motivation to improve English writing proficiency, particularly to prepare for the College English Test which is compulsory for all the non-English-major Chinese undergraduates in University X.

3.9 The Main Study

Research questions, variables, and research design for the main study have been presented in sections 3.3–3.7 of this chapter. This section will continue to describe the main study regarding its population and sample, instruments, and procedures for collecting and analysing data.

3.9.1 Population and sample

The population was Chinese upper-intermediate university EFL students. The sample was drawn from the first-year undergraduates at University X in China. In this university, English is a compulsory subject for all first-year non-English majors. English classes are organised based on students' majors and English levels, English levels being decided by students' scores on a placement test. The students take the test, developed by the Dept. of Applied Foreign Language Studies of the university, upon their entrance into the university. Based on their scores, they are classified into Level 1, Level 2, and Level 3, in the order of decreasing proficiency. Usually the majority go to Level 2 and Level 3. The test consists of listening, reading, and writing, and writing accounts for 30% of the total score. All the non-English majors, irrespective of the level they are placed in, take a Reading & Writing course and a Listening & Speaking course with two

corresponding teachers. In the two semesters of their first academic year, they have the two English courses on separate days per week, two hours for each course.

To exert an initial control for participants' English proficiency and potential extraneous factors from participants' teachers, the sample was selected from four Level 3 classes taught by the same Reading & Writing lecturer (the one whose other two classes participated in the pilot session) and the same Listening & Speaking lecturer. According to both teachers, the students in the four classes had a similar level of English proficiency and most of them could be classified as upper-intermediate English learners in China. I read essays of those students for an in-class writing assignment and agreed with the teachers' judgment. It would have been good to carry out a more accurate verification, though.

This study was conducted during the students' extracurricular time, as a supplement to their English lessons. The study lasted from Week Six to Week Ten of the 15-week semester according to the University's calendar.

Altogether 94 students were originally recruited, including 20 History students, 23 Business Management students, 24 Sociology students and 27 Engineering students. The participants aged between 17 and 19, and all had received their previous English education in mainland China. They had on average learned English for around 8 years. The ratio of female to male was 64% to 36%.

The 94 students were randomly classified into the five group conditions based on their English proficiency manifested in Task 1 (see section 3.9.3.1 for the grouping procedure). As a result 18 students were assigned to each treatment condition and 22 students to the control condition.

To counteract possible attrition, the students late to a session were allowed to complete it, and if possible alternative time periods were arranged for those who could not attend a prescribed session but indicated willingness to make up for that. Throughout the study the number of participants reduced moderately, though. Several students decided to withdraw their participation during the data collection, and some others missed certain sessions and did not make up for them. In addition, since this study was conducted as extracurricular practice, the students might not feel obliged to attend every session as they would for their normal classes. As a result, 77 students completed all the writing-

revising tasks. These students' data were used for data analysis, and the other students were removed. Table 3.6 illustrates the number of participants in each group condition of the study. I would clarify here that the attrition did not affect the results of initial stratified sampling based on participants' English proficiency. Subsequent statistical analyses indicated that no significant differences emerged across the five groups in the proficiency of any of the textual levels assessed.

Table 3.6 Number of participants in each group condition

Treatment condition				Control condition
ID+RFD	ID-RFD	IS+RFD	IS-RFD	
<i>n</i> = 15	<i>n</i> = 16	<i>n</i> = 15	<i>n</i> = 14	<i>n</i> = 17

3.9.2 Instruments

Two types of instruments were employed to collect data: a questionnaire and writing-revising tasks (see Appendix B).

3.9.2.1 Questionnaire

A background questionnaire was used to collect the participants' information regarding their age, gender, major, length of English learning, the place where they learned English before attending the university, and their experience of living in English-speaking environment. The questionnaire was anonymous and written in Chinese.

3.9.2.2 Writing-revising tasks

This section will explain the genre of the writing-revising tasks and the selection of writing topics and then provide examples of prompts for the tasks.

Argumentative writing was chosen for the writing tasks. For one thing, it is the genre that university students, including the participants in this study, frequently deal with in the academic context including standard English proficiency tests such as the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL). For another, it poses the greatest difficulty for most

university students in both English-as-an-L2 and English-as-an-L1 countries (Butler & Britt, 2011; Qin & Karabacak, 2010; Song & Ferretti, 2013).

The writing topics were selected following a few steps. First, 14 potential topics were devised; then two Reading & Writing teachers and two Listening & Speaking teachers at University X commented whether their students would be interested in and knowledgeable about these topics. The teachers also confirmed that the students had not written and would not write on any of these topics for their coursework during the semester. Based on their comment, 4 of the topics were removed and the remaining 10 topics were revised and further evaluated in terms of difficulty level. Specifically, the four teachers and some of their students assigned a number from 1 to 9 to each topic, indicating its difficulty level (with 1 representing the greatest ease and 9 the greatest difficulty). The most difficult and easiest topics were excluded, and the remaining six topics were reserved for the study. Only five topics were actually used because of the reduction of one writing-revising task (see section 3.7).

Each of the five tasks required participants to write an essay and revise it. The writing prompts were basically the same except for the topic, while the revising prompts differed with respect to the revising condition (i.e., control, +RFD, -RFD). The following are some examples.

A writing prompt:

You are allowed 35 minutes on this task.
Write at least 150 words. Do NOT consult any dictionaries, books, or other references.

Write about the following statement:

When college graduates seek jobs, interest should be their top concern because one will not devote himself to his work unless he is doing what he really loves.

To what extent do you agree or disagree with this statement? You can respond to both the claim and the supporting reason in the statement.

Audience/Readers: Your teacher and fellow students

Give reasons for your answer and include any relevant examples from your own knowledge or experience. A title for your essay is not necessary.

You may take notes on the blank sheet. Please write your essay on the lined sheet every second line.

A revising prompt for the control condition:

You will have 35 minutes for revising your first draft.
Do NOT consult any dictionaries, books, or other references.

Here is a copy of your first draft.

First, please read through your original draft. Then, revise your draft and pay attention to such aspects as content, organisation, and language (grammar and word choice).

You may take notes on the blank sheet. You may also write on the copy of your first draft. When you finish, please transcribe your revised version on the lined sheet provided to you. Please write on the lined sheet every second line.

A revising prompt for the –RFD condition (ID–RFD, IS–RFD):

You will have 35 minutes for revising your first draft.
Do NOT consult any dictionaries, books, or other references.

Here is a copy of your first draft with teacher feedback on the content, organisation, and language in your writing.

First, please read through your original draft and the feedback on it. Then, revise your draft and pay attention to such aspects as content, organisation, and language (grammar and word choice).

You may take notes on the blank sheet. You may also write on the copy of your first draft. When you finish, please transcribe your revised version on the lined sheet provided to you. Please write every second line.

A revising prompt for the +RFD condition (ID+RFD, IS+RFD):

You will have altogether 35 minutes for revising your first draft.
Do NOT consult any dictionaries, books, or other references.

You will first receive a copy of your first draft with teacher feedback on the content and organisation in your writing. Please read through your original draft and the feedback. Then, you have around 15 minutes to revise the content and organisation of your draft.

Next, you will receive another copy of your first draft with teacher feedback on the language in your writing. Again please read through your original draft and the feedback on it. Then, you have around 15 minutes to improve your language (grammar and word choice).

You may take notes on the blank sheet. You may also write on the copies of your first draft. When you finish, please transcribe your revised version on the lined sheet provided to you. Please write every second line.

3.9.3 The procedure for data collection

The data were collected through three stages over five consecutive weeks (see Table 3.7). In Stage 1 after being informed of the ethical issues as well as the research design, the participants signed the consent form according to the regulations of the University of Auckland Human Participants Ethics Committee (see Appendix A for the documents).

Then the participants completed the background questionnaire. Next they completed the first writing-revising task. In Stage 2 the participants completed Tasks 2 –4 after being assigned to treatment and control groups. In Stage 3 the participants completed Task 5. The writing topics are provided in Appendix B.

Table 3.7 Data collection procedure for the main study

Stage 1 pretreatment	Week 1: Task 1 (interest and job seeking)	Day 1	Ethics documents; Background questionnaire; Writing session	Grouping the participants
		Day 4	Revising session	
Stage 2 treatment	Week 2: Task 2 (giving arms to beggars)	Day 1	Writing session	treatment
		Day 4	Revising session	treatment
	Week 3: Task 3 (ancient Chinese language)	Day 1	Writing session	treatment
		Day 4	Revising session	treatment
	Week 4: Task 4 (use of social media)	Day 1	Writing session	treatment
		Day 4	Revising session Perception questionnaire	treatment
Stage 3 posttreatment	Week 5: Task 5 (animal testing)	Day 1	Writing session	
		Day 4	Revising session	

Note. The key words of the writing topics are put in parentheses.

As Table 3.7 presents, the writing and revising sessions for each task had a three-day interval. Based on my discussion with the participants before the data collection started, the writing and revising sessions took place on Monday and Thursday afternoons respectively, when they had no classes. All the writing sessions started at 4:30 p.m. The revising session in Task 1 and Task 5 started at 4:15 p.m.; those in Tasks 2–4 started at 4:15 p.m. for the +RFD groups and 5:15 p.m. for the –RFD groups and the control group, making it feasible to administer the sessions according to different revising conditions.

However, there were a few cases of make-up sessions for the participants who failed to comply with the schedule.

The major steps in the data collection procedure are presented in detail as follows with respect to the grouping procedure, the writing sessions, the feedback procedure, and the revising sessions.

3.9.3.1 Grouping the participants

With respect to the grouping of participants into treatment/control conditions, related studies adopted different methods to ensure similar target language proficiency across conditions at the outset of treatment. Often they inspected student performance on a certain pretest, such as use of a target feature in a writing task (Frear, 2012), performance on a vocabulary test (Van Beuningen et al., 2012), and global scores on a set of writing tasks (Ong & Zhang, 2010). It seems that there is no standardised or ideal method to predict a learner's language proficiency.

In this study the initially recruited participants were grouped based on the linguistic accuracy in their first draft of Task 1. Specifically, the measurement of the students' accuracy in grammar and lexis by the method used in the pilot session resulted in an accuracy score (i.e., the total number of grammatical and lexical errors per 100 words) for each student. The students were stratified into three tiers of linguistic proficiency according to their accuracy score, high (21 students with a score of less than 5), average (62 students with a score of between 5 and 10), and low (11 students with a score of larger than 10). Then the students of each tier were randomly assigned to the five group conditions. A few more participants were placed in the control group because of a predicted higher attrition rate in that group due to the lack of external feedback on their writing during the treatment stage.

It was recognised that the result of this grouping was subject to change because of the withdrawal of the students throughout the data collection process. It was hard to exert an absolute control for the equal linguistic proficiency across the groups.

3.9.3.2 The writing session for the five tasks

The writing sessions for the five tasks were administered following the same procedure. First, each participant received a package of materials, including the writing prompt, a blank sheet for note taking and a lined sheet for drafting. In the meantime, they received oral directions about the writing task and were informed of the chance for revision three days later. After they started writing, the participants were monitored and reminded of the remaining time. Finally, their notes and first drafts were collected.

3.9.3.3 Giving feedback for Tasks 2–4

During Stage 2 feedback was given to the first drafts of treatment groups according to their conditions, i.e., ID+RFD, ID–RFD, IS+RFD, and IS–RFD. Meanwhile the linguistic error types targeted for feedback (see Table 3.8) as well as the standards for distinguishing discourse-level ID feedback from IS feedback (see Tables 3.9 and 3.10) were referred to, which were developed and trialled during the pilot session.

As in Ashwell's (2000) study the amount of feedback offered to any piece of draft was controlled to mitigate its possible intervening effect on student revision and text quality change. While Ashwell imposed a uniform time limit within which feedback was given to each draft, such a manipulation might not apply here because the ID feedback might consume more time than the IS method. Instead the numbers of feedback points were controlled at discourse and linguistic levels, although this arrangement might not be an exact way of standardising the amount of feedback.

A maximum of 5 feedback points were given on discourse levels and a maximum of 10 feedback points on language, as practiced in the pilot session. This operation was informed by a browse of the students' coursework essays and the University teacher's knowledge about the average number of problems in a student's essay. In addition, it was practical considering the workload involved. Severe and repeated problems were prioritised during the process of feedback provision. However, I did not give feedback on each of the recurring problems; I tended to comment on one of them hoping that the students could understand the feedback and apply the information to revising other problems of the same nature. The feedback work was all completely by me, which took approximately 15 hours in total for one task with around 15 minutes for a single draft. Sample drafts and feedback are provided in Appendix D.

Table 3.8 Linguistic error categories targeted for feedback

Category	Description and examples
Grammar	
- part of speech	Using the wrong part of speech <i>He successfully became a speak. / What he likes particular is teaching. / protect animals from being extincting</i>
- verb	All errors in verb tense or form, including relevant subject-verb agreement errors <i>I will lost all my enthusiasm. / oppose to treat them in this manner / after being testing</i>
- noun ending	Plural or possessive ending incorrect, omitted, or redundant; includes relevant subject-verb agreement errors <i>many deadly illness / many researches / Their organs are similar to people. / in the history of human's development</i>
- article	Article or other determiner incorrect, omitted, or redundant <i>choose the job in bank / in a long run</i>
- pronoun	Pronoun incorrect, omitted, or redundant <i>all you life / hard for us to find a job what we really love</i>
- preposition	Preposition incorrect, omitted, or redundant <i>Despite whether what we do is what we really love, we should try out best to make it best. / make decisions which we will never regret of / We can't risk of harming a person.</i>
- the comparative/ superlative	Errors in the comparative or superlative <i>more cruel</i>
- sentence structure	Errors in sentence/clause boundaries (run-ons, fragments, comma splices), word order, omitted words or phrases, redundant words or phrases, dangling modifiers, other unidiomatic sentence construction <i>Because you love it, you keep doing it, you will be successful at last. / When seeking jobs, interest should be considered first. / Scientists choose to do tests on animals is out of the notion that they should reduce the harm to people when doing experiments on human beings.</i>
Lexis	

- wrong word/phrase	All specific lexical errors in word or phrase choice, including relevant preposition and pronoun errors; errors involving connotation, collocation, redundant or incomplete information; spelling errors only included if the misspelling resulted in an actual English word <i>The number of job-hunters is much more than that of jobs. / as the saying says / insist of working / Many medicines have been tested by animals.</i>
---------------------	---

Note. Examples are put in italics.

Table 3.9 Criteria and examples for providing ID and IS feedback on content

Feedback on content	ID	IS
Distinction	Stating or reiterating the expectations and rubrics; not explicitly relating the rating criteria to student text.	Telling students what to do with close reference to their text; offering specific suggestions; not including the rating criteria, i.e., the reason behind the suggestions.
Examples		
1.	<i>English argumentation requires a statement of a central idea which specifies the writer's overall opinion on the topic addressed.</i>	<i>Summarise your opinions expressed in different parts of your writing and put them in one sentence as your central idea.</i>
2.	<i>In English argumentation, a writer is expected to develop two or more main reasons to support his/her central idea.</i>	<i>Think about another reason to support your central idea. You may consider the cultural value of Ancient Chinese.</i>
3.	<i>Each reason should be elaborated by examples, reasoning, or statistics.</i>	<i>Add something to support this reason. You may take Tang poems as an example.</i>
4.	<i>The supporting materials should be related to and logically support your reason.</i>	<i>Replace this example with another one, because it is not related to the idea you want to support.</i>
5.	<i>Effective argumentation includes counterarguments and rebuttals.</i>	<i>Some people may propose counterarguments, e.g., "We can learn the message in classic works by reading translated versions." How do you argue against such counterarguments?</i>

Table 3.10 Criteria and examples for providing ID and IS feedback on organisation

Feedback on organisation	ID	IS
Distinction	Stating or reiterating the expectations and rubrics; not explicitly relating the rating criteria to student text.	Telling students what to do with close reference to their text; offering specific suggestions; not including the rating criteria, i.e., the reason behind the suggestions.
Examples		
1.	<i>English argumentation includes an introduction, which arouses readers' interest and leads them to the topic and argument.</i>	<i>Add an introduction to your article. You may write a brief story, a quote, or some background information.</i>
2.	<i>The sequencing of supporting materials should be logical.</i>	<i>Reverse the order of the two examples.</i>
3.	<i>The reasons to support the central idea should be clearly organised.</i>	<i>Reorganise your reasons and supporting materials in this long paragraph, so that distinct reasons can stand out.</i>
4.	<i>The propositions should be connected to each other.</i>	<i>Develop the topic in the previous sentence, so it naturally transfers to this sentence.</i>
5.	<i>The connection between adjacent sentences should be clear and evident to readers.</i>	<i>Use a transitional word or phrase here. / Here repeat the subject in the previous sentence instead of using "it".</i>
6.	<i>In the body part, one paragraph should develop only one idea.</i>	<i>This long paragraph includes two ideas. Put them into separate paragraphs.</i>
7.	<i>English argumentation includes a conclusion, which often reiterates the main arguments or makes an appeal.</i>	<i>Add a conclusion paragraph. You may summarise your major arguments.</i>

3.9.3.4 The revising session for the five tasks

The revising session was administered based on the design of the three stages and group conditions. In Stage 1 (Task 1) and Stage 3 (Task 5) all the participants revised their writing with no WF. In Stage 2 (Tasks 2–4) the control group revised drafts with no WF while the treatment groups completed the revising session with WF under either the +RFD or the –RFD condition.

3.9.3.4.1 The revising session with no WF

The students received a package of materials, including a copy of their first draft with no WF, a revising prompt, a blank sheet for note taking and a lined sheet for transcribing their second draft. After receiving oral directions, the students started the revising process. They were informed when 10 minutes and 25 minutes passed. Thirty-five minutes later, the participants' notes, second drafts, and copies of the first draft were all collected.

3.9.3.4.2 The revising session for the +RFD condition (ID+RFD, IS+RFD)

The students received a package of materials, including a copy of their first draft with WF on discourse levels, a revising prompt, a blank sheet for note taking and a lined sheet for transcribing their second draft. Meanwhile they received oral directions. Then the students started revising. Around 15 minutes later they received another copy of their first draft with WF on linguistic levels and were reminded to start working on their language. They were informed of the passing time. When the entire revising process lasted for 35 minutes, the relevant materials were collected.

3.9.3.4.3 The revising session for the –RFD condition (ID–RFD, IS–RFD)

The students received a similar package of materials to that received by the students under the +RFD condition. The only item different was that the copy of their first draft contained WF on both discourse and linguistic levels. The students started revising after receiving oral directions. They were also reminded of time. Thirty-five minutes later, all the relevant materials were collected.

3.9.4 Data analysis

Analysing the text quality of the collected drafts preceded other types of analysis. Based on the outcome of the textual analysis, statistical analysis was performed to answer the research questions. Next, a supplementary textual analysis was conducted focusing on the discrete-point changes between the first and the second drafts; the purpose was to triangulate the results for RQ1 (i.e., the immediate effect of treatment on student revision) which derived from statistical analysis.

3.9.4.1 Textual analysis of drafts

The assistant in the pilot session helped me with the textual analysis of the drafts. We dealt with the five stacks of drafts collected from the five tasks one by one. Since the five stacks of drafts were approached following the same procedure basically, only the procedure for analysing the drafts collected from one task will be presented.

The textual analysis of the draft pairs targeted the text quality of each draft. The drafts were photocopied, and photocopies were mixed and then stapled with information about the student's name, treatment condition, and draft sequence being concealed. Next the copies of drafts were numbered for convenience of record. As a result, we did not know whether we were rating an initial draft or a second one when rating each essay. The text quality on discourse and linguistic levels was rated in turn (see sections 3.9.4.1.1–3.9.4.1.3).

3.9.4.1.1 Assessing quality of content

Content was assessed using the modified version of Ong's (2013) scheme that had been piloted and adjusted (see Appendix C: Rating scheme for quality of content used in the main study). Ong's scheme was drawn on because it was designed to assess ideas in argumentative writing and conceptualised the quality of ideas in the same way as this study did, i.e., focusing on the overall persuasiveness. In addition, it was developed for a similar population, i.e., Asian university EFL students, or specifically, Chinese students. Further, it was relatively specific and simple to compute compared with some other scales for assessing content (e.g., East, 2009; Ferretti et al., 2000; Nimehchisalem & Mukundan, 2011).

We each independently rated a batch of 15 drafts (i.e., 9.7% of the total 154 drafts) after re-examining the rating scheme, reaching an adequate interrater reliability (.707) computed using intraclass correlation coefficient. After solving the disagreements, we rated another 15 drafts individually and reached a higher agreement (.754). Then we discussed the differences and shared the marking of the remaining drafts. Hence each draft received a content score.

3.9.4.1.2 Assessing quality of organisation

Organisation was measured on the macrostructure and connectedness in writing. It was measured against the scale (see Appendix C: Rating scheme for quality of organisation used in the main study) which was primarily developed based on related research (Lee, 2002; Liu & Braine, 2005; Nimehchisalem & Mukundan, 2011; Watson Todd, 2007) and later modified after being piloted.

Following the procedure for rating content, we rated the drafts in three batches. Our independent rating of the two batches of drafts resulted in acceptable reliability, the intraclass correlation being .712 and .815 respectively. As a result, each draft was given an organisation score.

3.9.4.1.3 Assessing linguistic accuracy

Both grammatical accuracy and lexical accuracy were measured as in the pilot session; the index was the total number of errors per 100 words, as the common measurement for linguistic accuracy in CF studies (e.g., Ashwell, 2000; Chandler, 2003). The index of error-free T-units was not used because of its disregard for the number of errors within a T-unit (Polio, 1999). The word length of every draft was counted before the error-coding work.

With respect to error-coding, we began with reviewing the guidelines (see Appendix C) amended during the pilot session, which were initially developed according to the error categories targeted by the WF and Yang's (2016) error-coding scheme (for Yang's scheme, see Appendix C). Next we performed error-coding with a batch of 10 drafts (i.e., 6.5% of the total 154 drafts) independently, and the intraclass correlation was .765 for grammatical accuracy and .679 for lexical accuracy. We discussed and solved the disagreements before continuing with another 10 drafts. This time the intraclass

correlation improved to .828 and .718. Finally, we each marked half of the remaining drafts. Therefore, each draft got a grammatical accuracy score as well as a lexical accuracy score.

3.9.4.1.4 An example of a student's drafts and their scores

This subsection provides an example of a student's essay and its scores for different aspects. The following are two drafts a student completed for Task 2, which required students to argue about the statement: People should not give alms to beggars because many of them have the ability to make a living. This student wrote in the IS+RFD condition. The first draft had a score of 6 for both content and organisation, and the second draft had a score of 7 for both content and organisation. The grammatical accuracy scores for Draft 1 and Draft 2 were respectively 3.06 and 2.25. The lexical accuracy scores for Draft 1 and Draft 2 were respectively 2.55 and 1.35. As explained in section 3.9.4.1.3, each accuracy score was obtained by first dividing the number of errors by the word length and then multiplying the quotient by 100. To label the linguistic errors in the following sample drafts the grammatical errors are underlined and numbered, and the lexical errors are italicised and numbered.

Draft 1: 196 words, 6 grammatical errors, 5 lexical errors

There are ¹*more* beggars in the streets and we don't know when it started. Are more people becoming poor today? Obviously ²*we* are not. ³*That's* because more people want to get money without any pain.

Should we give alms to beggars? I'm afraid that we shouldn't. Let me explain.

¹In addition that many of the beggars have ability to make a living, the most crucial thing is that some adults hit some children and make them disabled and then let them beg in the streets. That's really incredible!

You may say there are some beggars who are really disabled. But I have to say, our government has taken action to guarantee their living places and ⁴*costs*. As people ²becoming rich, they give too much money to the beggars and that ³leads beggars even richer than ordinary people. So it is not ⁵*amazing* that some people join in beggars though they are healthy.

To help beggars, what we should do is ⁴help them with their living ⁵not giving them too much money. At the same time, our government should take more measures

to create a more comfortable environment for them. So no giving alms⁶, just help them.

Draft 2: 222 words, 5 grammatical errors, 3 lexical errors

There are more beggars in the streets than before and we don't know when the phenomenon started. Are people becoming poor today? Obviously no. ¹*That's* because people want to get money without any effort. Should we give alms to beggars? I'm afraid that we shouldn't.

First, many beggars have ability to make a living. But they don't want to work. Instead, they choose to beg. And as people become rich, they give too much money to beggars instead of help and they think money is the best help they can give. That's terrible.

In addition, something more than cruel ¹happened ²*since then*, surprising the whole society. A few adults beat some homeless children or even their own children and made them disabled. Then they forced the little children to beg. I couldn't say a word but ²felt ³extreme angry when I heard this.

You may say there are some true beggars who are really disabled. But I have to say, our government has taken action to guarantee their habitation and income. Their living ⁴are no longer problems.

What we ought to do is to help beggars with their living ⁵not only giving money. At the same time, it is necessary for the government to take more measures to create a more comfortable environment for them. So no giving alms. Just help them.

3.9.4.2 Statistical analysis

This section gives an overview of the statistical methods used in this study; in the following chapters that report results in response to specific research questions, relevant statistical analysis will be reviewed briefly.

The data resulting from the textual analysis were submitted to a series of statistical analysis using the IBM Statistical Package for Social Sciences (SPSS) 22 to address the research questions. Specifically, the numeric data were transcribed into a dataset, containing the first-draft (Draft 1) and second-draft (Draft 2) scores for text quality on the four textual levels in each of the five writing-revising tasks.

3.9.4.2.1 Addressing RQ1

The data related to Tasks 2–4 were used to answer RQ1 concerning the immediate effect of treatment on student revision. The data for each of the three tasks were submitted to two sets of repeated-measures mixed model ANOVAs and relevant pairwise comparisons, which were applied to each of the four textual levels. The first set of ANOVAs was meant to compare the effects of the five conditions on text quality across drafts (RQ1.1), with draft (Draft 1 and Draft 2) being the within-subjects variable and group condition being the between-subjects variable. The second set of ANOVAs was used to assess the effects of WF type and \pm RFD on revision by treatment groups (RQ1.2), and data of the control group were excluded. These ANOVAs had draft as the within-subjects variable and WF type (ID, IS) as well as \pm RFD as between-subjects variables. To further explore within-subjects differences and between-subjects differences, pairwise comparisons with Bonferroni adjustment were also computed by editing SPSS syntax (Larson-Hall, 2010). This computation was chosen instead of post hoc *t* tests, which some studies deployed to explore within- or between-subjects differences, because it was more efficient. Although pairwise comparisons were conducted automatically for the two-way ANOVAs, it is acknowledged that they were only warranted if the two-way ANOVAs yielded significant effects. Before each statistical test was run, the related data were examined in terms of distribution of normality by values of skewness and kurtosis⁴.

The mediation of textual level in treatment effect (RQ1.3) was answered by synthesising the results from statistical tests regarding each textual level. The treatment effect on the four textual levels was compared in terms of relevant statistically significant data and effect sizes.

3.9.4.2.2 Addressing RQ2

The data related to Task 1 and Task 5 were employed to answer RQ2 concerning the treatment effect on student autonomous revision. Effect on autonomous revision, assessed in terms of between-draft text quality change before and after the treatment

⁴ According to Field (2009), the conversion of the values into *z*-scores indicates whether the data were normally distributed. If the *z*-score of skewness or kurtosis is not greater than 1.96, the distribution is approximately normal.

stage, was operationalised as gain scores across Stage 1 (i.e., Task 1) and Stage 3 (i.e., Task 5). A gain score refers to the between-draft difference in the scores on a certain text level. Hence gain scores were calculated for Task 1 and Task 5 by subtracting the scores of Draft 1 from the scores of Draft 2 in each task.

Next the gain scores were subjected to statistical tests, which basically followed the practice adopted to address RQ1. Therefore, only the different part is elaborated below. Two sets of ANOVAs and relevant pairwise comparisons were performed for each of the four textual levels. The first set of tests was meant to compare the effect of the five conditions on autonomous revision (RQ2.1), with stage (Stage 1 and Stage 3) being the within-subjects variable and group condition being the between-subjects variable. The second set of tests to assess the effect of WF type and \pm RFD on autonomous revision of treatment groups (RQ2.2) had stage as the within-subjects variable and WF type as well as \pm RFD as between-subjects variables. Similar to the way RQ1.3 was addressed, RQ2.3 was answered by synthesising the results from statistical tests regarding each textual level.

3.9.4.2.3 Addressing RQ3

The first-draft scores in Tasks 1–5 were used to address RQ3 regarding the treatment effect on student new writing. Specifically, the first-draft scores in Task 1 and Task 2 were averaged to obtain baseline scores (coded as scores in Task 1/2), which were supposed to represent student text quality in new writing before the treatment. The baseline scores, together with the first-draft scores in Tasks 3–5 were submitted to statistical tests.

Similar to the practice described above, two sets of ANOVAs and pairwise comparisons were run addressing the effect of group conditions (RQ3.1) and the effect of WF type and \pm RFD on new writing (RQ3.2) respectively. In the first set of tests, task (Task 1/2, Task 3, Task 4, and Task 5) was the within-subjects variable and group condition the between-subjects variable. In the second set of tests, task was still the within-subjects variable, and WF type and \pm RFD were between-subjects variables. Finally, RQ3.3 was approached by synthesising the results from statistical tests regarding each textual level.

3.9.4.3 Post hoc textual analysis of drafts to address RQ1

To triangulate results for RQ1 from the statistical analysis, post hoc textual analysis was performed which targeted discrete-point changes between drafts. For convenience of analysis, only data collected from Task 3 were used as exemplary evidence. This approach was thought to be acceptable in view of the similar results across the three tasks in Stage 2. The analysis comprised key steps including: (a) identifying and counting discrete changes between each pair of drafts in Task 3, (b) categorising the changes as discourse/linguistic changes, and (c) matching the discrete changes with the feedback points in order to find whether a change was attributable to feedback and whether a feedback point was acted upon.

With respect to the textual level of between-draft revisions, the schemes of Ziv (1984) and Butler and Britt (2011) were consulted and followed. Revisions were classified into discourse changes, i.e., changes on conceptual and structural levels that brought a major change to content or organisation of the draft, and linguistic changes, i.e., changes on sentential and lexical levels that did not bring a major change to content or organisation. Changes to content and organisation were not distinguished because both were at the relatively global level, and furthermore, in some cases a change in content could alter the organisation and vice versa. For instance, addition of a conclusion paragraph improved the overall structure and meanwhile might inject new ideas to the essay. Similarly, grammatical and lexical changes were lumped together as linguistic changes. It was expected that at times students might correct a grammatical error by a lexical means and vice versa, where it would be difficult to categorise the change and to judge students' intention behind.

A list of guidelines (see Appendix C) was drawn up for coding between-draft revisions based on an examination of the students' draft pairs and other researchers' practice (Butler & Britt, 2011; Faigley & Witte, 1981; Graham et al., 1995; Wallace & Hayes, 1991). The guidelines specified how to count an instance of a between-draft change and how to categorise a change in terms of textual level.

Adhering to the guidelines, I identified between-draft revisions in the draft pairs and coded them in terms of textual level (discourse, linguistic). Finally, I calculated the number of between-draft revisions in total and on each of the two broad textual levels.

I also matched the discrete changes with the feedback points in order to find whether a change was attributable to a feedback point and whether a feedback point was acted upon.

CHAPTER FOUR

RESULTS (1): THE IMMEDIATE EFFECT OF TREATMENT ON REVISION

4.1 Overview

This chapter addresses the first set of research questions which focused on the immediate effect of treatment on student revision. The data used to answer the questions were based on the textual analysis of the written texts the participants produced in the treatment stage (Stage 2) of the study which encompassed Tasks 2–4.

RQ1 Does treatment have any immediate effect on student revision?

RQ1.1 What are the relative effects of different treatment methods on student revision?

This research question attempted to examine whether the treatment groups varied from each other and in comparison to the control group in terms of between-draft text quality change during the treatment stage. Two-way repeated-measures ANOVA tests were run to calculate the effects of draft and group condition on text quality. Following each ANOVA test, pairwise comparisons by editing SPSS syntax were conducted to examine whether significant within-subjects differences (i.e., between-draft text quality change by each group) and between-subjects differences (i.e., text quality difference between the groups for a certain draft) existed. The statistical analyses were performed for each task in Stage 2 (S2) and for each of the four textual levels.

Two alternative approaches had been adopted to test the immediate effect of treatment on revision in terms of text quality across drafts, and the results confirmed those of the tests which were chosen to be reported in this chapter. One alternative approach dealt with the three tasks in S2 as a whole instead of separately by taking the average of the three sets of first-draft scores as the original score (Score 1) and the average of the three sets of second-draft scores as the score after revision (Score 2). The other approach was based on a comparison of gain score (i.e., the difference between Score 1 and Score 2).

As an example, one test based on the first approach used two-way repeated-measures ANOVA with score (Score 1 and Score 2) as the within-subjects factor and group condition as the between-subjects factor. For another example, a test based on the

second approach employed one-way ANOVA with gain score (i.e., the difference between Score 1 and Score 2) as the dependent variable and group condition as the factor.

Both approaches might fail to reveal some detailed and important results. The defect of the first approach lied in that it masked the possible differences between the three tasks in the effect of treatments. The second approach, focusing on between-subjects difference (i.e., gain score), could not find significant between-draft text quality change within each single group. Therefore, it was chosen to report the results of the repeated-measures ANOVAs that addressed the three tasks separately.

RQ1.2 Do the factors of WF type and \pm revision-focus direction play a role in the immediate effect of treatment methods on student revision?

This research question investigated, on one hand, whether students receiving ID varied from students receiving IS in between-draft text quality change, and on the other hand, whether the +RFD students varied from the -RFD students¹. The analyses started with ANOVAs testing the effects of draft, WF type, and \pm RFD on text quality. Then pairwise comparisons examined whether there were significant within-subjects differences (i.e., between-draft text quality change by students receiving either type of WF, and that by students either receiving RFD or not receiving RFD) and between-subjects differences (i.e., text quality difference between students receiving ID and students receiving IS, and text quality difference between the +RFD and the -RFD students). The analyses also itemised the textual levels.

RQ1.3 Does textual level mediate the immediate effect of treatment?

This research question investigated whether the effects of different treatment methods vary with textual level, and whether the effects of WF type and \pm RFD vary with textual level. To answer this question, the results for all the textual levels in relation to RQ1.1 and RQ1.2 were synthesised; comparisons were drawn between the treatment effects on the textual levels by observing the relevant statistically significant data and effect sizes.

¹ In the chapters that report and discuss results, participants who received RFD and those who did not will be respectively termed as “the +RFD students” and “the -RFD students” for convenience of report.

From section 4.2 to section 4.5, results for the four textual levels will be successively reported. The report on each textual level will start with the outcome of statistical analyses regarding the effect of group condition (RQ1.1) and proceed to the outcome of analyses regarding the effects of WF type and \pm RFD (RQ1.2). Sections 4.6 and 4.7 will summarise the results in relation to RQ1.1 and RQ1.2 and present the results with respect to the mediation of textual level in treatment effect (RQ1.3).

4.2. Results for Content Quality

4.2.1 The immediate effect of group conditions on content quality across drafts

Table 4.1 displays the descriptive results for content quality in the three tasks in S2 by draft and group condition. It shows that all the five groups increased their content scores on the 9-point scale from Draft 1 to Draft 2 in each of the three tasks. The z -scores of skewness and kurtosis (see Footnote 4 of Chapter Three) for abnormal distribution are noted in footnotes for this table as well as the other tables reporting descriptive statistics.

Table 4.1 Means and standard deviations for content quality in Tasks 2–4 by draft and group condition

Group	Draft	Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	D1	5.33	1.11	5.87	1.13	5.93	0.80
	D2	5.87	0.64	6.67	1.05	6.73	0.70
ID–RFD (<i>n</i> = 16)	D1	5.50	1.16	5.44	1.15	5.88	0.96
	D2	5.81	1.05	5.94	1.18	6.50	0.89
IS+RFD (<i>n</i> = 15)	D1	5.60	1.12	5.67	1.18	5.80	1.01
	D2	6.13 ²	0.92	6.67	1.05	6.47	0.92
IS–RFD (<i>n</i> = 14)	D1	5.64	0.93	5.86	0.95	5.86	1.03
	D2	6.14	0.86	6.14	0.86	6.36	0.84
Control (<i>n</i> = 17)	D1	5.53	0.87	5.41	0.80	5.41	0.62
	D2	5.71	0.77	5.65	0.70	5.65	0.70

Figure 4.1 reflects the mean scores presented in Table 4.1. As the graph shows, the five groups differed in terms of between-draft improvement. In Task 2, the ID+RFD, the IS+RFD, and the IS–RFD groups made relatively bigger increase. In Task 3, the ID+RFD and the IS+RFD groups showed the biggest increase. In Task 4, compared with the control group, the treatment groups made relatively bigger increase in a similar degree.

² $Z_{skewness} = 2.42$, $Z_{kurtosis} = 3.14$

Figure 4.1 Mean content scores for Drafts 1 and 2 over Tasks 2–4 by group condition

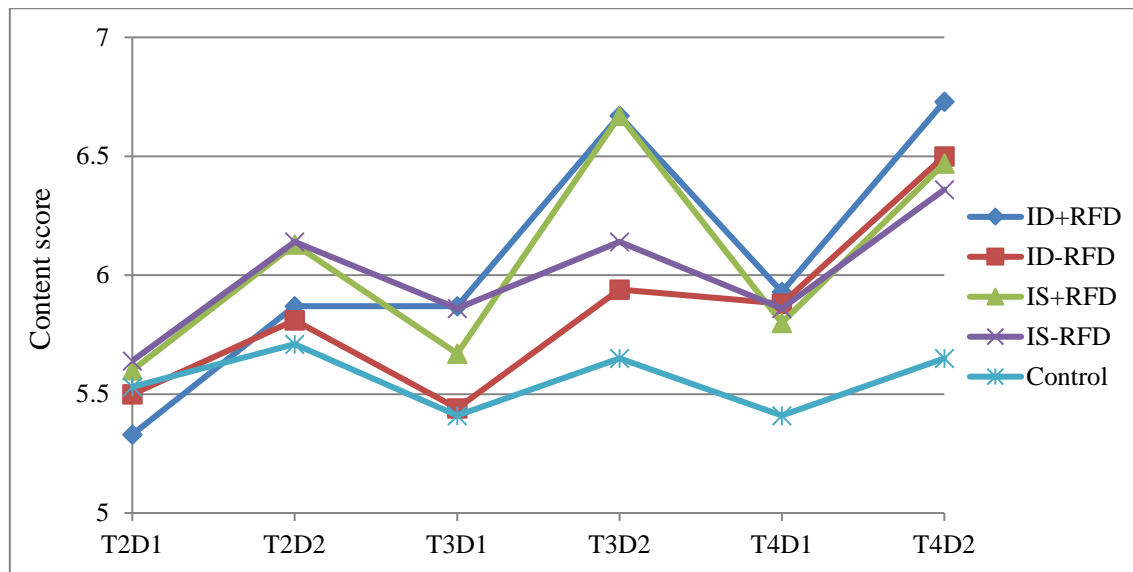


Table 4.2 Effects of draft and group condition on content quality in Tasks 2–4

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	44.70***	.000	.383	57.96***	.000	.446	73.40***	.000	.505
group	.36	.839	.019	1.70	.160	.086	2.43^	.055	.119
draft \times group	1.42	.237	.073	3.96**	.006	.180	2.21	.077	.109

Note. The *df* was 1 for draft, and the *df* was 4 for group condition as well as the two-way interaction. par. η^2 = partial η^2 .

*** $p < .001$

** $p < .01$

^ $p < .06$

Two-way repeated-measures ANOVAs were run to calculate the effect of draft and condition on content quality in the three tasks of S2, and the results are shown in Table 4.2³. As explained earlier, accompanying the ANOVAs pairwise comparisons with

³ Box's tests of equality of covariance matrices and Levene's tests of equal variances showed no violation of the assumptions for the three ANOVAs.

Bonferroni adjustment were also computed by editing SPSS syntax to further explore within- and between-subjects differences. It should be noted that the results are not included in Table 4.2 and the other tables that display ANOVA results. Instead, the results from pairwise comparisons are represented in text.

In Task 2 there was a main effect for draft, $F(1, 72) = 44.70, p = .000$, partial $\eta^2 = .383$, but neither group condition nor the two-way interaction was significant⁴. The nonsignificant interaction suggested that overall the five groups performed similarly across drafts. Pairwise comparisons revealed that the five groups produced in general similar content quality in Draft 1 as well as in Draft 2. However, they exhibited differences in terms of between-draft quality change. Significant improvement occurred in the four treatment groups but not in the control group ($p = .181$). A closer observation showed that the ID+RFD condition ($p = .000, d = 0.59$), the IS+RFD condition ($p = .000, d = 0.52$), and the IS–RFD condition ($p = .000, d = 0.56$) had medium effect sizes, and the ID–RFD condition ($p = .023, d = 0.28$) had a small effect size⁵.

In Task 3 there was again a significant effect of draft, $F(1, 72) = 57.96, p = .000$, partial $\eta^2 = .446$, but no significant effect of group condition. However, the interaction effect between draft and group condition was significant, $F(4, 72) = 3.96, p = .006$, partial $\eta^2 = .180$, suggesting that the five groups differed greatly in their content quality across drafts. Pairwise comparisons indicated that in Draft 1 the five groups' content scores were similar but in Draft 2 the ID+RFD group ($p = .045, d = 1.16$) and the IS+RFD group ($p = .045, d = 1.16$) significantly outperformed the control group with large effect sizes. The other two treatment groups did not show such an advantage over the control group. In addition, the ID+RFD group ($p = .000, d = 0.73$), the ID–RFD group ($p = .003, d = 0.43$), and the IS+RFD group ($p = .000, d = 0.90$) made a significant between-draft improvement with different effect sizes. The IS–RFD group ($p = .104, d = 0.32$) and the control group ($p = .139, d = 0.32$) did not make significant improvement.

⁴ The alpha value was set at .05 for the report of results of this study, except that for the Box's test, which was set at .001.

⁵ According to Larson-Hall (2010), the effect size for post hoc contrasts after ANOVA tests can be calculated by Cohen's d . The formula adopted in this study was Cohen's $d = (x_1 - x_2) / \{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2] / (n_1 + n_2 - 2)\}^{1/2}$. The outcome can be classified as small ($d = 0.2$), medium ($d = 0.5$), or large ($d = 0.8$).

In Task 4 draft had a significant effect, $F(1, 72) = 57.96, p = .000$, partial $\eta^2 = .446$, and the effect of group condition was marginally significant, $F(4, 72) = 2.43, p = .055$, partial $\eta^2 = .119$. The interaction effect between draft and group condition was considerable though not significant, $F(4, 72) = 2.21, p = .077$, partial $\eta^2 = .109$. Pairwise comparisons showed that the five groups' content scores did not show significant differences in Draft 1 but did in Draft 2. Specifically, the ID+RFD group ($p = .003, d = 1.55$) and the ID-RFD group ($p = .036, d = 1.07$) significantly outperformed the control group. The IS+RFD group ($p = .059, d = 1.01$) almost significantly outperformed the control group. In terms of a between-draft increase, all the treatment groups produced significant improvement, but the control group did not ($p = .098, d = 0.36$). A closer observation found that the ID+RFD group ($p = .000, d = 1.06$), the ID-RFD group ($p = .000, d = 0.68$), the IS+RFD group ($p = .000, d = 0.68$), and the IS-RFD group ($p = .002, d = 0.53$) made improvement with different effect sizes.

4.2.2 The immediate effect of WF type and \pm RFD on content quality across drafts

Table 4.3 shows the descriptive statistics on treatment groups' content scores in S2 by draft, WF type, and \pm RFD. It shows that participants with either of the WF types and under either of the \pm RFD conditions made some improvement from Draft 1 to Draft 2 in each of the three tasks.

Table 4.3 Means and standard deviations for content quality in Tasks 2–4 by draft, WF type, and \pm RFD

	Draft	Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID (<i>n</i> = 31)	D1	5.42	1.12	5.65	1.14	5.90	0.87
	D2	5.84	0.86	6.29	1.16	6.61	0.80
IS (<i>n</i> = 29)	D1	5.62	1.02	5.76	1.06	5.83	1.00
	D2	6.14	0.88	6.41	0.98	6.41 ⁶	0.87
+RFD (<i>n</i> = 30)	D1	5.47	1.11	5.77	1.14	5.87	0.90
	D2	6.00	0.79	6.67	1.03	6.60 ⁷	0.81
–RFD (<i>n</i> = 30)	D1	5.57	1.04	5.63	1.07	5.87	0.97
	D2	5.97	0.96	6.03	1.03	6.43	0.86

The figures below illustrate the mean content scores over S2 by WF type and by \pm RFD condition, respectively. Figure 4.2 indicates that students receiving either type of feedback improved content from Draft 1 to Draft 2 on a similar magnitude in each of the tasks. Figure 4.3 illustrates that the +RFD and the –RFD students improved from Draft 1 to Draft 2 in a similar degree in Task 2 and Task 4; however, the +RFD students made a greater improvement than the –RFD students in Task 3.

⁶ $Z_{\text{skewness}} = 2.69$, $Z_{\text{kurtosis}} = 2.23$

⁷ $Z_{\text{kurtosis}} = 2.03$

Figure 4.2 Mean content scores for Drafts 1 and 2 over Tasks 2–4 by WF type

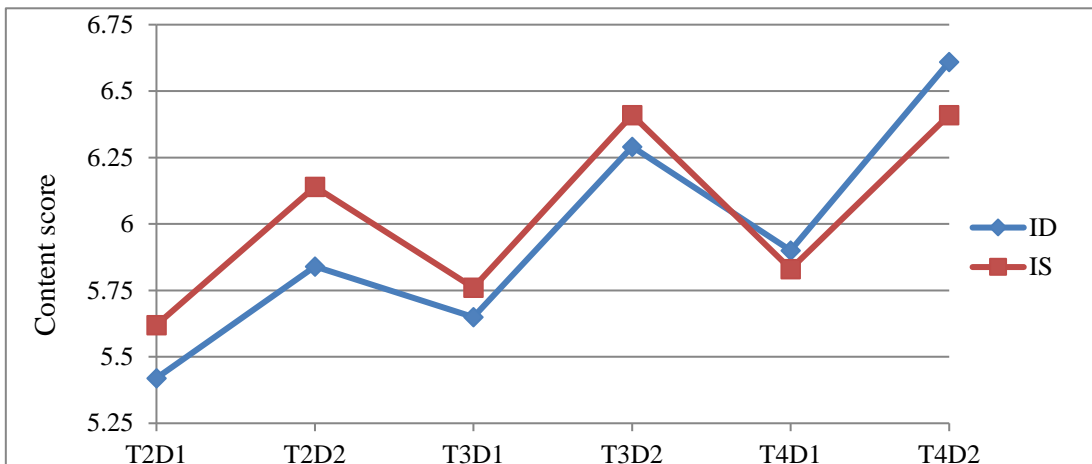


Figure 4.3 Mean content scores for Drafts 1 and 2 over Tasks 2–4 by ±RFD

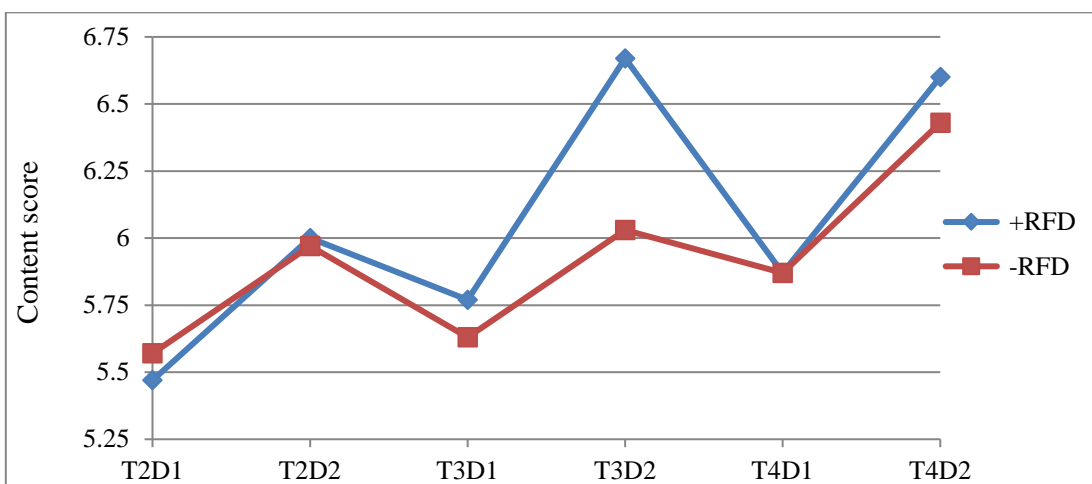


Table 4.4 displays the main effect of draft, WF type and \pm RFD on content quality in the three tasks of S2, which were obtained from three-way repeated-measures ANOVAs⁸. It is found that draft had a significant effect with large and ascending effect sizes over the tasks, and that the interaction between draft and \pm RFD was statistically significant in Task 3.

⁸ Box's tests and Levene's tests showed no violation of the assumptions for the three ANOVAs used.

Table 4.4 Effects of draft, WF type, and \pm RFD on content quality in Tasks 2–4

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	40.20***	.000	.418	55.47***	.000	.498	74.14***	.000	.570
WFT	1.06	.309	.018	.16	.690	.003	.41	.527	.007
\pm RFD	.03	.867	.001	1.99	.164	.034	.15	.697	.003
draft \times WFT	.40	.530	.007	.00	.967	.000	.74	.394	.013
draft \times \pm RFD	.74	.395	.013	8.54**	.005	.132	1.23	.261	.022
WFT \times \pm RFD	.00	.951	.000	.61	.439	.011	.07	.787	.001
draft \times WFT \times \pm RFD	.40	.530	.007	1.42	.238	.025	.00	.978	.000

Note. WFT = WF type. The *df* was 1 for all the *F* values. par. η^2 = partial η^2 .

*** $p < .001$

** $p < .01$

In Task 2 only draft had a significant effect, $F(1, 56) = 40.20$, $p = .000$, partial $\eta^2 = .418$. Neither the effect of WF type nor that of \pm RFD was significant, and none of the two-way or three-way interaction effect was significant. This overall outcome suggested that the treatment groups altogether improved significantly across drafts, but their improvement did not vary according to WF type or \pm RFD. Students with ID ($p = .000$, $d = 0.42$) and students with IS ($p = .000$, $d = 0.54$) both significantly improved with similar effect sizes. The +RFD students ($p = .000$, $d = 0.55$) and the –RFD students ($p = .000$, $d = 0.41$) also improved significantly with similar effect sizes. There were no significant between-subjects differences concerning either WF types or \pm RFD conditions.

In Task 3 there was a significant effect of draft, $F(1, 56) = 55.47$, $p = .000$, partial $\eta^2 = .498$, but there was no significant main effect of WF type or \pm RFD. The interaction

between draft and \pm RFD was significant, $F(1, 56) = 8.54, p = .005$, partial $\eta^2 = .132$, indicating that the difference caused by the factor of draft varied across \pm RFD conditions. There were no other significant interaction effects. Students with ID ($p = .000, d = 0.57$) and student with IS ($p = .000, d = 0.63$) both improved significantly with a medium effect size. There were no between-subjects differences concerning WF types. On the other hand, while both the +RFD students ($p = .000, d = 0.83$) and the –RFD students ($p = .002, d = 0.37$) made a significant between-draft improvement, the former’s improvement was on a larger magnitude. In addition, the \pm RFD students did not differ in content quality of Draft 1, but the +RFD students significantly outperformed the –RFD students in Draft 2 ($p = .024, d = 0.61$). These comparisons suggested the advantage of the +RFD condition irrespective of WF type.

In Task 4 only draft had a significant effect, $F(1, 56) = 74.14, p = .000$, partial $\eta^2 = .570$. Neither the effect of WF type nor that of \pm RFD was significant, and none of the two-way or three-way interactions was significant. Pairwise comparisons revealed that students with ID ($p = .000, d = 0.85$) as well as students with IS ($p = .000, d = 0.62$) significantly improved content quality. Similarly, both the +RFD students ($p = .000, d = 0.86$) and the –RFD students ($p = .000, d = 0.61$) significantly improved. No between-subjects differences were observed concerning either WF types or \pm RFD conditions.

4.2.3 Summary of results for content quality

(1) Statistically significant contrasts concerning the five group conditions

Table 4.5 displays the significant within-subjects contrasts of the five group conditions in S2. The ID+RFD condition made significant improvement in all the three tasks, and the effect size grew steadily from a medium magnitude ($d = 0.59$) to a large one ($d = 1.06$). The IS+RFD group also made significant improvement in all the three tasks. The effect size grew from a medium magnitude ($d = 0.52$) to a large one ($d = 0.90$) but dropped a little bit ($d = 0.68$) in Task 4. The ID–RFD group produced significant between-draft increase in each of the three tasks, and the effect size increased gradually from a small one ($d = 0.28$) to a medium to large one ($d = 0.68$). The IS–RFD group significantly improved content in Task 2 ($d = 0.56$) and Task 4 ($d = 0.53$) with medium effect sizes. Finally, the control group did not produce significant improvement in any of the three tasks.

Table 4.5 Significant within-subjects contrasts of the group conditions regarding content quality across drafts in Tasks 2–4

ID+RFD		ID–RFD		IS+RFD		IS–RFD		Control
Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task
***T2	0.59	*T2	0.28	***T2	0.52	***T2	0.56	
***T3	0.73	**T3	0.43	***T3	0.90			
***T4	1.06	***T4	0.68	***T4	0.68	**T4	0.53	

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

Table 4.6 summarises the significant between-subjects contrasts across the five group conditions. It basically supports the information in Table 4.5. The ID+RFD group performed similarly to the control group in Draft 2 of Task 2; it significantly outperformed the control group in Draft 2 of Task 3 ($d = 1.16$) and Task 4 ($d = 1.55$), and the relevant effect sizes were large. The IS+RFD group performed similarly to the ID+RFD group with respect to the comparison with the control group, except that in Draft 2 of Task 4 it outperformed the control group almost significantly ($p = .059$, $d = 1.01$). The ID–RFD group only significantly outperformed the control group in Draft 2 of Task 4 with a large effect size ($d = 1.07$). The IS–RFD group showed no significant advantage over the control group.

Table 4.6 Significant between-subjects contrasts across the group conditions regarding content quality in Tasks 2–4

	Task 2	Task 3	<i>d</i>	Task 4	<i>d</i>
Draft 1					
Draft 2		*ID+RFD > CG	1.16	*ID+RFD > CG	1.55
		*IS+RFD > CG	1.16	*ID–RFD > CG	1.07
				^IS+RFD > CG	1.01

Note. The blank cells indicate the absence of applicable results. CG = control group

* $p < .05$

^ $p < .06$

In summary, treatment conditions were more effective than the control condition in helping students to produce better content in a revised text. Among the treatment conditions, the ID+RFD and the IS+RFD conditions showed the biggest advantage, followed by the ID–RFD condition and the IS–RFD condition in order.

(2) Statistically significant contrasts concerning the factors of WF type and ±RFD

Table 4.7 presents the significant within-subjects contrasts concerning WF types and ±RFD conditions. Both feedback types brought significant improvement in all the three tasks. The effect size of the ID feedback increased from a small to medium magnitude ($d = 0.42$) to a large one ($d = 0.85$), while the effect size of IS remained at the medium magnitude. Both ±RFD conditions brought significant improvement in all the three tasks. The effect size was medium ($d = 0.55$) to large ($d = 0.86$) for the +RFD condition while small ($d = 0.37$) to medium ($d = 0.61$) for the –RFD condition.

Table 4.7 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding content quality in Tasks 2–4

ID		IS		+RFD		-RFD	
Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>
***T2	0.42	***T2	0.54	***T2	0.55	***T2	0.41
***T3	0.57	***T3	0.63	***T3	0.83	**T3	0.37
***T4	0.85	***T4	0.62	***T4	0.86	***T4	0.61

*** $p < .001$

** $p < .01$

With respect to the significant contrasts across WF types and \pm RFD conditions regarding content quality in each draft of the three tasks, only in Draft 2 of Task 3 did the +RFD students significantly outperform the -RFD group with a medium effect size ($p = .024$, $d = 0.61$).

In summary, the factor of WF type did not play a role in content quality improvement between drafts although the ID feedback type exhibited an increasing effect size. On the other hand, the \pm RFD factor did show some significant effect. The +RFD condition seemed to lead to larger content quality improvement.

4.3 Results for Organisation Quality

4.3.1 The immediate effect of group conditions on organisation quality across drafts

Table 4.8 provides the means and standard deviations for organisation quality in S2 by draft and group condition. The scores of all the five groups showed some increase on the 9-point scale from Draft 1 to Draft 2 in each of the three tasks.

Table 4.8 Mean and standard deviations for organisation quality in Tasks 2–4 by draft and group condition

Group	Draft	Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	D1	5.27	1.03	5.93	1.10	6.13	0.74
	D2	6.27	0.80	7.00	1.07	7.00	0.76
ID–RFD (<i>n</i> = 16)	D1	5.63	0.96	5.63	1.36	6.06	0.85
	D2	6.13	1.03	6.31	1.08	6.75	0.93
IS+RFD (<i>n</i> = 15)	D1	5.20	1.37	5.87	0.99	5.93	0.88
	D2	6.27 ⁹	1.10	6.80 ¹⁰	0.86	6.60	0.99
IS–RFD (<i>n</i> = 14)	D1	5.79	1.25	5.93	1.27	6.14	0.86
	D2	6.14	0.95	6.36	0.93	6.71	0.83
Control (<i>n</i> = 17)	D1	5.59	0.87	5.41	0.87	5.47	0.80
	D2	5.82	1.02	5.76 ¹¹	0.83	5.71	0.85

Figure 4.4 depicts the five groups' mean scores recorded in Table 4.8. The graph shows that the five groups performed differently in terms of between-draft improvement over Tasks 2–4. In Task 2 and Task 3, the ID+RFD and IS+RFD group exhibited the biggest increase in the mean score. In Task 4, all the four treatment groups improved a little more than the control group.

⁹ $Z_{\text{kurtosis}} = 2.40$

¹⁰ $Z_{\text{skewness}} = 2.08$

¹¹ $Z_{\text{skewness}} = 2.25$

Figure 4.4 Mean organisation scores for Drafts 1 and 2 over Tasks 2–4 by group condition

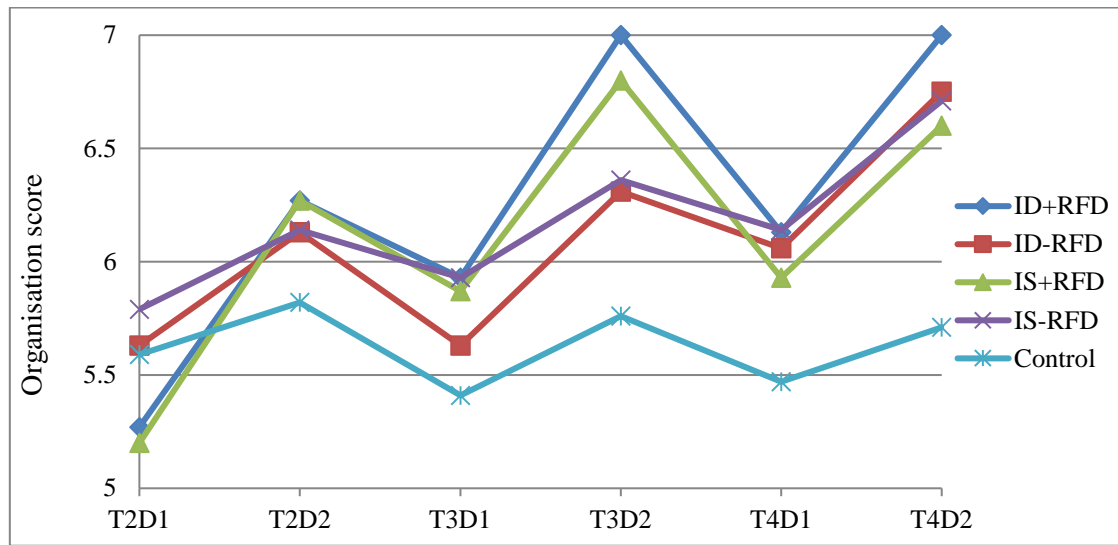


Table 4.9 lists the main effects of draft and group condition on organisation quality in S2¹². Draft had a consistently large effect over S2, but the effect of group condition and that of the interaction between draft and group condition varied across tasks.

Table 4.9 Effects of draft and group condition on organisation quality in Tasks 2–4

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	62.59***	.000	.465	86.79***	.000	.547	69.81***	.000	.492
group	.18	.948	.010	1.92	.117	.096	3.91**	.006	.178
draft \times group	4.54**	.003	.201	3.47*	.012	.162	2.19	.079	.108

Note. The *df* was 1 for draft, and the *df* was 4 for group condition as well as the two-way interaction. par. η^2 = partial η^2 .

*** $p < .001$

** $p < .01$

* $p < .05$

¹² Box's tests and Levene's tests showed no violation of the assumptions for the ANOVAs involved.

In Task 2 there was a significant main effect of draft, $F(1, 72) = 62.59, p = .000$, partial $\eta^2 = .465$, but no significant effect for group condition. The interaction effect between draft and group condition was significant, $F(4, 72) = 4.54, p = .003$, partial $\eta^2 = .201$, suggesting that the between-draft change varied across the groups. Pairwise comparisons revealed that the five groups produced in general similar organisation quality in Draft 1 as well as in Draft 2. However, they exhibited differences in terms of between-draft improvement. The control group did not improve significantly ($p = .170$). Significant improvement occurred in the ID+RFD group ($p = .000, d = 1.08$), the ID-RFD group ($p = .006, d = 0.50$), and the IS+RFD group ($p = .000, d = 0.86$). The improvement of the IS-RFD group approached significance ($p = .060, d = 0.32$).

In Task 3 there was again a significant effect of draft, $F(1, 72) = 86.79, p = .000$, partial $\eta^2 = .547$, and no significant effect of group condition. The interaction between draft and group condition was significant, $F(4, 72) = 3.47, p = .012$, partial $\eta^2 = .162$, suggesting that the five groups differed greatly in their between-draft organisation quality change. Pairwise comparisons revealed that for Draft 1 the five groups' scores were not significantly different, but for Draft 2 the ID+RFD group ($p = .005, d = 1.30$) and the IS+RFD group ($p = .032, d = 1.23$) significantly outperformed the control group. In terms of within-subjects differences, all the five groups produced significant improvement. A closer observation found that the ID+RFD condition ($p = .000, d = 0.98$) and the IS+RFD condition ($p = .000, d = 1.01$) resulted in relatively larger effect sizes than the ID-RFD condition ($p = .000, d = 0.56$), the IS-RFD condition ($p = .016, d = 0.39$), and the control condition ($p = .029, d = 0.37$).

In Task 4 draft had a significant effect, $F(1, 72) = 69.81, p = .000$, partial $\eta^2 = .492$, and so did group condition, $F(4, 72) = 3.91, p = .006$, partial $\eta^2 = .178$. The two-way interaction effect was considerable though not significant, $F(4, 72) = 2.19, p = .079$, partial $\eta^2 = .108$. Pairwise comparisons showed that the five groups' organisation scores of Draft 1 were still not significantly different, but for Draft 2 the ID+RFD group ($p = .001, d = 1.60$), the ID-RFD group ($p = .010, d = 1.18$), and the IS-RFD group ($p = .021, d = 1.20$) significantly outperformed the control group. The IS+RFD group ($p = .051, d = 0.97$) almost significantly outperformed the control group. In terms of between-draft improvement, all the treatment groups produced significant improvement, but the control group did not ($p = .131$). A closer observation found that the ID+RFD

condition ($p = .000$, $d = 1.16$), the ID–RFD condition ($p = .000$, $d = 0.77$), the IS+RFD condition ($p = .000$, $d = 0.71$), and the IS–RFD condition ($p = .001$, $d = 0.68$) generated different effect sizes.

4.3.2 The immediate effect of WF type and \pm RFD on organisation quality across drafts

Table 4.10 records the descriptive results for the treatment groups’ organisation scores by the factors of WF type and \pm RFD. Figures 4.5 and 4.6 reflect the mean scores presented in this table.

Table 4.10 Means and standard deviations for organisation quality in Tasks 2–4 by draft, WF type, and \pm RFD

		Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID ($n = 31$)	D1	5.45	1.00	5.77	1.23	6.10	0.79
	D2	6.19	0.91	6.65	1.11	6.87	0.85
IS ($n = 29$)	D1	5.48	1.33	5.90	1.11	6.03	0.87
	D2	6.21	1.01	6.59	0.91	6.66	0.90
+RFD ($n = 30$)	D1	5.23	1.19	5.90	1.03	6.03	0.81
	D2	6.27 ¹³	0.94	6.90	0.96	6.80	0.89
–RFD ($n = 30$)	D1	5.70	1.09	5.77	1.31	6.10	0.85
	D2	6.13	0.97	6.33	0.99	6.73	0.87

Figure 4.5 indicates that students receiving either ID or IS improved organisation from Draft 1 to Draft 2 on a similar magnitude in any of the three tasks. Figure 4.6 shows that the +RFD students produced a bigger improvement than the –RFD students in Task 2 and Task 3, but their between-draft improvement was on a similar magnitude in Task 4.

¹³ $z_{\text{kurtosis}} = 2.48$

Figure 4.5 Mean organisation scores for Drafts 1 and 2 over Tasks 2–4 by WF type

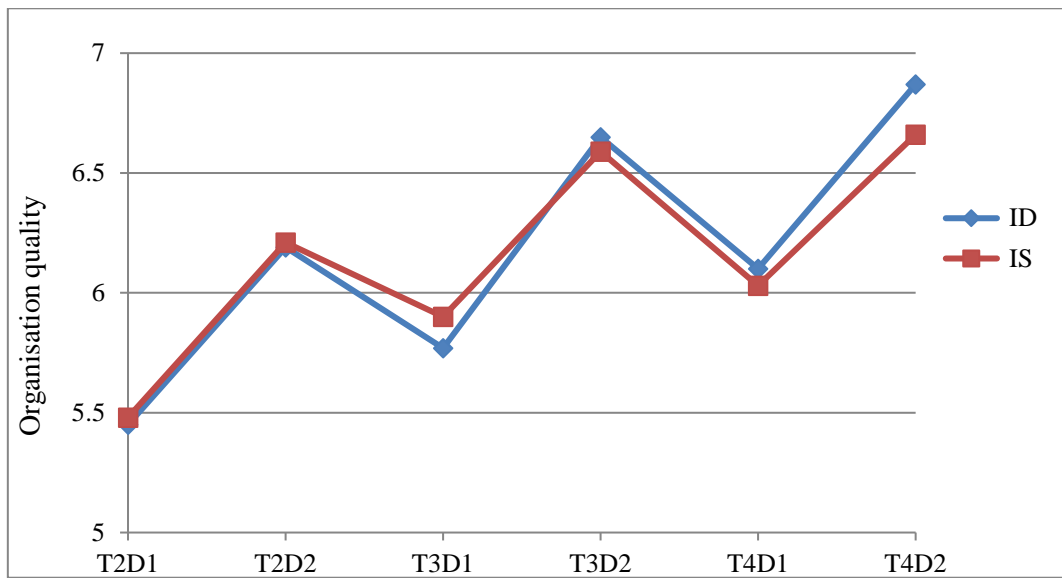


Figure 4.6 Mean organisation scores for Drafts 1 and 2 over Tasks 2–4 by ±RFD

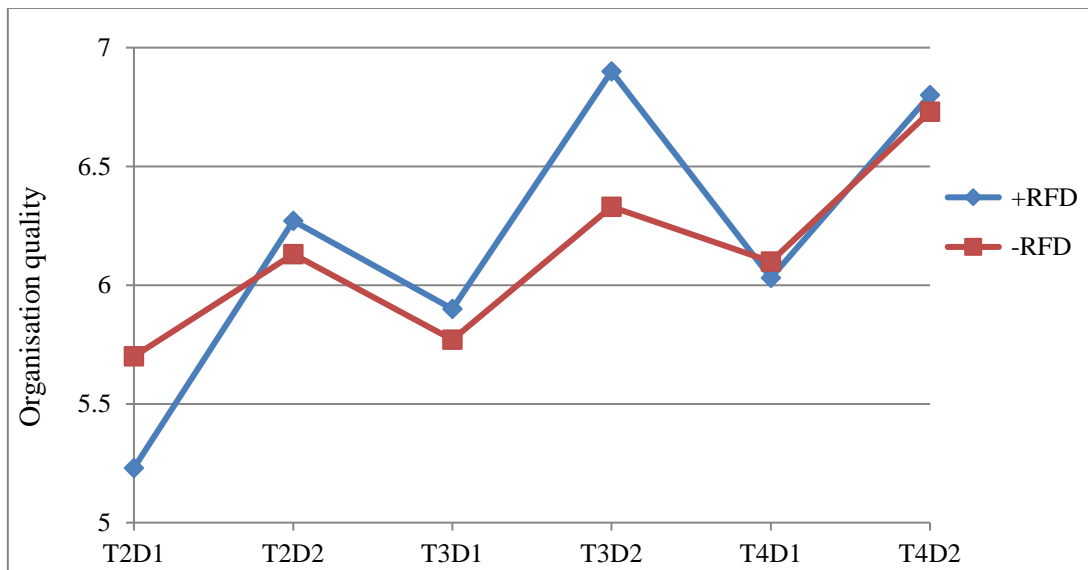


Table 4.11 summarises the main effect of draft, WF type, and \pm RFD on organisation quality in the three tasks of S2¹⁴. With respect to statistical significance, draft had a

¹⁴ Box's tests and Levene's tests showed no violation of the assumptions for the three ANOVAs used.

consistent significant effect over the tasks, and the interaction effect between draft and \pm RFD was significant in Task 2 and Task 3.

Table 4.11 Effects of draft, WF type, and \pm RFD on organisation quality in Tasks 2–4

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	63.64***	.000	.532	89.50***	.000	.615	68.26***	.000	.549
WFT	.01	.915	.000	.01	.940	.000	.46	.502	.008
\pm RFD	.42	.519	.007	1.61	.209	.028	.00	.997	.000
draft \times WFT	.04	.836	.001	1.42	.239	.025	.88	.354	.015
draft \times \pm RFD	10.89**	.002	.163	7.20**	.010 ¹⁵	.114	.66	.420	.012
WFT \times \pm RFD	.06	.816	.001	.32	.573	.006	.62	.436	.011
draft \times WFT \times \pm RFD	.33	.570	.006	.15	.704	.003	.06	.805	.001

Note. WFT = WF type. The *df* was 1 for all the *F* values. par. η^2 = partial η^2 .

*** $p < .001$

** $p < .01$

In Task 2 draft had a significant effect, $F(1, 56) = 63.64$, $p = .000$, partial $\eta^2 = .532$, and the interaction effect between draft and \pm RFD was significant, $F(1, 56) = 10.89$, $p = .002$, partial $\eta^2 = .163$, suggesting that the difference caused by the factor of draft differed significantly across \pm RFD conditions. Pairwise comparisons found that students with either ID ($p = .000$, $d = 0.78$) or IS ($p = .000$, $d = 0.60$) produced a significant improvement with similar effect sizes. However, the +RFD students ($p = .000$, $d = 0.96$) made improvement with a large effect size whereas the –RFD students ($p = .002$, $d = 0.42$) improved with a small to medium effect size. No significant between-subjects

¹⁵ $p = .0096$ if rounded to four decimals

differences were observed concerning either WF types or \pm RFD conditions for either Draft 1 or Draft 2.

In Task 3 there was again a significant effect of draft, $F(1, 56) = 89.50, p = .000$, partial $\eta^2 = .615$. There was a significant interaction between draft and \pm RFD, $F(1, 56) = 7.20, p = .010$, partial $\eta^2 = .114$. As in Task 2, both students receiving ID ($p = .000, d = 0.75$) and students receiving IS ($p = .000, d = 0.67$) made a significant improvement, and their organisation scores did not differ significantly in either Draft 1 or Draft 2. In terms of \pm RFD, both the +RFD students ($p = .000, d = 1.00$) and the -RFD students ($p = .000, d = 0.48$) made a significant improvement, but the effect size of the former was much larger. The \pm RFD students' organisation quality did not differ significantly for Draft 1 but did for Draft 2 ($p = .032, d = 0.58$).

In Task 4 only draft had a significant effect, $F(1, 56) = 68.26, p = .000$, partial $\eta^2 = .549$. Pairwise comparisons showed that students receiving either ID ($p = .000, d = 0.95$) or IS ($p = .000, d = 0.70$) made a significant improvement. The +RFD students ($p = .000, d = 0.90$) and the -RFD students ($p = .000, d = 0.73$) both improved significantly. No significant between-subjects difference was found for either Draft 1 or Draft 2 with respect to feedback types and \pm RFD conditions.

4.3.3 Summary of results for organisation quality

(1) Statistically significant contrasts concerning the five group conditions

Table 4.12 displays the significant within-subjects contrasts of the five group conditions in S2. Both ID+RFD and IS+RFD groups produced significant improvement in all the three tasks with overall large effect sizes. The ID-RFD group produced significant between-draft increase in each of the tasks, and the effect size grew gradually from a medium magnitude ($d = 0.50$) to a marginally large one ($d = 0.77$). The IS-RFD group significantly improved in Task 3 ($d = 0.39$) and Task 4 ($d = 0.68$), and the effect size was basically medium. Finally, the control group made significant improvement only in Task 3 with a small to medium effect size ($d = 0.37$).

Table 4.12 Significant within-subjects contrasts of the group conditions regarding organisation quality across drafts in Tasks 2–4

ID+RFD		ID–RFD		IS+RFD		IS–RFD		Control	
Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>
***T2	1.08	**T2	0.50	***T2	0.86	^T2	0.32		
***T3	0.98	***T3	0.56	***T3	1.01	*T3	0.39	*T3	0.37
***T4	1.16	***T4	0.77	***T4	0.71	**T4	0.68		

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

Table 4.13 summarises the significant between-subjects contrasts across the five group conditions. It corroborates the information in Table 4.12. The ID+RFD group significantly outperformed the control group in Draft 2 of Task 3 ($d = 1.30$) and Task 4 ($d = 1.60$), and the relevant effect sizes were large. The IS+RFD group performed similarly to the ID+RFD group with respect to the comparison with the control group; the difference was that for Draft 2 of Task 4 it outperformed the control group almost significantly, and the effect size was smaller ($p = .051$, $d = 0.97$). The ID–RFD and the IS–RFD groups only significantly outperformed the control group for Draft 2 of Task 4, but the effect sizes were large ($d = 1.18$ for ID–RFD and 1.20 for IS–RFD).

Table 4.13 Significant between-subjects contrasts across the group conditions regarding organisation quality in Tasks 2–4

	Task 2	Task 3	<i>d</i>	Task 4	<i>d</i>
Draft 1					
Draft 2		**ID+RFD > CG	1.30	**ID+RFD > CG	1.60
		*IS+RFD > CG	1.23	*ID–RFD > CG	1.18
				^IS+RFD > CG	0.97
				*IS–RFD > CG	1.20

Note. The blank cells indicate the absence of applicable results. CG = control group

** $p < .01$

* $p < .05$

^ $p < .06$

Overall, of the five group conditions the ID+RFD and the IS+RFD conditions had the biggest advantage in terms of the effect on organisation quality improvement between drafts, followed by the ID–RFD condition, the IS–RFD condition, and the control condition in order.

(2) Statistically significant contrasts concerning the factors of WF type and \pm RFD

Table 4.14 presents the significant within-subjects contrasts concerning WF types and \pm RFD conditions. Both WF methods brought significant improvement in all the three tasks. The effect size of the ID type was basically large ($d = 0.78, 0.75,$ and 0.95), while the effect size of the IS type was medium to large ($d = 0.60, 0.67,$ and 0.70). Both \pm RFD students made significant improvement in all the three tasks, but on average the +RFD condition generated an evidently larger effect size ($d = 0.96, 1.00,$ and 0.90).

Table 4.14 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding organisation quality in Tasks 2–4

ID		IS		+RFD		-RFD	
Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>
***T2	0.78	***T2	0.60	***T2	0.96	**T2	0.42
***T3	0.75	***T3	0.67	***T3	1.00	***T3	0.48
***T4	0.95	***T4	0.70	***T4	0.90	***T4	0.73

*** $p < .001$

** $p < .01$

With respect to the significant between-subjects contrasts across WF types and \pm RFD conditions regarding organisation quality for each draft of the three tasks, only one case of such significant contrasts was found; for Draft 2 of Task 3 the +RFD students significantly outperformed the -RFD students with a medium effect size ($p = .032$, $d = 0.58$).

In summary, the factor of WF type did not show an effect on organisation quality improvement between drafts although the ID type exhibited a slightly larger effect size. On the other hand, the \pm RFD factor did play a role, with the +RFD condition having an advantage over the -RFD condition.

4.4 Results for Grammatical Accuracy

4.4.1 The immediate effect of group conditions on grammatical accuracy across drafts

Table 4.15 displays the means and standard deviations for grammatical accuracy in S2 by draft and group condition. It shows that all the five groups made some reduction in the number of their grammatical errors (per 100 words) from Draft 1 to Draft 2 in each of the three tasks.

Table 4.15 Means and standard deviations for grammatical accuracy^a in Tasks 2–4 by draft and group condition

Group	Draft	Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	D1	5.90 ¹⁶	2.43	5.60	2.33	5.29	1.59
	D2	4.31	2.23	4.20	1.28	4.59	1.94
ID–RFD (<i>n</i> = 16)	D1	5.80	2.05	5.67	2.61	5.91	1.98
	D2	3.57	1.68	3.74	0.99	4.39	1.87
IS+RFD (<i>n</i> = 15)	D1	5.39	2.48	5.44	2.50	5.60	2.11
	D2	4.36	2.19	4.57	2.16	4.56	2.21
IS–RFD (<i>n</i> = 14)	D1	5.55	1.83	5.64	2.43	5.71	1.93
	D2	3.68	1.46	3.26	1.67	3.78	1.41
Control (<i>n</i> = 17)	D1	5.94	1.51	6.12	1.82	5.99	2.15
	D2	5.10	1.30	5.29	1.77	5.26	1.76

^a Number of grammatical errors per 100 words.

Figure 4.7 depicts the five groups' mean scores recorded in Table 4.15. It can be seen from the graph that the five groups differed in between-draft improvement over Tasks 2–4. In all the three tasks, the ID–RFD and the IS–RFD groups made a relatively bigger decrease in errors, while the IS+RFD and the control groups made a smaller reduction.

¹⁶ $z_{\text{skewness}} = 2.15$, $z_{\text{kurtosis}} = 2.66$

Figure 4.7 Mean grammatical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by group condition

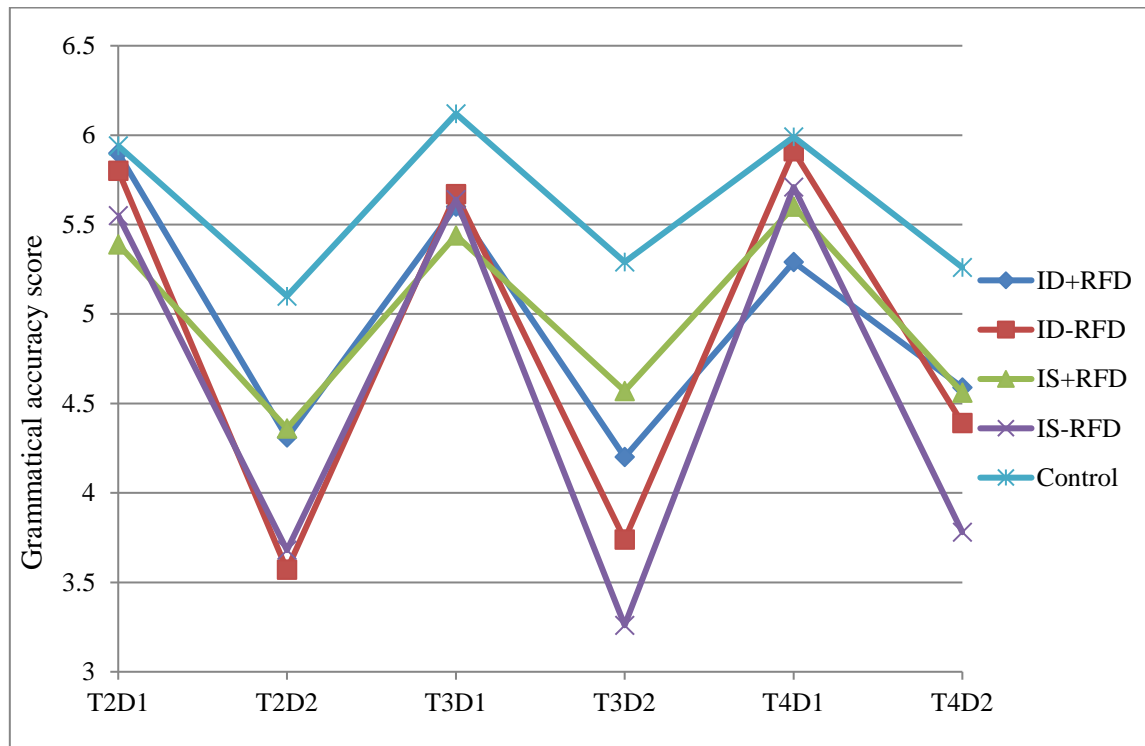


Table 4.16 Effects of draft and group condition on grammatical accuracy in S2

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	41.41***	.000	.365	47.72***	.000	.399	48.93***	.000	.405
group	.79	.536	.042	1.11	.358	.058	.55	.699	.030
draft \times group	1.26	.293	.065	1.95	.111	.098	1.91	.117	.096

Note. The *df* for draft was 1, and the *df* was 4 for group condition as well as the two-way interaction. par. η^2 = partial η^2 .

*** $p < .001$

Table 4.16 presents the effect of draft and group condition on grammatical accuracy in S2¹⁷. A glimpse finds that only draft rendered a significant effect. In Task 2 draft produced a significant effect, $F(1, 72) = 41.41, p = .000$, partial $\eta^2 = .365$, but neither group condition nor the two-way interaction had a significant effect. Pairwise comparisons revealed that the five groups wrote with generally similar grammatical accuracy in Draft 1 as well as in Draft 2. In terms of within-subjects differences, the control group did not improve significantly over drafts ($p = .099$). Significant improvement occurred in the ID+RFD group ($p = .004, d = 0.68$), the ID–RFD group ($p = .000, d = 1.19$), and the IS–RFD group ($p = .001, d = 1.13$). The improvement of the IS+RFD group approached statistical significance ($p = .055, d = 0.44$). The effect sizes of the ID–RFD and the IS–RFD conditions were large.

In Task 3 again there was only a significant effect of draft, $F(1, 72) = 48.93, p = .000$, partial $\eta^2 = .405$. Pairwise comparisons showed that the five groups had no significant differences in grammatical accuracy for Draft 1, but for Draft 2, the IS–RFD group significantly outperformed the control group ($p = .009, d = 1.18$). In terms of within-subjects differences, significant improvement occurred in the ID+RFD group ($p = .005, d = 0.75$), the ID–RFD group ($p = .000, d = 0.98$), and the IS–RFD group ($p = .000, d = 1.14$). The improvement of the IS+RFD group ($p = .079, d = 0.37$) and the control group ($p = .072, d = 0.46$) was considerable but not statistically significant. The effect sizes of the ID–RFD and the IS–RFD conditions were large.

In Task 4 only the effect of draft was significant, $F(1, 72) = 48.93, p = .000$, partial $\eta^2 = .405$. No significant between-subjects differences were found for either Draft 1 or Draft 2. In terms of within-subjects differences, significant improvement occurred in the ID–RFD group ($p = .000, d = 0.79$), the IS+RFD group ($p = .009, d = 0.48$), the IS–RFD group ($p = .000, d = 1.14$), as well as the control group ($p = .045, d = 0.37$). The improvement of the ID+RFD group ($p = .069, d = 0.40$) was considerable but not

¹⁷ Box's tests showed no violation of the assumption. Levene's tests showed that the scores of Draft 2 of Task 3 violated homogeneity of variance ($p = .018$). A further check found that the violation was somewhat serious because among the five groups' *SDs* the largest *SD* (2.16) was twice bigger than the smallest one (.99) (see Table 4.15). According to Lowie and Seton (2013), if the *SD* of a group is less than twice as big as that of another group, the variance is approximately the same. Given that only one of the Levene's tests showed violation, the output of the ANOVAs was adopted.

statistical. As in Task 2 and Task 3, the effect sizes of the ID–RFD and the IS–RFD conditions were large.

4.4.2 The immediate effect of WF type and \pm RFD on grammatical accuracy across drafts

Table 4.17 displays the descriptive data on treatment groups’ grammatical accuracy in Tasks 2–4 by draft, WF type, and \pm RFD. It shows that the participants with either WF type and under either of \pm RFD conditions made some improvement in accuracy from Draft 1 to Draft 2 in each of the three tasks.

Table 4.17 Means and standard deviations for grammatical accuracy^a in Tasks 2–4 by draft, WF type, and \pm RFD

	Draft	Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID (<i>n</i> = 31)	D1	5.84	2.21	5.64 ¹⁹	2.44	5.61	1.80
	D2	3.93 ¹⁸	1.97	3.96	1.14	4.48	1.87
IS (<i>n</i> = 29)	D1	5.47	2.16	5.54	2.42	5.65	1.99
	D2	4.03	1.87	3.94	2.02	4.18	1.87
+RFD (<i>n</i> = 30)	D1	5.64 ²⁰	2.43	5.52	2.38	5.44	1.84
	D2	4.33	2.17	4.39	1.76	4.58	2.04
–RFD (<i>n</i> = 30)	D1	5.68	1.92	5.66	2.48	5.82	1.93
	D2	3.62	1.56	3.52	1.35	4.10	1.67

^aNumber of grammatical errors per 100 words.

The figures below illustrate the mean grammatical accuracy scores over Tasks 2–4 by WF type and by \pm RFD, respectively. Figure 4.8 illustrates that in all the three tasks students receiving ID and students receiving IS reduced their errors from Draft 1 to Draft

¹⁸ $z_{\text{skewness}} = 2.42$

¹⁹ $z_{\text{skewness}} = 2.04$

²⁰ $z_{\text{skewness}} = 2.29$, $z_{\text{kurtosis}} = 2.21$

2 on a similar magnitude. Figure 4.9 shows that the -RFD students had a greater reduction of errors across drafts than the +RFD students throughout the three tasks.

Figure 4.8 Mean grammatical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by WF type

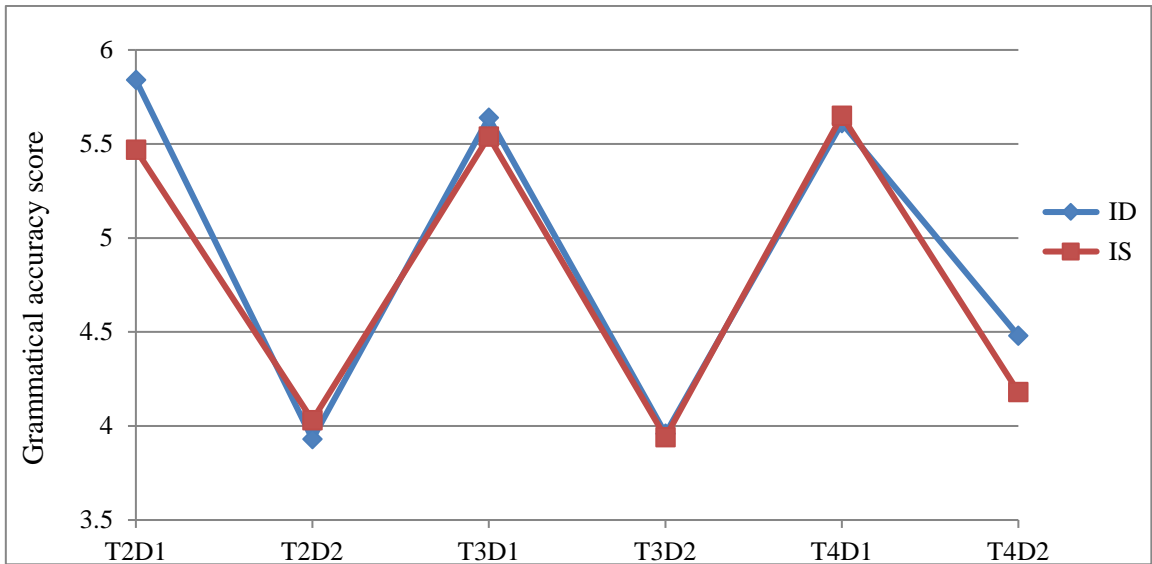


Figure 4.9 Mean grammatical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by ±RFD

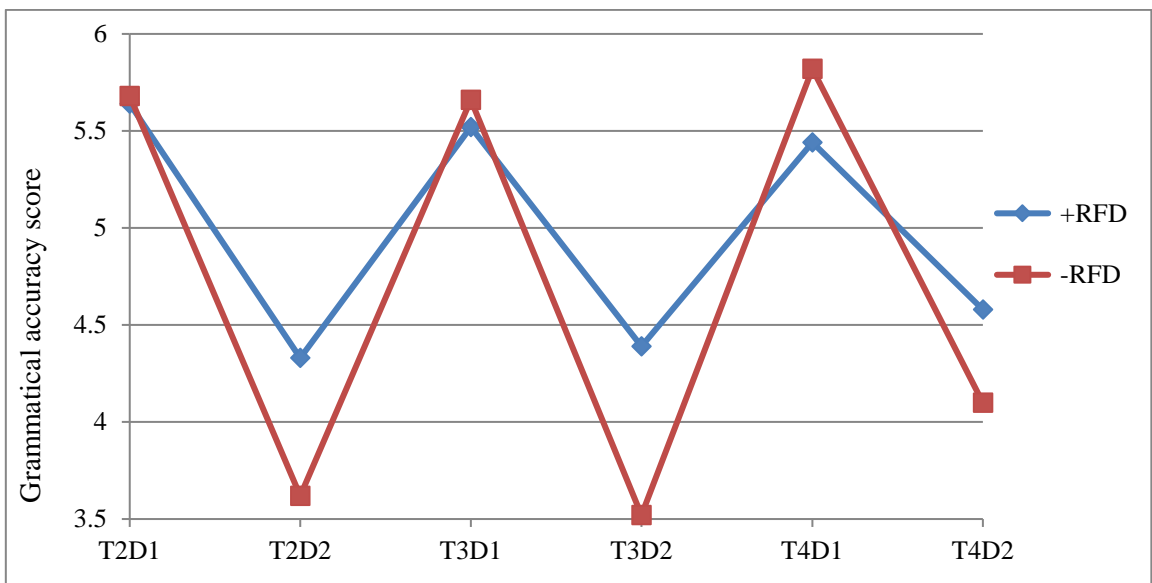


Table 4.18 summarises the main effect of draft, WF type, and \pm RFD on grammatical accuracy in the three tasks of S2²¹. A brief look finds that draft had a significant effect with a consistently large effect size over the tasks, and the interaction between draft and \pm RFD was significant in Task 4 and marginally significant in Task 3.

Table 4.18 Effects of draft, WF type, and \pm RFD on grammatical accuracy in Tasks 2–4

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	32.58***	.000	.368	41.41***	.000	.425	44.96***	.000	.445
WFT	.01	.744	.002	.03	.872	.000	.09	.769	.002
\pm RFD	.57	.454	.010	.63	.432	.011	.02	.886	.000
draft \times WFT	.59	.445	.010	.01	.935	.000	.89	.351	.016
draft \times \pm RFD	1.57	.216	.027	4.00 [^]	.050	.067	4.91*	.031	.081
WFT \times \pm RFD	.03	.859	.001	.15	.703	.003	.37	.548	.006
draft \times WFT \times \pm RFD	.03	.862	.001	.95	.335	.017	.012	.913	.000

Note. WFT = WF type. The *df* was 1 for all the *F* values. par. η^2 = partial η^2 .

*** $p < .001$

* $p < .05$

[^] $p < .06$ ($p = .0501$ if rounded to 4 decimals)

²¹ Box's tests showed no violation of the assumption. Levene's tests indicated that the scores of Draft 2 of Task 3 violated homogeneity of variance ($p = .010$). A further check found that the violation was somewhat serious because among the four groups' *SDs* in question (see Table 4.17) the largest *SD* (2.44) was twice bigger than the smallest one (1.14). Given that only one of the Levene's tests showed violation, the output of the ANOVAs was still adopted for reporting results (see Footnote 17 of Chapter Four).

In Task 2 only the factor of draft had a significant effect, $F(1, 56) = 32.58, p = .000$, partial $\eta^2 = .368$. Pairwise comparisons revealed that students receiving either ID ($p = .000, d = 0.91$) or IS ($p = .001, d = 0.72$) produced a significant improvement with similar effect sizes. On the other hand, the +RFD students ($p = .003, d = 0.57$) made a significant increase with a medium effect size whereas the -RFD group ($p = .000, d = 1.17$) made improvement with a large effect size. There was no significant between-subjects difference in either Draft 1 or Draft 2 with respect to feedback type or \pm RFD.

In Task 3 there was a significant effect of draft, $F(1, 56) = 41.41, p = .000$, partial $\eta^2 = .425$. The interaction effect between draft and \pm RFD approached statistical significance, $F(1, 56) = 4.00, p = .050$, partial $\eta^2 = .067$, which suggested that the between-draft difference differed across \pm RFD conditions. Pairwise comparisons showed that both ID ($p = .000, d = 0.87$) and IS ($p = .000, d = 0.73$) rendered a significant accuracy increase with similar effect sizes. Students with ID and students with IS did not differ significantly in grammatical accuracy for either draft. On the other hand, both the +RFD students ($p = .003, d = 0.54$) and the -RFD students ($p = .000, d = 1.08$) made a significant improvement, but the effect size of the -RFD condition was larger. In addition, the -RFD students significantly outperformed the +RFD students for Draft 2 ($p = .034, d = 0.56$).

In Task 4 there was a significant effect of draft, $F(1, 56) = 44.96, p = .000$, partial $\eta^2 = .445$, and a significant interaction between draft and \pm RFD, $F(1, 56) = 4.91, p = .031$, partial $\eta^2 = .081$, which generally repeated the result of Task 3. Again both ID ($p = .000, d = 0.61$) and IS ($p = .000, d = 0.77$) rendered a significant increase. On the other hand, both the +RFD students ($p = .002, d = 0.45$) and the -RFD students ($p = .000, d = 0.96$) made a significant increase, but the effect size of the -RFD condition was evidently larger. Moreover, no significant between-subjects difference was observed with respect to WF types and \pm RFD conditions for either draft.

4.4.3 Summary of results for grammatical accuracy

(1) Statistically significant contrasts concerning the five group conditions

Table 4.19 displays the significant within-subjects contrasts of the five group conditions in S2. Both the IS–RFD and ID–RFD conditions rendered significant grammatical accuracy increase in all the three tasks, and their respective effect sizes were consistently large (the smallest one being 0.79). The ID+RFD condition produced significant between-draft increases in Task 2 and Task 3, and the corresponding effect sizes were both medium to large ($d = 0.68$ and 0.75). Participants in the IS+RFD condition and the control condition significantly improved only in Task 4 with small to medium effect sizes.

Table 4.19 Significant within-subjects contrasts of the group conditions regarding grammatical accuracy across drafts in Tasks 2–4

ID+RFD		ID–RFD		IS+RFD		IS–RFD		Control	
Task	d	Task	d	Task	d	Task	d	Task	d
T2	0.68	*T2	1.19	^T2	0.44	**T2	1.13		
T3	0.75	*T3	0.98			***T3	1.14		
		***T4	0.79	**T4	0.48	***T4	1.14	*T4	0.37

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

^ $p < .06$

With respect to the significant between-subjects contrasts across the five group conditions, significant contrast only occurred for Draft 2 of Task 3, when the IS–RFD group outperformed the control group with a large effect size ($p = .009$, $d = 1.18$).

In summary, of the five conditions the IS–RFD condition had the biggest advantage in terms of the effect on grammatical accuracy improvement between drafts, followed by the ID–RFD condition, the ID+RFD condition, the IS+RFD condition, and the control condition in order.

(2) Statistically significant contrasts concerning the factors of WF type and \pm RFD

Table 4.20 presents the significant within-subjects contrasts concerning WF types and \pm RFD conditions. Both WF methods rendered significant accuracy increase in all the three tasks. On average, the effect sizes of the two feedback types were similar. Both +RFD students and –RFD students made significant improvement in all the three tasks. The +RFD condition generated in general medium effect sizes ($d = 0.57, 0.54,$ and 0.45), while the –RFD condition had consistently large effect sizes ($d = 1.17, 1.08,$ and 0.96).

Table 4.20 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding grammatical accuracy in Tasks 2–4

ID		IS		+RFD		–RFD	
Task	d	Task	d	Task	d	Task	d
***T2	0.91	**T2	0.72	**T2	0.57	***T2	1.17
***T3	0.87	***T3	0.73	**T3	0.54	***T3	1.08
***T4	0.61	***T4	0.77	**T4	0.45	***T4	0.96

*** $p < .001$

** $p < .01$

With respect to between-subjects differences across WF types and \pm RFD conditions regarding grammatical accuracy in each draft of the three tasks, only in Draft 2 of Task 3 occurred a significant between-subjects contrast; the –RFD students significantly outperformed the +RFD students with a medium effect size ($p = .034, d = 0.56$).

In summary, the factor of WF type did not show an effect on grammatical accuracy improvement between drafts in S2. On the other hand, the \pm RFD factor played a role; the –RFD condition had an advantage over the +RFD condition.

4.5 Results for Lexical Accuracy

4.5.1 The immediate effect of group conditions on lexical accuracy across drafts

Table 4.21 displays the means and standard deviations for lexical accuracy in S2 by draft and group condition. It shows that all the five groups made some reduction in their lexical errors (per 100 words) from Draft 1 to Draft 2 in each of the three tasks.

Table 4.21 Means and standard deviations for lexical accuracy^a in Tasks 2–4 by draft and group condition

Group	Draft	Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	D1	2.77	1.67	2.91	1.22	3.08	1.07
	D2	2.43	1.34	2.40	1.32	2.75	1.36
ID–RFD (<i>n</i> = 16)	D1	2.81	1.29	2.87	1.62	2.93	1.59
	D2	2.35	1.54	2.38	1.54	2.37	1.33
IS+RFD (<i>n</i> = 15)	D1	2.88	1.20	2.72	1.08	2.96	1.76
	D2	2.17	1.01	2.46	1.28	2.69	1.43
IS–RFD (<i>n</i> = 14)	D1	2.75	1.24	2.93	1.13	2.90	1.45
	D2	2.22	1.32	2.42	0.78	2.19	0.94
Control (<i>n</i> = 17)	D1	2.78	0.84	3.01	1.06	2.98	1.05
	D2	2.57	0.87	2.74	0.86	2.77	1.03

^aNumber of lexical errors per 100 words.

Figure 4.10 depicts the five groups' mean scores recorded Table 4.21. The graph reveals that the five groups showed some differences in between-draft improvement over Tasks 2–4. In Task 2, the IS+RFD group made the sharpest decrease in lexical errors (per 100 words). In Task 3, the ID+RFD, the ID–RFD, and the IS–RFD groups made relatively sharper reduction of errors. In Task 4, the IS–RFD and the ID–RFD groups made the largest decrease.

Figure 4.10 Mean lexical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by group condition

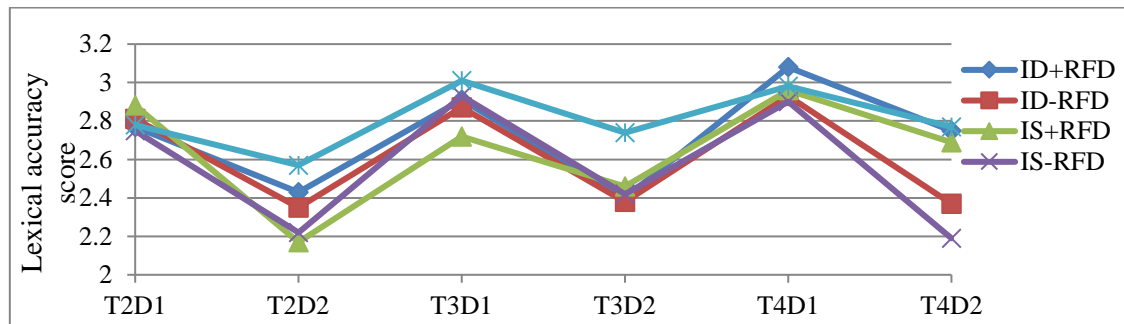


Table 4.22 lists the main effect of draft and group condition on lexical accuracy in S2²². In Task 2 draft had a significant effect on lexical accuracy, $F(1, 72) = 12.91, p = .001$, partial $\eta^2 = .152$, but neither the effect of group condition nor the interaction effect between draft and group was significant. Pairwise comparisons revealed that the five groups wrote with generally similar lexical accuracy for both Draft 1 and Draft 2. In terms of within-subjects differences, significant improvement only occurred in the IS+RFD group ($p = .014, d = 0.64$) and the improvement of the IS–RFD group approached significance ($p = .076, d = 0.41$).

In Task 3 again there was only a significant effect of draft on lexical accuracy, $F(1, 72) = 9.31, p = .003$, partial $\eta^2 = .115$. Pairwise comparisons showed no significant differences of the five groups for either draft, and no significant between-draft improvement occurred in any of the five groups.

In Task 4 only draft produced a significant effect on lexical accuracy, $F(1, 72) = 19.77, p = .000$, partial $\eta^2 = .215$. Pairwise comparisons still showed no significant differences in lexical accuracy of the five groups for either draft. In terms of within-subjects differences, significant improvement occurred in the ID–RFD group ($p = .008, d = 0.38$) and the IS–RFD group ($p = .002, d = 0.58$).

²² Box’s tests showed no violation of the assumption. Levene’s tests found a violation of equal variances in Draft 2 of Task 3 ($p = .009$). The violation was acceptable, though, because the largest *SD* (1.54) was not twice as big as the smallest one (.78).

Table 4.22 Effects of draft and treatment condition on lexical accuracy in Tasks 2–4

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	12.91**	.001	.152	9.31**	.003	.115	19.77***	.000	.215
group	.07	.992	.004	.18	.948	.010	.24	.915	.013
draft \times group	.46	.762	.025	.20	.938	.011	.99	.416	.052

Note. The *df* for draft was 1, and the *df* was 4 for group condition as well as the two-way interaction. par. η^2 = partial η^2 .

*** $p < .001$

** $p < .01$

4.5.2 The immediate effect of WF type and \pm RFD on lexical accuracy across drafts

Table 4.23 lists the descriptive results for the treatment groups' lexical accuracy. It shows that participants with either WF type and under either of the \pm RFD conditions produced less lexical errors (per 100 words) from Draft 1 to Draft 2 in each of the three tasks.

Table 4.23 Means and standard deviations for lexical accuracy^a in Tasks 2–4 by draft, WF type, and \pm RFD

	Draft	Task 2		Task 3		Task 4	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ID (<i>n</i> = 31)	D1	2.79	1.46	2.88	1.42	3.01	1.34
	D2	2.39	1.43	2.39	1.41	2.56	1.33
IS (<i>n</i> = 29)	D1	2.82	1.20	2.82	1.09	2.93	1.59
	D2	2.19	1.15	2.44	1.05	2.45	1.23
+RFD (<i>n</i> = 30)	D1	2.83	1.43	2.81	1.14	3.02	1.43
	D2	2.30	1.17	2.43	1.27	2.72	1.37
-RFD (<i>n</i> = 30)	D1	2.78	1.25	2.90	1.39	2.92	1.50
	D2	2.29	1.42	2.40	1.22	2.29	1.15

^aNumber of lexical errors per 100 words.

The following two figures demonstrate the mean accuracy scores over Tasks 2–4 by WF type and by \pm RFD, respectively. Figure 4.11 shows that in general students with different feedback types reduced errors on a similar magnitude. Figure 4.12 illustrates that the \pm RFD students decreased the error ratio in a similar degree in Task 2 and Task 3, but that in Task 4 the -RFD students made a bigger reduction in error ratio.

Figure 4.11 Mean lexical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by WF type

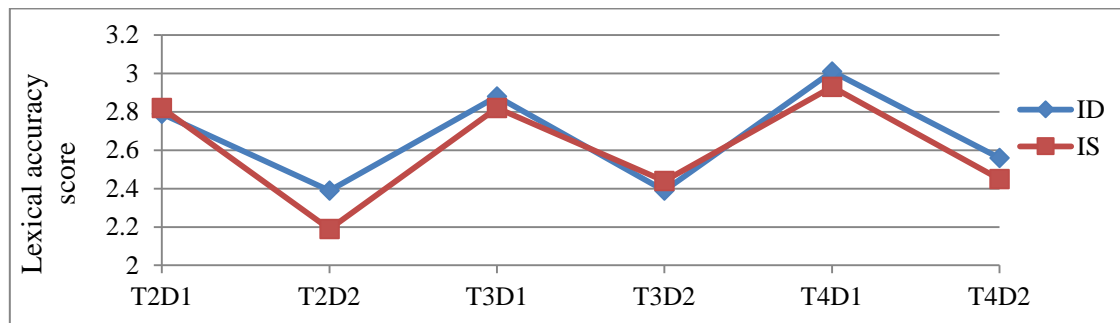
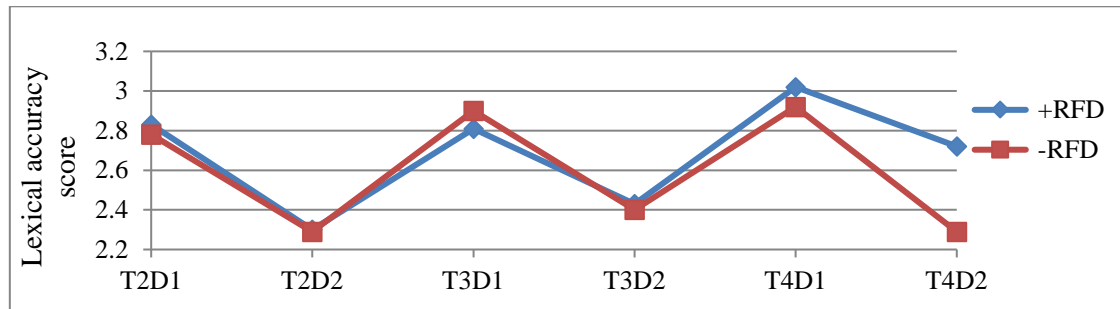


Figure 4.12 Mean lexical accuracy scores for Drafts 1 and 2 over Tasks 2–4 by \pm RFD



A group of ANOVAs were run to assess the effect of draft, WF type, and \pm RFD on lexical accuracy in S2²³ (see Table 4.24). A brief observation of the output found that only the factor of draft produced significant effect. In Task 2 draft had a significant effect, $F(1, 56) = 11.49, p = .001, \text{partial } \eta^2 = .170$. Pairwise comparisons revealed that students receiving ID made a considerable accuracy increase but not significantly ($p = .062, d = 0.28$), while students receiving IS ($p = .006, d = 0.53$) produced a significant improvement. On the other hand, both the +RFD students ($p = .016, d = 0.40$) and the –RFD students ($p = .025, d = 0.37$) made a significant improvement with similar effect sizes. There was no significant between-subjects difference in either Draft 1 or Draft 2 with respect to feedback type or \pm RFD.

In Task 3 again draft had a significant effect, $F(1, 56) = 8.00, p = .006, \text{partial } \eta^2 = .125$. Students receiving ID improved significantly ($p = .025, d = 0.35$) with a small effect size, while students receiving IS ($p = .092, d = 0.36$) did not produce a significant improvement. The +RFD students ($p = .089, d = 0.31$) did not improve significantly, but the –RFD students did ($p = .027, d = 0.38$). Again, no significant between-subjects difference was found for either draft with respect to feedback type or \pm RFD.

In Task 4 only draft rendered a significant effect, $F(1, 56) = 19.60, p = .000, \text{partial } \eta^2 = .259$. Students receiving either ID ($p = .004, d = 0.33$) or IS ($p = .002, d = 0.34$) improved significantly with a small effect size. The +RFD students ($p = .049, d = 0.21$) improved significantly with a small effect size, and the –RFD students ($p = .000, d = 0.47$) improved significantly with a medium effect size. In addition, there was no

²³ Box's tests and Levene's tests showed no violation of the assumptions.

significant between-subjects difference for either draft with respect to feedback type or \pm RFD.

Table 4.24 Effects of draft, WF type, and \pm RFD on lexical accuracy in Tasks 2–4

	Task 2			Task 3			Task 4		
	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2	<i>F</i>	<i>p</i>	par. η^2
draft	11.49*	.001	.170	8.00*	.006	.125	19.60***	.000	.259
WFT	.08	.785	.001	.00	.984	.000	.09	.768	.002
\pm RFD	.01	.923	.000	.01	.919	.000	.64	.429	.011
draft \times WFT	.55	.461	.010	.14	.715	.002	.05	.831	.001
draft \times \pm RFD	.02	.900	.000	.15	.704	.003	2.49	.120	.043
WFT \times \pm RFD	.00	.975	.000	.04	.850	.001	.00	.978	.000
draft \times WFT \times \pm RFD	.24	.627	.004	.21	.647	.004	.23	.637	.004

Note. WFT = WF type. The *df* was 1 for all the *F* values. par. η^2 = partial η^2 .

*** $p < .001$

* $p < .05$

4.5.3 Summary of results for lexical accuracy

(1) Statistically significant contrasts concerning the five group conditions

Only three groups made significant between-draft improvement, and their significant improvement only occurred in one of the three tasks. The IS+RFD group made a significant increase in Task 2 with a medium effect size ($p = .014$, $d = 0.64$); the ID–RFD group ($p = .008$, $d = 0.38$) and the IS–RFD group improved significantly ($p = .002$, $d = 0.58$) in Task 4 with similar effect sizes.

Unlike the results for the other three textual levels (i.e., content, organisation, and grammatical accuracy), no significant between-subjects contrasts were found across the five group conditions for lexical accuracy in either draft of any tasks in S2.

In summary, none of the five conditions showed a clear advantage in the immediate effect on lexical accuracy improvement between drafts. The IS–RFD, the IS+RFD, and the ID–RFD conditions showed similarly slight effectiveness.

(2) Statistically significant contrasts concerning the factors of WF type and ±RFD

Table 4.25 presents the significant within-subjects contrasts concerning WF types and ±RFD conditions. Students receiving either type of feedback produced significant accuracy increase in two of the three tasks, and the corresponding effect sizes were small or medium ($d = 0.33$ and 0.35 for ID; $d = 0.34$ and 0.53 for IS). The +RFD students made significant increase in two of the three tasks with generally small effect sizes ($d = 0.21$ and 0.40), while the –RFD students made significant increase in all the three tasks with small to medium effect sizes ($d = 0.37$, 0.38 , and 0.47).

Table 4.25 Significant within-subjects contrasts concerning WF types and ±RFD conditions regarding lexical accuracy across drafts in Tasks 2–4

ID		IS		+RFD		–RFD	
Task	d	Task	d	Task	d	Task	d
		**T2	0.53	*T2	0.40	*T2	0.37
*T3	0.35					*T3	0.38
T4	0.33	**T4	0.34	*T4	0.21	*T4	0.47

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

No significant between-subjects contrasts were observed across the two WF types or across ±RFD conditions for lexical accuracy in either draft of any task in S2.

In summary, neither the factor of WF type nor \pm RFD showed a significant effect on lexical accuracy improvement between drafts in S2. The ID and the IS feedback methods had overall similar effect. The $-$ RFD condition exhibited a slight advantage over the $+$ RFD condition.

4.6 Summary of Results in Relation to the Immediate Effects of Group Conditions on Text Quality across Drafts

This section summarises the effect of group conditions on revision in terms of text quality across drafts, so it includes the results in relation to all the four textual levels. Table 4.26 and Table 4.27 respectively display the within-subjects and the between-subjects significant contrasts concerning the four text levels in Tasks 2–4 of the five conditions. The following two subsections successively compare the effects of the treatment conditions with that of the control condition and compare the effects of the treatment conditions between each other and by textual level. The term “the average effect size” of a certain treatment condition is meant to refer to the average of its effect sizes for the two or three cases of significant improvement it brought during the three tasks in Stage 2.

4.6.1 The effects of treatment conditions compared with the control condition

In general each of the treatment conditions was more effective than the control condition in helping participants improve text quality in revised drafts. Table 4.26 shows that each treatment condition rendered two to three cases of significant between-draft improvements for content quality, organisation quality, and grammatical accuracy, while the control condition brought merely one case of significant improvement for organisation quality and for grammatical accuracy with small effect sizes ($d = 0.37$ and 0.37). The advantage of treatment was not obvious for lexical accuracy, however. Only the ID–RFD, the IS+RFD, and the IS–RFD conditions brought one case of significant increase in lexical accuracy. Table 4.27 also indicates that the treatment conditions were superior to the control condition. Starting from Task 3, some of the treatment groups produced better content and organisation in Draft 2 although they did not outperform the control group in Draft 1, and the corresponding effect sizes were all large. The IS–RFD group produced higher grammatical accuracy than the control group in Draft 2 of

Task 3 with a large effect size ($d = 1.18$). However, such an advantage was not evidenced in terms of lexical accuracy.

4.6.2 The mediation of textual level in the effects of group conditions

The effect of the four treatment conditions varied with textual level. As Table 4.26 and Table 4.27 show, the treatment conditions taken together were more effective for revision in organisation than for revision in content, and they were more effective for grammatical accuracy increase than lexical accuracy increase. For three of the treatment conditions, the average effect size for between-draft organisation quality improvement in S2 tasks was somewhat larger than that for between-draft content quality improvement ($d = 1.07$ vs. 0.79 for ID+RFD, 0.61 vs. 0.46 for ID-RFD, and 0.86 vs. 0.70 for IS+RFD). The IS-RFD condition brought significant improvement in content quality in two of the three tasks with an average medium effect size ($d = 0.55$); as for organisation quality, it brought significant improvement in two tasks and one marginally significant improvement in one task, the average effect size being approximately medium ($d = 0.46$).

A closer inspection of the statistically significant differences and the corresponding effect sizes in the tables reveals that the four treatment conditions differed from one another in terms of their effect for each textual level. For both content and organisation quality improvement across drafts, the ID+RFD and the IS+RFD conditions were the most effective, closely followed by ID-RFD and IS-RFD. For grammatical accuracy increase across drafts, the IS-RFD and the ID-RFD conditions were the most effective, brought significant increase in three tasks with large effect sizes (on average $d = 1.14$ for IS-RFD and 0.99 for ID-RFD). The ID+RFD and the IS+RFD conditions followed in order. For lexical accuracy increase across drafts, the ID-RFD, the IS+RFD, and the IS-RFD conditions showed similarly slight effect, while the ID+RFD and the control conditions were not effective.

It is notable there were no significant differences in text quality for Draft 1 or Draft 2 across the treatment conditions on any textual level. As indicated by the contrasts in Table 4.27, the significant between-subjects contrasts were only located between treatment conditions and the control condition.

Table 4.26 Significant within-subjects contrasts of the group conditions regarding text quality across drafts in Tasks 2–4

	ID+RFD		ID–RFD		IS+RFD		IS–RFD		Control	
	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>
Content quality	***T2	0.59	*T2	0.28	***T2	0.52	***T2	0.56		
	***T3	0.73	**T3	0.43	***T3	0.90				
	***T4	1.06	***T4	0.68	***T4	0.68	**T4	0.53		
Organisation quality	***T2	1.08	**T2	0.50	***T2	0.86	^T2	0.32		
	***T3	0.98	***T3	0.56	***T3	1.01	*T3	0.39	*T3	0.37
	***T4	1.16	***T4	0.77	***T4	0.71	**T4	0.68		
Grammatical accuracy	**T2	0.68	***T2	1.19	^T2	0.44	**T2	1.13		
	T3	0.75	*T3	0.98			***T3	1.14		
			***T4	0.79	**T4	0.48	***T4	1.14	*T4	0.37
lexical accuracy					*T2	0.64				
			**T4	0.38			**T4	0.58		

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

^ $p < .06$

Table 4.27: Significant between-subjects contrasts across the group conditions regarding text quality in Tasks 2–4

		Task 2	Task 3	<i>d</i>	Task 4	<i>d</i>
Content	Draft 1					
	Draft 2		*ID+ > CG	1.16	*ID+ > CG	1.55
			*IS+ > CG	1.16	*ID- > CG	1.07
					^IS+ > CG	1.01
Organisa- -tion	Draft 1					
	Draft 2		**ID+ > CG	1.30	**ID+ > CG	1.60
			*IS+ > CG	1.23	*ID- > CG	1.18
					^IS+ > CG	0.97
					*IS- > CG	1.20
Gramma- -tical accuracy	Draft 1					
	Draft 2		**IS- > CG	1.18		
Lexical accuracy	Draft 1					
	Draft 2					

Note. The blank cells indicate the absence of applicable results. ID+ = ID+RFD, ID- = ID-RFD, IS+ = IS+RFD, IS- = IS-RFD, CG = control group.

** $p < .01$

* $p < .05$

^ $p < .06$

4.7 Summary of Results in Relation to the Immediate Effects of the Factors of WF Type and \pm RFD on Text Quality across Drafts

As section 4.6, this section summarises the effect of WF type and \pm RFD on revision in terms of text quality across drafts by synthesising the results for all the textual levels. Table 4.28 displays the relevant within-subjects significant contrasts. Data of the control condition was excluded. As no significant interaction was found between WF type and \pm RFD, the effect of the two factors will be reported separately in the following two subsections.

4.7.1 The effect of WF type and the mediation of textual level

Overall, the factor of WF type did not produce an evident effect on between-draft text quality change for any of the four textual levels. Compared with each other, the two feedback methods were similarly effective, irrespective of textual level (see Table 4.28). For example, both ID and IS brought significant between-draft improvement in content quality for the three tasks; the average effect sizes were both medium ($d = 0.61$ for ID and 0.60 for IS).

Table 4.28 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding text quality across drafts in Tasks 2–4

	ID		IS		+RFD		-RFD	
	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>	Task	<i>d</i>
Content quality	***T2	0.42	***T2	0.54	***T2	0.55	***T2	0.41
	***T3	0.57	***T3	0.63	***T3	0.83	**T3	0.37
	***T4	0.85	***T4	0.62	***T4	0.86	***T4	0.61
Organisation quality	***T2	0.78	***T2	0.60	***T2	0.96	**T2	0.42
	***T3	0.75	***T3	0.67	***T3	1.00	***T3	0.48
	***T4	0.95	***T4	0.70	***T4	0.90	***T4	0.73
Grammatical accuracy	***T2	0.91	**T2	0.72	**T2	0.57	***T2	1.17
	***T3	0.87	***T3	0.73	**T3	0.54	***T3	1.08
	***T4	0.61	***T4	0.77	**T4	0.45	***T4	0.96
Lexical accuracy			**T2	0.53	*T2	0.40	*T2	0.37
	*T3	0.35					*T3	0.38
	T4	0.33	**T4	0.34	*T4	0.21	*T4	0.47

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

4.7.2 The effect of \pm RFD and the mediation of textual level

The factor of \pm RFD played a clearer role than the factor of WF type. Although both \pm RFD conditions brought significant between-draft improvement for each textual level, they differed between each other in their effect size according to the textual levels (for a comparison of their effect sizes, see Table 4.28). The +RFD condition was more effective than the -RFD condition for content quality improvement (on average $d = 0.75$

for +RFD and 0.46 for –RFD). The same was true for organisation quality improvement (on average $d = 0.95$ for +RFD and 0.54 for –RFD). By contrast, their relative effectiveness reversed for revision on linguistic levels, with the –RFD condition having an advantage. For example, for grammatical accuracy improvement the average effect size of +RFD was medium ($d = 0.52$) and the average effect size of –RFD was large ($d = 1.07$).

In addition, Table 4.29 shows that for Draft 2 of Task 3 the +RFD students performed significantly better than the –RFD students on discourse levels with medium effect sizes ($d = 0.61$ and 0.58) while the –RFD students produced significantly higher grammatical accuracy than the +RFD students with a medium effect size ($d = 0.56$). These data also indicated the role of the \pm RFD factor in relation to textual levels.

Table 4.29 Significant between-subjects contrasts across WF types and \pm RFD conditions regarding text quality in Tasks 2–4

	Task 2	Task 3	d	Task 4
Content	Draft 1			
	Draft 2	*+RFD > –RFD	0.61	
Organisa- -tion	Draft 1			
	Draft 2	*+RFD > –RFD	0.58	
Gramma- -tical accuracy	Draft 1			
	Draft 2	*–RFD > +RFD	0.56	
Lexical accuracy	Draft 1			
	Draft 2			

Note. The blank cells indicate the absence of applicable results.

* $p < .05$

CHAPTER FIVE

**RESULTS (2): THE EFFECT OF TREATMENT ON
AUTONOMOUS REVISION**

5.1 Overview

This chapter addresses the second set of research questions which investigated the effect of treatment on student autonomous revision. The data used to answer the questions were based on the textual analysis of the written texts participants produced in the pretreatment stage (Stage 1, i.e., Task 1) and the posttreatment stage (Stage 3, i.e., Task 5) of the study.

RQ2 Does treatment have any effect on student autonomous revision?

RQ2.1 What are the relative effects of different treatment methods on student autonomous revision?

This research question examined whether the treatment groups varied from one another and in comparison to the control group in terms of autonomous revision. Two-way repeated-measures ANOVA tests were run to calculate the effects of stage and group condition on the gain score (i.e., the difference between first-draft and second-draft scores; for an introduction to the concept, see section 3.9.4.2.2). Following each ANOVA test, pairwise comparisons were used to examine whether significant within-subjects differences (i.e., the difference in gain scores across stages by each group) and between-subjects differences (i.e., the difference in gain scores between the groups for a certain stage) existed. Such statistical analyses were performed for each textual level.

An alternative approach to statistical analysis was tried by including the data from the treatment stage (Stage 2), and its output confirmed the results reported in this chapter. Specifically, the alternative ANOVAs used stage (Stage 1, Stage 2, and Stage 3) as the within-subjects variable, group condition as the between-subjects variable, and gain score as the dependent variable. In addition, a preliminary transformation of data was completed for Stage 2 before the ANOVAs were performed. The first-draft scores of Tasks 2, 3, and 4 were averaged to represent the first-draft score of Stage 2, and similarly

a second-draft score was obtained for Stage 2. This transformation was conducted because no significant effect was found for task or its interaction with group condition on gain score.

RQ2.2 Do the factors of WF type and revision-focus direction play a role in the effects of treatment methods on student autonomous revision?

This research question investigated, on one hand, whether students receiving ID varied from students receiving IS in autonomous revision, and on the other hand, whether the +RFD students varied from the –RFD students. The analyses started with ANOVAs testing the effects of stage, WF type, and \pm RFD on gain score. Then pairwise comparisons found out the significant within-subjects differences (i.e., the difference in gain scores across stages by students receiving either type of WF, and that by students either receiving or not receiving RFD) and between-subjects differences (i.e., the difference in gain scores between students receiving ID and students receiving IS, and that between the +RFD and the –RFD students). The analyses also itemised the textual levels.

RQ2.3 Does textual level mediate the effect of treatment?

This research question investigated whether the effects of different treatment methods vary with textual level, and whether the effects of WF type and \pm RFD vary with textual level. To answer this question, the results for all the textual levels in relation to RQ2.1 and RQ2.2 were synthesised; comparisons were drawn between the treatment effects on the textual levels by observing the relevant statistically significant data and effect sizes.

From section 5.2 to section 5.5, results for the four textual levels will be successively reported. The report for each textual level will start with the outcome of statistical analyses regarding the effect of group condition (RQ2.1) and proceed to the outcome of analyses regarding the effects of WF type and \pm RFD (RQ2.2). Section 5.6 will summarise the results in relation to RQ2.1 and RQ2.2 and meanwhile report the results with respect to the mediation of textual level in treatment effect (RQ2.3).

5.2 Results for Content Quality

5.2.1 The effect of group conditions on autonomous revision: Content gain score across stages

Table 5.1 displays the descriptive statistics for the content gain score in Stage 1 (S1; i.e., Task 1) and Stage 3 (S3; i.e., Task 5) by group condition. The first- and second-draft scores in Task 1 as well as the second-draft scores in Task 5 on the four textual levels are put in Appendix E. The first-draft scores in Task 5 are reported in Chapter Six. Figure 5.1 depicts the mean gain scores of the five groups. It can be seen that the ID+RFD, the IS+RFD, and the IS–RFD groups made relatively larger content gains in S3 than in S1, while the control group made a smaller gain in S3.

Table 5.1 Means and standard deviations for the content gain score by stage and group condition

Group	Stage	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	S1	0.27 ¹	0.46
	S3	0.60	0.51
ID–RFD (<i>n</i> = 16)	S1	0.19	0.54
	S3	0.25 ²	0.45
IS+RFD (<i>n</i> = 15)	S1	0.13	0.52
	S3	0.33	0.49
IS–RFD (<i>n</i> = 14)	S1	0.21 ³	0.43
	S3	0.36 ⁴	0.63
Control (<i>n</i> = 17)	S1	0.24	0.56
	S3	0.18 ⁵	0.39

¹ $z_{\text{skewness}} = 2.03$

² $z_{\text{skewness}} = 2.27$

³ $z_{\text{skewness}} = 2.62$

⁴ $z_{\text{skewness}} = 2.83$

⁵ $z_{\text{skewness}} = 3.39$

Figure 5.1 Mean content gain scores in S1 and S3 by group condition

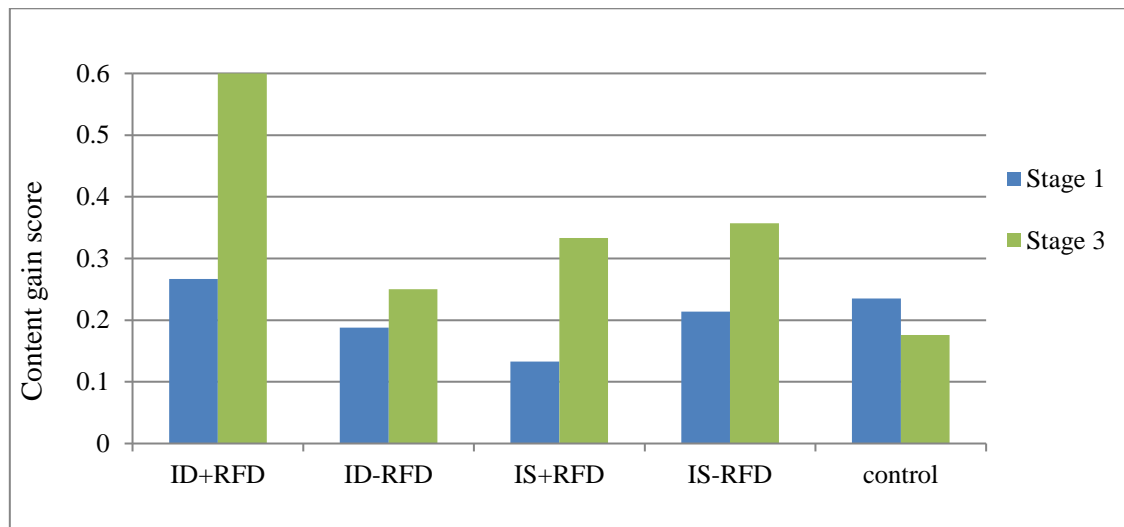


Table 5.2 Effects of stage and group condition on the content gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	2.89	.094	.039
group	1.04	.391	.055
stage \times group	.70	.594	.037

Note. The *df* was 1 for stage, and the *df* was 4 for group condition as well as the two-way interaction.

Table 5.2 shows the effect of stage and group condition on the content gain score⁶. As the practice in Chapter Four, results for pairwise comparisons, which were computed by editing SPSS syntax, are not included in the tables that present results from ANOVAs. They are reported in text.

As Table 5.2 shows, there was no main effect for stage, suggesting that the five groups as a whole did not produce different content gain scores across the two stages. There was neither a main effect for group condition, suggesting that the five groups produced

⁶ Box's tests of equality of covariance matrices and Levene's tests of equal variances showed no violation of the assumptions for the ANOVA.

in general similar gain scores if the factor of stage was ignored. The two-way interaction effect was not significant, indicating that the difference between the groups in terms of content gain score did not vary greatly according to the factor of stage.

Post hoc pairwise comparisons revealed that none of the groups made a significantly larger gain in S3 than in S1. However, the ID+RFD group's gain score in S3 was larger, though not statistically, than that in S1 ($p = .069$, $d = 0.69$). In addition, the five groups' gain scores were neither significantly different in S1 nor in S3. It was noted that for S3 the ID+RFD group differed from the control group to a large extent ($p = .182$, $d = 0.94$). These test results indicated an effect of the ID+RFD condition on student autonomous revision in terms of content quality. By contrast, the other conditions did not gender a clear effect on this aspect.

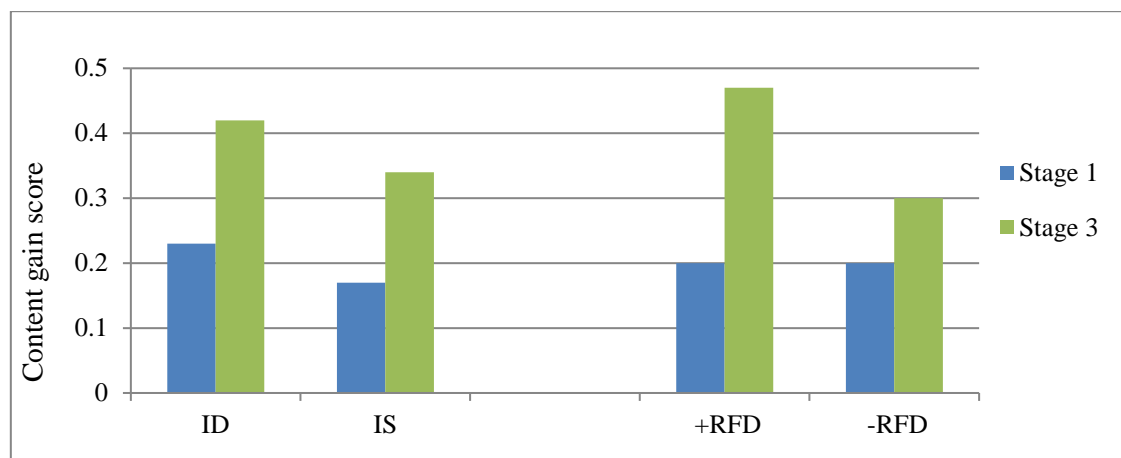
5.2.2 The effect of WF type and \pm RFD on autonomous revision: Content gain score across stages

Table 5.3 includes the descriptive data on treatment groups' content gain scores by stage, WF type, and \pm RFD. Figure 5.2 represents the mean gain scores listed in Table 5.3. The figure shows that students receiving either ID or IS produced larger content gain scores in S3 than in S1 in a similar degree. On the other hand, both the +RFD and the -RFD students made larger gain scores in S3; however, the +RFD students' mean gain score increased much more than that of the -RFD students.

Table 5.3 Means and standard deviations for the content gain score by stage, WF type, and \pm RFD

	Stage	<i>M</i>	<i>SD</i>
ID (<i>n</i> = 31)	S1	0.23	0.50
	S3	0.42 ⁷	0.50
IS (<i>n</i> = 29)	S1	0.17	0.47
	S3	0.34 ⁸	0.55
+RFD (<i>n</i> = 30)	S1	0.20	0.48
	S3	0.47 ⁹	0.51
-RFD (<i>n</i> = 30)	S1	0.20	0.48
	S3	0.30 ¹⁰	0.54

Figure 5.2 Mean content gain scores in S1 and S3 by WF type and \pm RFD



⁷ $z_{\text{kurtosis}} = 2.46$

⁸ $z_{\text{skewness}} = 3.12$

⁹ $z_{\text{kurtosis}} = 2.55$

¹⁰ $z_{\text{skewness}} = 3.80, z_{\text{kurtosis}} = 2.34$

Table 5.4 Effect of stage, WF type, and \pm RFD on the content gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	4.33*	.042	.072
WFT	.48	.491	.009
\pm RFD	.72	.401	.013
stage \times WFT	.02	.882	.000
stage \times \pm RFD	.85	.359	.015
WFT \times \pm RFD	1.94	.169	.033
stage \times WFT \times \pm RFD	.36	.549	.006

Note. WFT = WF type. The *df* was 1 for all the *F* values.

* $p < .05$

Table 5.4 lists the effect of stage, WF type, and \pm RFD on the content gain score¹¹. There was a main effect for stage, suggesting that the treatment groups as a whole produced different content gain scores across the two stages. There was no main effect for WF type or \pm RFD, suggesting that neither of the two factors played a role in content gain scores if the factor of stage was ignored. In addition, none of the two-way or three-way interactions was significant.

Post hoc pairwise comparisons revealed that neither the ID or the IS feedback type rendered an effect on content gain score across S1 and S3. However, a role of \pm RFD was found. The +RFD condition brought a significantly larger gain in S3 than in S1 ($p = .038$, $d = 0.54$), while the -RFD condition did not. This result indicated that the +RFD condition enabled the participants to produce larger improvement in content in the case of autonomous revision. A closer inspection of the results of the pairwise comparisons that were run to break down the three-way interaction revealed that the benefit of +RFD was a little more obvious when it concurred with the ID feedback. This result was in line with the advantage of the ID+RFD condition, as reported in the preceding section (5.2.1).

¹¹ Box's test and Levene's test showed no violation of the assumptions for the ANOVA.

5.2.3 Summary of results for content quality

(1) Statistically significant contrasts concerning the five group conditions

The five groups did not show significantly different gain scores in either S1 or S3. None of the five conditions made a significantly different content gain score from S1 to S3. However, participants in the ID+RFD condition made a gain in S3 that approached statistical significance, and the effect size was medium to large ($p = .069$, $d = 0.69$). This result suggested a moderately positive effect of the ID+RFD condition on student autonomous revision with respect to content quality improvement.

(2) Statistically significant contrasts concerning the factors of WF type and \pm RFD

Neither WF type rendered a significantly larger gain score in S3 compared to the gain score they rendered in S1, and the two WF types did not show a significant difference between each other in content gain scores in either S1 or S3. Likewise, the +RFD and the -RFD conditions did not differ significantly between each other in either S1 or S3. However, the +RFD students produced a significantly larger gain score in S3 than in S1 with a medium effect size ($p = .038$, $d = 0.54$), which indicated an effect of the +RFD condition on autonomous revision with regard to content improvement.

5.3 Results for Organisation Quality

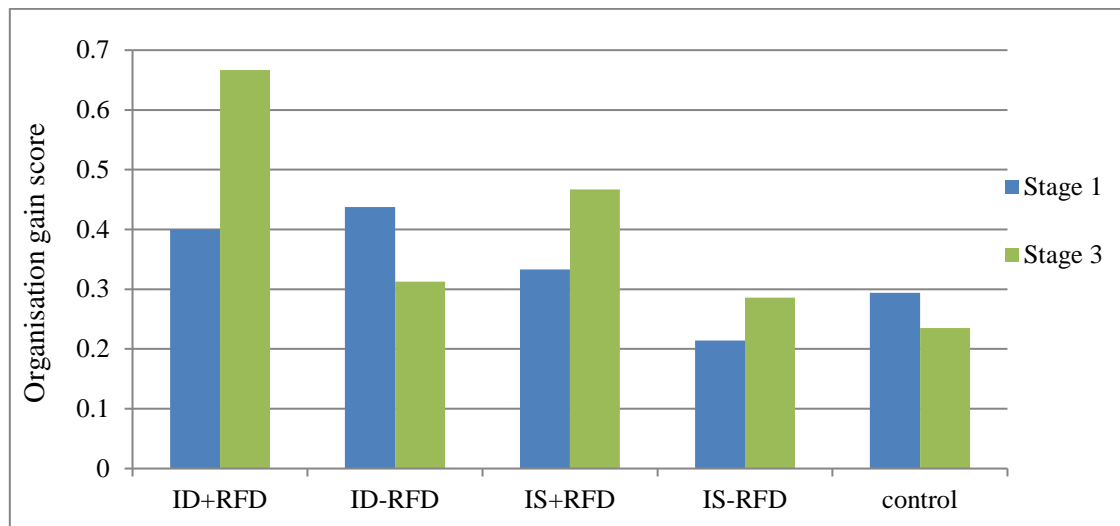
5.3.1 The effect of group conditions on autonomous revision: Organisation gain score across stages

Table 5.5 displays the descriptive results for the organisation gain score in S1 and S3 by group condition. Figure 5.3 depicts the mean gain scores of the five groups. The graph shows that the ID+RFD, the IS+RFD, and the IS-RFD groups made some increase in the gain score from S1 to S3, while the ID-RFD and the control groups made smaller gains in S3.

Table 5.5 Means and standard deviations for the organisation gain score by stage and group condition

Group	Stage	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	S1	0.40	0.63
	S3	0.67	0.62
ID-RFD (<i>n</i> = 16)	S1	0.44 ¹²	0.51
	S3	0.31	0.48
IS+RFD (<i>n</i> = 15)	S1	0.33	0.62
	S3	0.47 ¹³	0.52
IS-RFD (<i>n</i> = 14)	S1	0.21	0.70
	S3	0.29	0.47
Control (<i>n</i> = 17)	S1	0.29	0.69
	S3	0.24 ¹⁴	0.44

Figure 5.3 Mean organisation gain scores in S1 and S3 by group condition



¹² $z_{\text{kurtosis}} = 2.03$

¹³ $z_{\text{kurtosis}} = 2.06$

¹⁴ $z_{\text{skewness}} = 2.49$

Table 5.6 summarises the effect of stage and condition on the organisation gain score¹⁵. There was no main effect for either stage, suggesting that the five groups as a whole did not produce different gain scores across the two stages. There was neither a main effect for group condition, suggesting that the five groups produced in general similar gain scores if the factor of stage was ignored. The interaction effect between stage and group condition was not significant, indicating that the difference in gain scores across stages did not vary according to group condition.

Pairwise comparisons revealed that none of the five groups produced significantly larger gains in S3 than they did in S1. With respect to between-subjects differences, the five groups made in general similar gains during S1. They showed differences in S3, but none of the differences was statistically significant. The largest difference occurred between the ID+RFD group and the control group ($p = .186$, $d = 0.81$).

Table 5.6 Effects of stage and group condition on the organisation gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	.36	.549	.005
group	1.31	.275	.068
stage \times group	.54	.708	.029

Note. The *df* was 1 for stage, and the *df* was 4 for group condition as well as the two-way interaction.

5.3.2 The effect of WF type and \pm RFD on autonomous revision: Organisation gain score across stages

Table 5.7 includes the descriptive statistics on treatment groups' organisation gain scores by stage, WF type, and \pm RFD. Figure 5.4 demonstrates the mean gain scores. It can be seen from the graph that students receiving either ID or IS produced some increase in the organisation gain score from S1 to S3. On the other hand, the +RFD students' gain score increased from S1 to S3 while the -RFD students' gain reduced.

¹⁵ Box's test and Levene's test showed no violation of the assumptions for the ANOVA.

Table 5.7 Means and standard deviations for the organisation gain score by stage, WF type, and \pm RFD

	Stage	<i>M</i>	<i>SD</i>
ID (<i>n</i> = 31)	S1	0.42	0.56
	S3	0.48	0.57
IS (<i>n</i> = 29)	S1	0.28	0.65
	S3	0.38 ¹⁶	0.49
+RFD (<i>n</i> = 30)	S1	0.37	0.61
	S3	0.57	0.57
-RFD (<i>n</i> = 30)	S1	0.33	0.61
	S3	0.30 ¹⁷	0.47

Figure 5.4 Mean organisation gain scores in S1 and S3 by WF type and \pm RFD

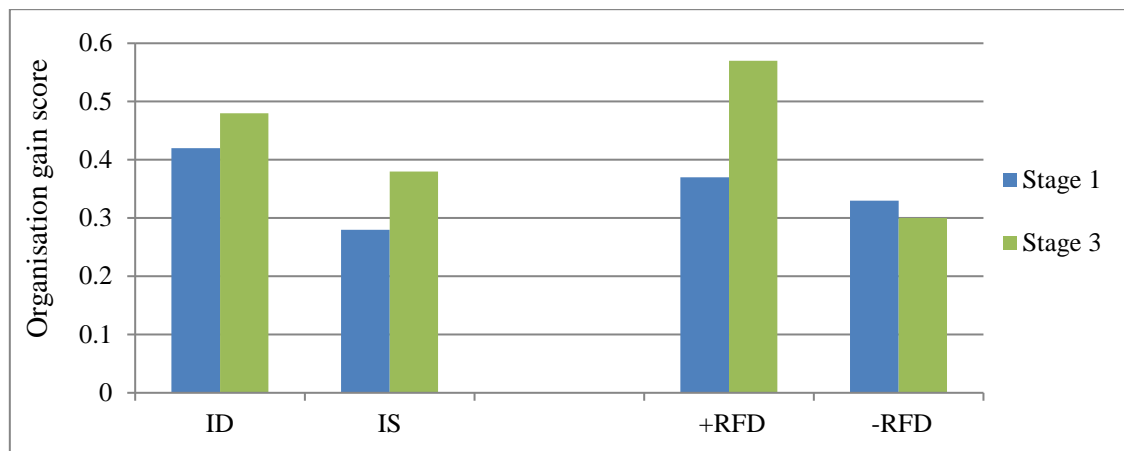


Table 5.8 lists the effect of stage, WF type, and \pm RFD on the organisation gain score¹⁸. It is shown that there was no main effect for stage, WF type, or \pm RFD, and there was no significant two-way or three-way interaction effect.

¹⁶ $z_{\text{kurtosis}} = 2.20$

¹⁷ $z_{\text{skewness}} = 2.15$

¹⁸ Box's tests and Levene's tests showed no violation of the assumptions for the ANOVA.

Post hoc pairwise comparisons to break up the interaction between stage and WF type suggested no significant within- or between-subjects differences. Therefore, WF type did not play a role in student autonomous revision. However, pairwise comparisons to explore the interaction between stage and \pm RFD showed that although the +RFD students and the -RFD students had similar gain scores in S1, the +RFD students' gain score in S3 was marginally significantly larger than that of the -RFD students ($p = .053$, $d = 0.51$). A closer examination of the three-way interaction found that the advantage of the +RFD was more evident when it concurred with the ID feedback, and this observation corresponded to the detected advantage of the ID+RFD condition in increasing organisation gain scores across stages.

Table 5.8 Effects of stage, WF type, and \pm RFD on the organisation gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	.71	.403	.013
WFT	1.48	.229	.026
\pm RFD	2.11	.152	.036
stage \times WFT	.02	.878	.000
stage \times \pm RFD	1.22	.274	.021
WFT \times \pm RFD	.00	.969	.000
stage \times WFT \times \pm RFD	.64	.425	.011

Note. WFT = WF type. The *df* was 1 for all the *F* values.

5.3.3 Summary of results for organisation quality

(1) Statistically significant contrasts concerning the five group conditions

None of the five groups made a significantly different gain score from S1 to S3. On the other hand, the five groups did not show significantly different gain scores in S1, but during S3 the ID+RFD group produced a larger mean gain score than the control group ($p = .186$, $d = 0.81$). This result suggested that the ID+RFD condition was somewhat

effective in promoting student revision skills with respect to organisation quality improvement.

(2) Statistically significant contrasts concerning the factors of WF type and \pm RFD

No significant within- or between-subjects contrast was found concerning the factor of WF type, and no significant within-subjects contrast was observed concerning \pm RFD conditions. However, the +RFD condition brought a larger gain than the –RFD condition for S3 ($p = .053$, $d = 0.51$), and the difference approached statistical significance. This advantage was more evident when the +RFD condition was combined with the ID feedback.

5.4 Results for Grammatical Accuracy

5.4.1 The effect of group conditions on autonomous revision: Grammatical accuracy gain score across stages

Table 5.9 shows the descriptive data on grammatical accuracy gain scores in S1 and S3 by group condition. Figure 5.5 illustrates the mean gain scores presented in the table. The bar graph shows that the ID+RFD, the ID–RFD, and the IS–RFD groups produced some increase in the gain scores across stages, while the IS+RFD and the control groups made smaller gains in S3.

Table 5.9 Means and standard deviations for the grammatical accuracy gain score by stage and group condition

Group	Stage	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	S1	0.49	0.81
	S3	0.70	0.64
ID-RFD (<i>n</i> = 16)	S1	0.53	1.06
	S3	0.70	0.72
IS+RFD (<i>n</i> = 15)	S1	0.67	0.83
	S3	0.63	1.01
IS-RFD (<i>n</i> = 14)	S1	0.51	1.36
	S3	0.78	0.56
Control (<i>n</i> = 17)	S1	0.53	1.03
	S3	0.45	1.05

Figure 5.5 Mean grammatical accuracy gain scores in S1 and S3 by group condition

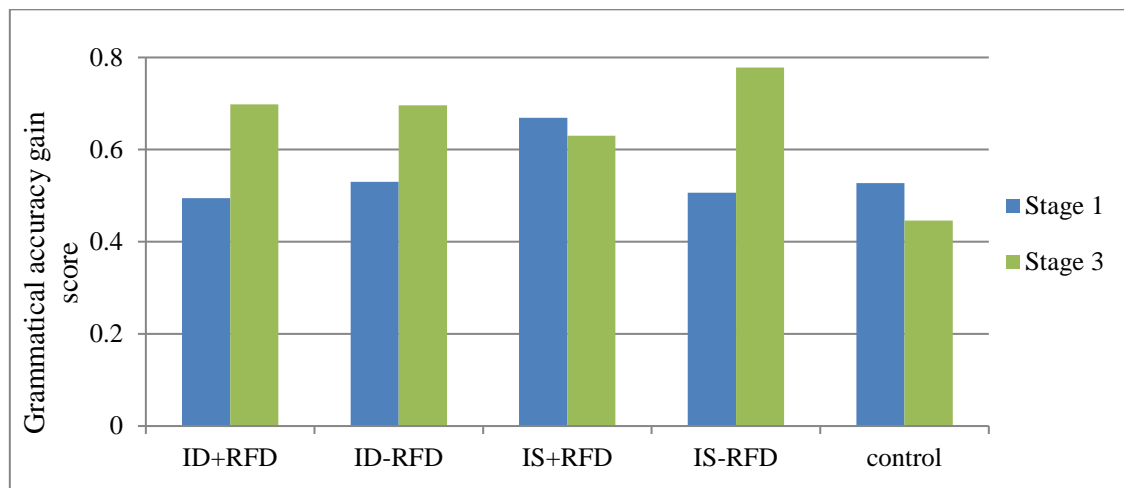


Table 5.10 presents the effect of stage and condition on grammatical accuracy gain score¹⁹. There was no main effect for stage, suggesting that the five groups as a whole did not produce different gain scores across the two stages. There was neither a main effect for group condition, suggesting that the five groups produced in general similar grammatical accuracy gain scores if the factor of stage was ignored. The interaction effect between stage and group condition was not significant, indicating that the difference in gain scores across stages did not vary according to group condition.

Post hoc pairwise comparisons showed that none of the groups made a significantly larger gain in S3 than in S1. In addition, the five groups' gain scores were neither significantly different in S1 nor in S3. To conclude, none of the group conditions had an effect on student autonomous revision with respect to grammatical accuracy increase.

Table 5.10: Effects of stage and group condition on the grammatical accuracy gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	.53	.468	.007
group	.14	.965	.008
stage \times group	.24	.916	.013

Note. The *df* was 1 for stage, and the *df* was 4 for group condition as well as the two-way interaction.

5.4.2 The effect of WF type and \pm RFD on autonomous revision: grammatical accuracy gain score across stages

Table 5.11 shows the descriptive statistics on treatment groups' grammatical accuracy gain scores by stage, WF type, and \pm RFD. Figure 5.6 illustrates the mean gain scores presented in the table. The bar graph shows that both students receiving ID and those receiving IS made some increase in the gain score across stages, and that both the +RFD and the -RFD students made some increase.

¹⁹ Box's tests of equality of covariance matrices and Levene's tests of equal variances showed no violation of the assumptions for the ANOVA.

Table 5.11 Means and standard deviations for the grammatical accuracy gain score by stage, WF type, and \pm RFD

	Stage	<i>M</i>	<i>SD</i>
ID (<i>n</i> = 31)	S1	0.51	0.93
	S3	0.70	0.67
IS (<i>n</i> = 29)	S1	0.59	1.10
	S3	0.70	0.82
+RFD (<i>n</i> = 30)	S1	0.58	0.81
	S3	0.66	0.83
-RFD (<i>n</i> = 30)	S1	0.52	1.19
	S3	0.73	0.64

Figure 5.6 Mean grammatical accuracy gain scores in S1 and S3 by WF type and \pm RFD

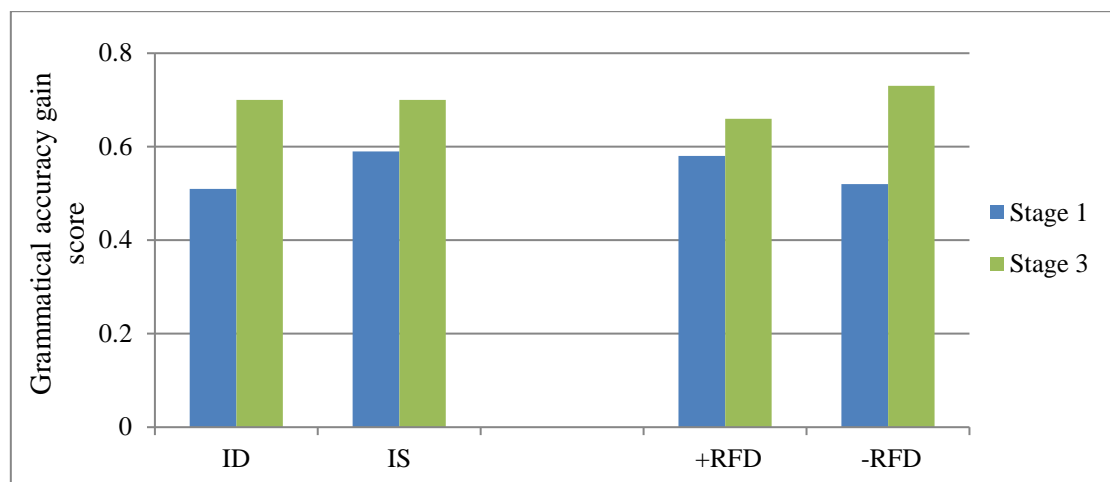


Table 5.12 summarises the effect of stage, WF type, and \pm RFD on grammatical accuracy gain score²⁰. As the table shows, there was no main effect for stage, WF type, or \pm RFD, and there was no significant two-way or three-way interaction effect.

Post hoc pairwise comparisons that were computed to break up the interaction between stage and WF type found no significant within- or between-subjects differences. Likewise, pairwise comparisons to explore the interaction between stage and \pm RFD found no significant differences. Therefore, neither WF type nor \pm RFD played a role in student autonomous revision in terms of grammatical accuracy increases.

Table 5.12 Effects of stage, WF type, and \pm RFD on the grammatical accuracy gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	.78	.382	.014
WFT	.07	.798	.001
\pm RFD	.00	.977	.000
stage \times WFT	.04	.843	.001
stage \times \pm RFD	.16	.689	.003
WFT \times \pm RFD	.01	.941	.000
stage \times WFT \times \pm RFD	.26	.612	.005

Note. WFT = WF type. The *df* was 1 for all the *F* values.

5.4.3 Summary of results for grammatical accuracy

No significant contrast was found in relation to the grammatical accuracy gain score concerning the five group conditions. This indicated that none of the group conditions helped to enhance student ability to increase grammatical accuracy when they revised their original draft without receiving external feedback.

Likewise, no significant contrast was found in relation to grammatical accuracy gain scores concerning the factors of WF type and \pm RFD. This indicated that neither factor

²⁰ Box's tests and Levene's tests showed no violation of the assumptions for the ANOVA.

was significant in the treatment on student ability to improve grammatical accuracy autonomously.

5.5 Results for Lexical Accuracy

5.5.1 The effect of treatment conditions on autonomous revision: Lexical accuracy gain score across stages

Table 5.13 lists the descriptive results for the lexical accuracy gain score in S1 and S3 by group condition. Figure 5.6 illustrates the mean gain scores. The bar graph illustrates that the ID–RFD, the IS–RFD, and the IS+RFD groups made some increase in the gain score from S1 to S3. The control group produced the same gain score in the two stages, while the ID+RFD group made a smaller gain in S3.

Table 5.13 Means and standard deviations for the lexical accuracy gain score by stage and group condition

Group	Stage	<i>M</i>	<i>SD</i>
ID+RFD (<i>n</i> = 15)	S1	0.30	.76
	S3	0.25	.63
ID–RFD (<i>n</i> = 16)	S1	0.21	1.01
	S3	0.37	0.65
IS+RFD (<i>n</i> = 15)	S1	0.28	0.88
	S3	0.33	0.94
IS–RFD (<i>n</i> = 14)	S1	0.29	0.87
	S3	0.41	0.72
Control (<i>n</i> = 17)	S1	0.21	1.02
	S3	0.21	0.90

Figure 5.7 Mean lexical accuracy gain scores in S1 and S3 by group condition

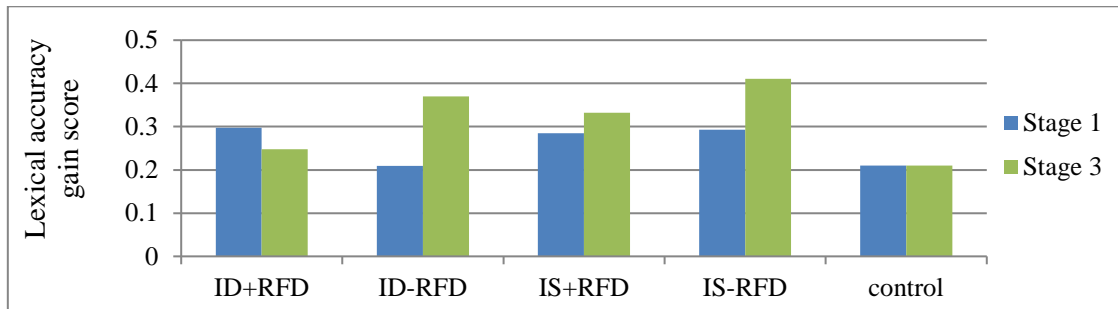


Table 5.14 lists the effect of stage and group condition on the lexical accuracy gain score²¹. It can be seen that there was no main effect for stage, suggesting that the five groups as a whole did not produce different gain scores across stages. There was neither a main effect for group condition, suggesting that the five groups produced in general similar lexical accuracy gain scores if the factor of stage was ignored. In addition, the interaction effect was not significant, indicating that the difference in gain scores across stages did not vary with group condition.

Table 5.14 Effects of stage and group condition on lexical accuracy gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	.19	.662	.003
group	.10	.983	.005
stage \times group	.09	.985	.369

Note. The *df* was 1 for stage, and the *df* was 4 for group condition as well as the two-way interaction.

Post hoc pairwise comparisons revealed no significant within- or between-subjects difference. To conclude, none of the group conditions had an effect on student autonomous revision with respect to lexical accuracy increase.

²¹ Box's tests of equality of covariance matrices and Levene's tests of equal variances showed no violation of the assumptions for the ANOVA.

5.5.2 The effect of WF type and \pm RFD on autonomous revision: Lexical accuracy gain score across stages

Table 5.15 shows the descriptive statistics on treatment groups' lexical accuracy gain scores by the factors of WF type and \pm RFD. Figure 5.8 illustrates the mean scores listed in the table. As the graph shows, students receiving either type of feedback made some increase in the gain score across stages. Moreover, the +RFD students' gain score remained the same across stages, while the -RFD students make some increase from S1 to S3.

Table 5.15 Means and standard deviations for the lexical accuracy gain score by stage, WF type, and \pm RFD

	Stage	<i>M</i>	<i>SD</i>
ID (<i>n</i> = 31)	S1	0.25	0.94
	S3	0.31	0.64
IS (<i>n</i> = 29)	S1	0.29	0.86
	S3	0.37	0.83
+RFD (<i>n</i> = 30)	S1	0.29	0.81
	S3	0.29	0.79
-RFD (<i>n</i> = 30)	S1	0.25	0.98
	S3	0.39	0.67

Figure 5.8 Mean lexical accuracy gain scores in S1 and S3 by WF type and \pm RFD

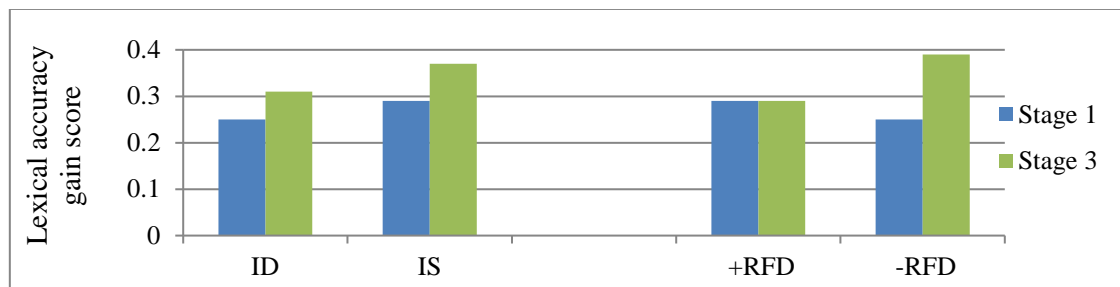


Table 5.16 shows the effect of stage, WF type and \pm RFD on the lexical accuracy gain score²². There was no main effect for stage, WF type, or \pm RFD, and there were no significant two-way or three-way interactions.

Table 5.16 Effects of stage, WF type, and \pm RFD on the lexical accuracy gain score

	<i>F</i>	<i>p</i>	partial η^2
stage	.24	.624	.004
WFT	.09	.767	.002
\pm RFD	.03	.855	.001
stage \times WFT	.01	.923	.000
stage \times \pm RFD	.25	.618	.004
WFT \times \pm RFD	.01	.937	.000
stage \times WFT \times \pm RFD	.06	.802	.001

Note. WFT = WF type. The *df* was 1 for all the *F* values.

Post hoc pairwise comparisons to break up the interaction between stage and WF type showed no significant within- or between-subjects differences. Likewise, pairwise comparisons to explore the interaction between stage and \pm RFD showed no significant differences. Therefore, neither feedback type was effective in improving student revision skills regarding lexical accuracy increase. In the same way, the factor of \pm RFD did not make a difference in affecting student autonomous revision.

5.5.3 Summary of results for lexical accuracy

As the results for grammatical accuracy, no significant contrast was observed in relation to the lexical accuracy gain score concerning the five group conditions. This indicated that none of the group conditions helped to enhance students' skills to increase lexical accuracy when they revised their draft without external feedback.

²² Box's tests and Levene's tests showed no violation of the assumptions for the ANOVA.

No significant contrast was found concerning the factors of WF type and \pm RFD. This suggested that neither factor played a role in the treatment effect on student autonomous revision in relation to lexical accuracy.

5.6 Summary of Results Regarding the Effect of Treatment on Autonomous Revision in Terms of Text Quality across Drafts

5.6.1 The effect of group conditions and the mediation of textual level

Unlike the results reported in Chapter Four that suggested definite immediate effect of treatment conditions on revision, the results in relation to the effect of group conditions on autonomous revision revealed less benefit gendered by treatment. Of the four treatment conditions, only the ID+RFD condition was found to be somewhat effective in helping students produce better autonomous revision outcome for content quality. Specifically, the content gain score of the ID+RFD group in S3 was larger than their content gain in S1, and the increase approached statistical significance ($p = .069$, $d = 0.69$).

5.6.2 The effect of WF type and \pm RFD and the mediation of textual level

Overall, the factor of WF type made no difference in the effect of treatment on autonomous revision. The ID and the IS feedback were similarly (in)effective for improving students' autonomous revision outcome for any of the four textual levels.

Neither did the factor of \pm RFD play a role in treatment effect on autonomous revision for grammatical accuracy or lexical accuracy. However, it made a difference in the autonomous revision on discourse levels with an advantage of the +RFD condition. The +RFD treatment enabled the participants to make a significantly larger improvement in content when they autonomously revised their draft in S3 compared to the improvement they made in S1 ($p = .038$, $d = 0.54$). With respect to revision in organisation, the +RFD students made larger between-draft improvement in S3 than the -RFD students, and the difference approached statistical significance ($p = .053$, $d = 0.51$).

CHAPTER SIX

RESULTS (3): THE EFFECT OF TREATMENT ON NEW WRITING

6.1 Overview

This chapter addresses the third set of research questions which investigated the effect of treatment on student new writing. The data used to answer the questions were based on the textual analysis of the written texts participants produced in the initial drafts of each writing-revising task (i.e., Tasks 1–5) over the study. As introduced in section 3.8.4.2.3, first-draft scores in Task 1 and Task 2 were averaged to represent student text quality in new writing before the treatment, coded as scores in Task 1/2. The baseline scores (i.e., scores in Task 1/2), as well as first-draft scores in Tasks 3–5 were submitted to statistical analysis.

RQ3 Does treatment have any effect on student new writing?

RQ3.1 What are the relative effects of different treatment methods on student new writing?

This research question attempted to examine whether the treatment groups varied from each other and in comparison to the control group in terms of text quality in new writing. ANOVA tests were run to calculate the effects of task and group condition on text quality. Following each ANOVA test, pairwise comparisons examined whether significant within-subjects differences (i.e., the difference in text quality across tasks by each group) and between-subjects differences (i.e., the difference in text quality between the groups for a certain task) existed.

RQ3.2 Do the factors of WF type and revision-focus direction play a role in the effects of treatment methods on student new writing?

This research question investigated, on one hand, whether students receiving ID varied from students receiving IS in text quality in new writing, and on the other hand, whether the +RFD students varied from the –RFD students. The analyses started with ANOVAs testing the effects of draft, WF type, and \pm RFD on text quality. Post hoc pairwise

comparisons examined whether there were significant within-subjects differences (i.e., the difference in text quality across tasks by students receiving either type of WF, and that by students either receiving RFD or not receiving RFD) and between-subjects differences (i.e., the difference in text quality between students receiving ID and students receiving IS, and the difference in text quality between the +RFD and the – RFD students for a certain task).

RQ3.3 Does textual level mediate the effect of treatment?

This research question investigated whether the effects of different treatment methods vary with textual level, and whether the effects of WF type and \pm RFD vary with textual level. To answer this question, the results for all the textual levels in relation to RQ1.1 and RQ1.2 were synthesised; comparisons were drawn between the treatment effects on the textual levels by observing the relevant statistically significant data and effect sizes.

Following the pattern to report results in Chapter Four and Chapter Five, this chapter will present results in sections 6.2–6.5 itemising textual levels. The report for each textual level will start with the outcome in response to RQ3.1 (effects of group conditions) and proceed to the outcome with respect to RQ3.2 (effects of WF type and \pm RFD). Finally, section 6.6 will summarise the results in relation to RQ3.1 and RQ3.2 and report the results with respect to RQ3.3 (the mediation of textual level in treatment effect).

6.2 Results for Content Quality in New Writing

6.2.1 The effect of group conditions on content quality in new writing

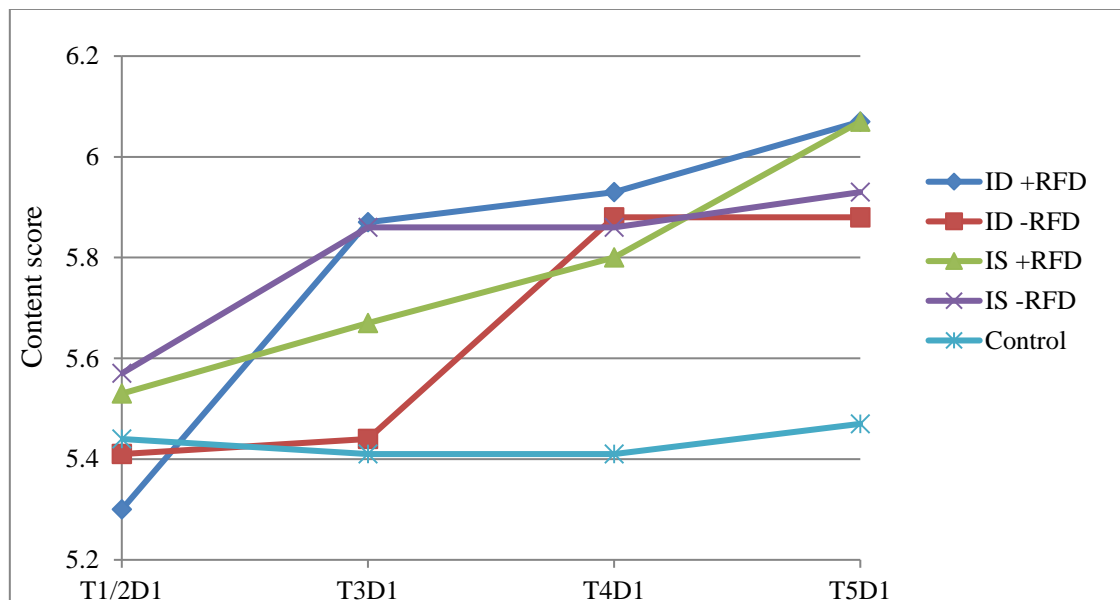
The descriptive statistics on first-draft content quality are presented in Table 6.1 and illustrated in Figure 6.1. It can be seen that on the whole the four treatment groups' first-draft scores showed a gradual increase from Task 1/2 to Task 5, while the control group's first-draft score remained relatively unchanged over the tasks.

Table 6.1 Means and standard deviations for first-draft content quality by task and group condition

Group	Task 1/2	Task 3	Task 4	Task 5
ID+RFD (<i>n</i> = 15)	5.30 (1.00)	5.87 (1.13)	5.93 (0.80)	6.07 (0.70)
ID-RFD (<i>n</i> = 16)	5.41 (0.71)	5.44 (1.15)	5.88 (0.96)	5.88 (1.15)
IS+RFD (<i>n</i> = 15)	5.53 ¹ (0.83)	5.67 (1.18)	5.80 (1.01)	6.07 (0.88)
IS-RFD (<i>n</i> = 14)	5.57 (0.87)	5.86 (0.95)	5.86 (1.03)	5.93 (1.07)
Control (<i>n</i> = 17)	5.44 (0.68)	5.41 (0.80)	5.41 (0.62)	5.47 ² (0.72)

Note. Standard deviations are put in parentheses.

Figure 6.1 Mean first-draft content scores over Tasks 1/2–5 by group condition



¹ $z_{skewness} = 2.55$

² $z_{skewness} = 2.30$

Table 6.2 Effects of task and group condition on first-draft content quality in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	4.88**	.003	.063
group	.87	.484	.046
task \times group	.65	.797	.035

Note. The *df* was 3 for task, 4 for condition, and 12 for the two-way interaction.
 ** $p < .01$

A two-way repeated-measures ANOVA was performed to assess the effects of task and group condition on the content quality of the first draft (see Table 6.2)³. There was a significant main effect of task, $F(3, 216) = 4.88, p = .003, \text{partial } \eta^2 = .063$, suggesting that all the groups taken together produced significantly different first-draft scores across tasks. However, neither group condition nor the two-way interaction was significant. The nonsignificant interaction effect indicated that overall the different content quality across tasks did not vary according to group conditions. Between-subjects pairwise comparisons revealed that the five groups produced in general similar content quality of the first draft in any of the tasks. Within-subjects pairwise comparisons found that for the ID+RFD condition the mean difference between scores in T1/2 and T5 was significant ($p = .014, d = 0.89$).

6.2.2 The effects of WF type and \pm RFD on content quality in new writing

Table 6.3 shows the descriptive statistics on content scores in new writing over Tasks 1/2–5 itemised by WF type and \pm RFD. It can be seen that participants with either WF type and under either of the \pm RFD conditions produced improvement across tasks. Figures 6.2 and 6.3 illustrate the mean scores displayed in Table 6.3 by WF type and \pm RFD, respectively. According to the graphs, students receiving either type of feedback

³ Box's test of equality of covariance matrices, Levene's test of equal variances, and Mauchly's test of sphericity showed no violation of the assumptions.

made some improvement from Task 1/2 to Task 5, and the +RFD students as well as the -RFD students produced some increase over time.

Table 6.3 Means and standard deviations for first-draft content quality by task, WF type, and \pm RFD

	Task 1/2	Task 3	Task 4	Task 5
ID	5.35	5.65	5.90	5.97
(<i>n</i> = 31)	(0.85)	(1.14)	(0.87)	(0.95)
IS	5.55	5.76	5.83	6.00
(<i>n</i> = 29)	(0.84)	(1.06)	(1.00)	(0.96)
+RFD	5.42	5.77	5.87	6.07
(<i>n</i> = 30)	(0.91)	(1.14)	(0.90)	(0.79)
-RFD	5.48	5.63	5.87	5.90
(<i>n</i> = 30)	(0.78)	(1.07)	(0.97)	(1.09)

Note. Standard deviations are put in parentheses.

Figure 6.2 Mean content scores of the first draft over Tasks 1/2–5 by WF type

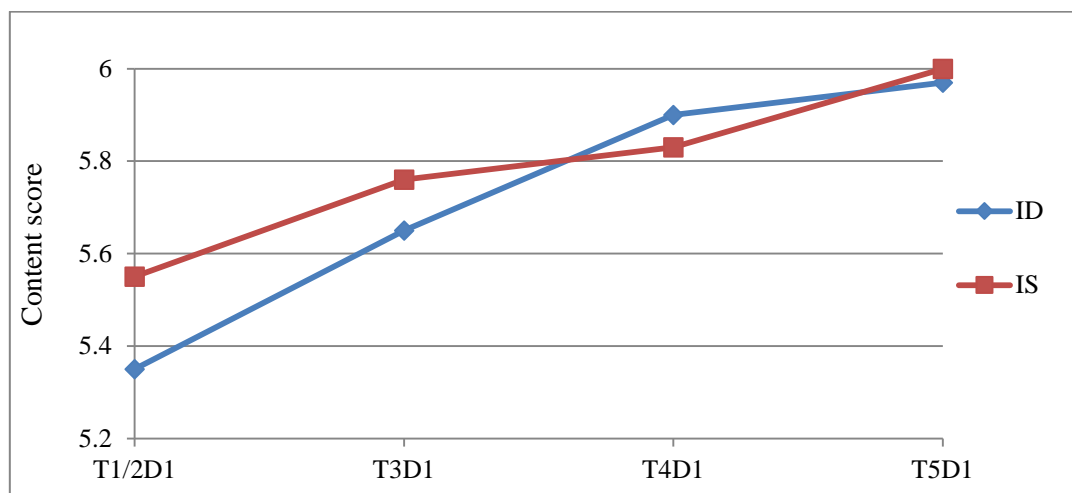


Figure 6.3 Mean content scores of the first draft over Tasks 1/2–5 by ±RFD

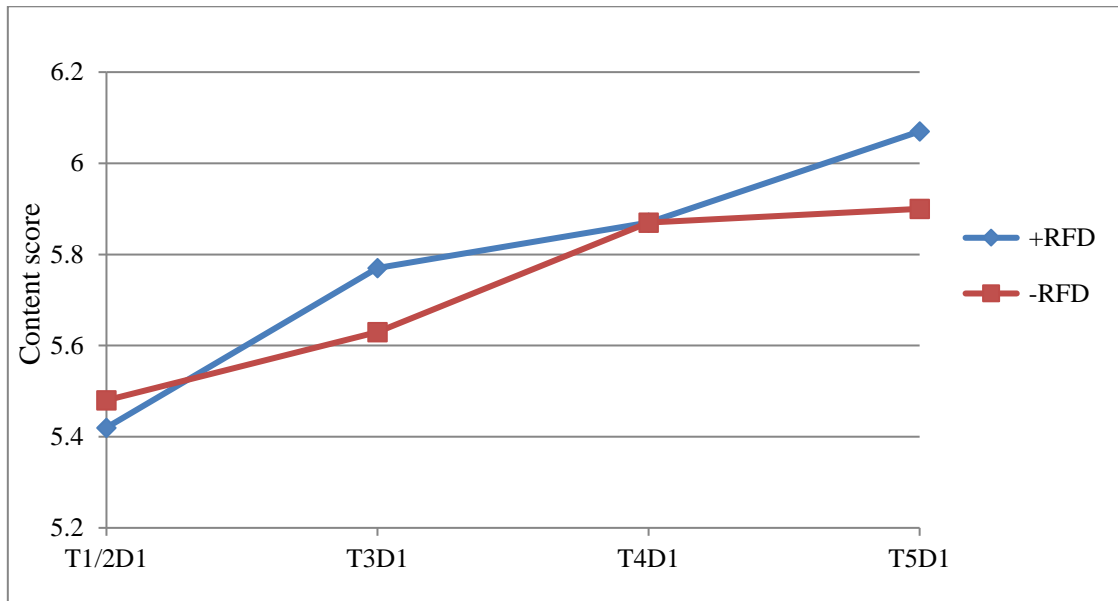


Table 6.4 displays the effects of task, WF type and ±RFD on content quality in new pieces of writing⁴. There was a significant main effect of task, $F(3, 168) = 5.44, p = .001$, partial $\eta^2 = .089$, but neither WF type nor ±RFD was significant. None of the two-way and three-way interactions was significant. This outcome indicated that although treatment groups, as a whole, produced significantly different scores across tasks, the differences did not vary with either WF type or ±RFD.

⁴ Box's test, Levene's test, and Mauchly's test found no violation of the assumptions.

Table 6.4 Effects of task, WF type, and \pm RFD on first-draft content quality in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	5.44**	.001	.089
WFT	.12	.727	.002
\pm RFD	.08	.776	.001
task \times WFT	.35	.786	.006
task \times \pm RFD	.30	.823	.005
WFT \times \pm RFD	.13	.724	.002
task \times WFT \times \pm RFD	.59	.622	.010

Note. WFT = WF type. The *df* was 3 for task as well as the two-way and three-way interactions, and the *df* was 1 for WF type and \pm RFD.

** $p < .01$

Pairwise comparisons showed no significant between-subjects difference with respect to feedback type or \pm RFD. Within-subjects pairwise comparisons for WF type suggested significant contrasts concerning the ID feedback. The mean difference between Task 1/2 and Task 4 ($p = .016$, $d = 0.64$) as well as that between Task 1/2 and Task 5 ($p = .005$, $d = 0.69$) was statistically significant. Within-subjects pairwise comparisons for the factor of \pm RFD found significant contrasts concerning the +RFD condition. The significant contrast occurred between the mean scores of Task 1/2 and Task 5 ($p = .003$, $d = 0.76$).

6.2.3 Summary of results for content quality

(1) Statistically significant contrasts concerning the five group conditions

With respect to the significant within-subjects contrasts in relation to group conditions across Tasks 1/2–5, only the ID+RFD group produced content that was significantly better than the content they produced before the treatment ($p = .014$, $d = 0.89$). From another perspective, the above-stated significant improvement in new writing only occurred after the treatment was provided three times (for three writing-revising tasks),

which suggested that the ID+RFD treatment needed to be repeated to generate a significant effect on content quality in new writing. In terms of between-subjects contrasts, no significant differences were found across group conditions in any of the tasks under examination. In summary, of the five conditions the ID+RFD condition was relatively more effective in helping students improve content quality in new writing.

(2) Statistically significant contrasts concerning the factors of WF type and ±RFD

The significant within-subjects contrasts concerning the feedback types and ±RFD conditions were listed in Table 6.5. It can be seen that after two cycles of treatment, students receiving ID produced content that was significantly better than what they produced before the treatment, and the effect size was medium to large ($d = 0.64$). Such significant improvement recurred after a third cycle of treatment ($d = 0.69$). Students receiving IS did not bring significant improvement in new writing. The +RFD students produced significant improvement after three cycles of treatment, and the effect size was marginally large ($d = 0.76$). The -RFD students did not bring any significant improvement. There was no significant between-subjects difference in any of the tasks with respect to feedback type and ±RFD.

Table 6.5 Significant within-subjects contrasts concerning WF types and ±RFD conditions regarding first-draft content quality across Tasks 1/2–5

ID	IS	+RFD	-RFD
*T1/2 vs. T4 $d = 0.64$		**T1/2 vs. T5 $d = 0.76$	
**T1/2 vs. T5 $d = 0.69$			

Note. The blank cells indicate the absence of applicable results.

** $p < .01$

* $p < .05$

In summary, both factors of WF type and ±RFD played a role in participants' content quality in new pieces of writing. The ID feedback was more effective than IS. On the other hand, the +RFD condition showed an advantage over the -RFD condition. In addition, the ID feedback and the +RFD condition need repeated provision to the participants to take effect.

6.3 Results for Organisation Quality in New Writing

6.3.1 The effect of group conditions on organisation quality in new writing

Table 6.6 presents the descriptive results for organisation scores in the first draft by group condition. Figure 6.4 depicts the mean scores displayed in Table 6.6. The graph shows that the four treatment groups made some improvement, while the scores of the control group remained relatively unchanged.

Table 6.6 Means and standard deviations for first-draft organisation quality by task and group condition

Group	Task 1/2	Task 3	Task 4	Task 5
ID+RFD (<i>n</i> = 15)	5.17 (0.72)	5.93 (1.10)	6.13 (0.74)	6.20 (0.94)
ID–RFD (<i>n</i> = 16)	5.47 ⁵ (0.76)	5.63 (1.36)	6.06 (0.85)	6.25 (0.78)
IS+RFD (<i>n</i> = 15)	5.17 (1.03)	5.87 (0.99)	5.93 (0.88)	6.20 (0.94)
IS–RFD (<i>n</i> = 14)	5.61 (0.84)	5.93 (1.27)	6.14 (0.86)	6.21 (0.89)
Control (<i>n</i> = 17)	5.44 (0.63)	5.41 (0.87)	5.47 (0.80)	5.47 (0.80)

Note. Standard deviations are put in parentheses.

⁵ $z_{\text{skewness}} = 2.26$, $z_{\text{kurtosis}} = 2.07$

Figure 6.4 Mean first-draft organisation scores over Tasks 1/2–5 by group condition

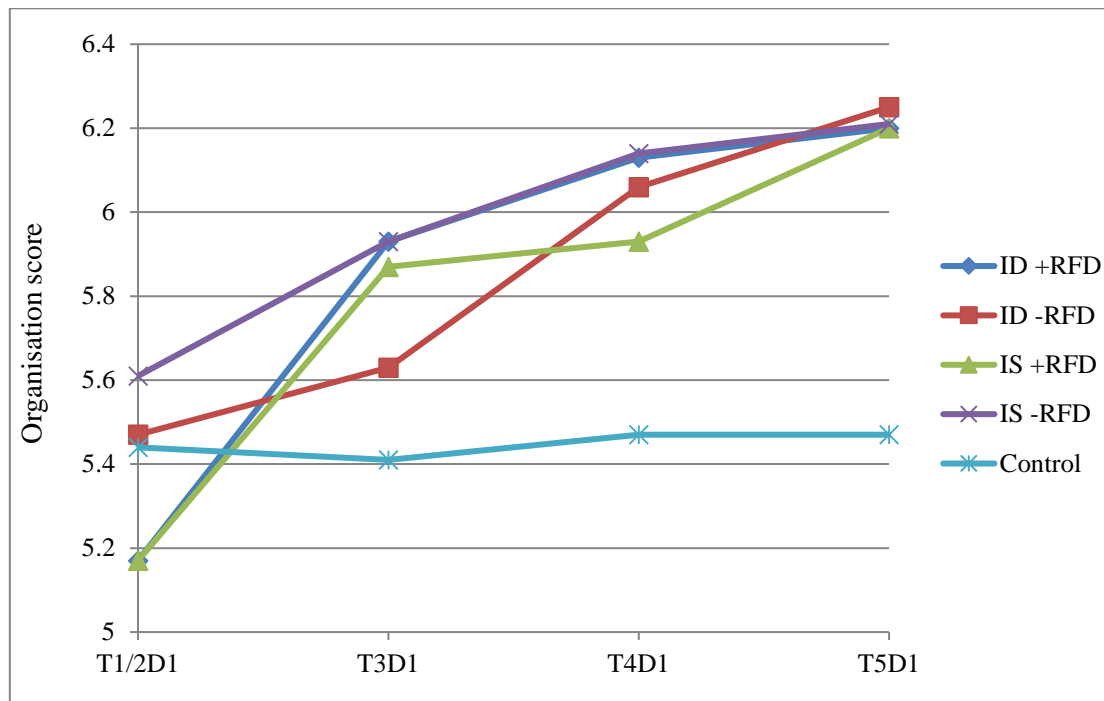


Table 6.7 lists the output of the ANOVA, i.e., effects of task and group condition on organisation quality in the first draft⁶. The only significant effect was found in task, $F(3, 216) = 14.98, p = .000, \text{partial } \eta^2 = .172$, indicating that the five groups altogether wrote significantly different first-draft scores across tasks, but that the differences did not vary according to group conditions. Between-subjects pairwise comparisons revealed that the five groups produced in general similar first-draft organisation quality in any of the tasks. However, within-subjects pairwise comparisons showed significant differences. The ID+RFD group's mean difference between T1/2 and T3 ($p = .018, d = 0.83$), between T1/2 and Task 4 ($p = .001, d = 1.32$), and between T1/2 and Task 5 ($p = .000, d = 1.23$) were all statistically significant. The ID–RFD group made a significant improvement from T1/2 to T5 ($p = .005, d = 1.01$). The IS+RFD group produced significant improvement from T1/2 to T3 ($p = .039, d = 0.69$), from T1/2 to T4 ($p = .012, d = 0.80$), and from T1/2 to T5 ($p = .000, d = 1.05$). However, the IS–RFD and the control groups made no significant improvement over time.

⁶ Box's test, Levene's test, and Mauchly's test showed no violation of the assumptions.

Table 6.7 Effects of task and group condition on first-draft organisation quality in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	14.98***	.000	.172
group	1.33	.267	.069
task \times group	1.37	.178	.071

Note. The *df* was 3 for task, 4 for condition, and 12 for the interaction between task and condition.

*** $p < .001$

6.3.2 The effects of WF type and \pm RFD on organisation quality in new writing

Table 6.8 lists the descriptive results for first-draft organisation quality over time. It shows that participants with either WF type and under either of the \pm RFD conditions produced improvement across tasks. Figures 6.5 and 6.6 provide a graphic illustration of the mean organisation scores by WF type and by \pm RFD, respectively. According to the figures, students receiving ID and students receiving IS made some improvement over time on a similar magnitude, and both the +RFD and the –RFD students produced some increase over tasks.

Table 6.8 Means and standard deviations for first-draft organisation quality by task, WF type, and \pm RFD

	Task 1/2	Task 3	Task 4	Task 5
ID	5.32	5.77	6.10	6.23
(n = 31)	(0.75)	(1.23)	(0.79)	(0.85)
IS	5.38	5.90	6.03	6.21
(n = 29)	(0.95)	(1.11)	(0.87)	(0.90)
+RFD	5.17	5.90	6.03	6.20
(n = 30)	(0.87)	(1.03)	(0.81)	(0.93)
-RFD	5.53	5.77	6.10	6.23
(n = 30)	(0.79)	(1.31)	(0.85)	(0.82)

Note. Standard deviations are put in parentheses.

Figure 6.5 Mean first-draft organisation scores over Tasks 1/2–5 by WF type

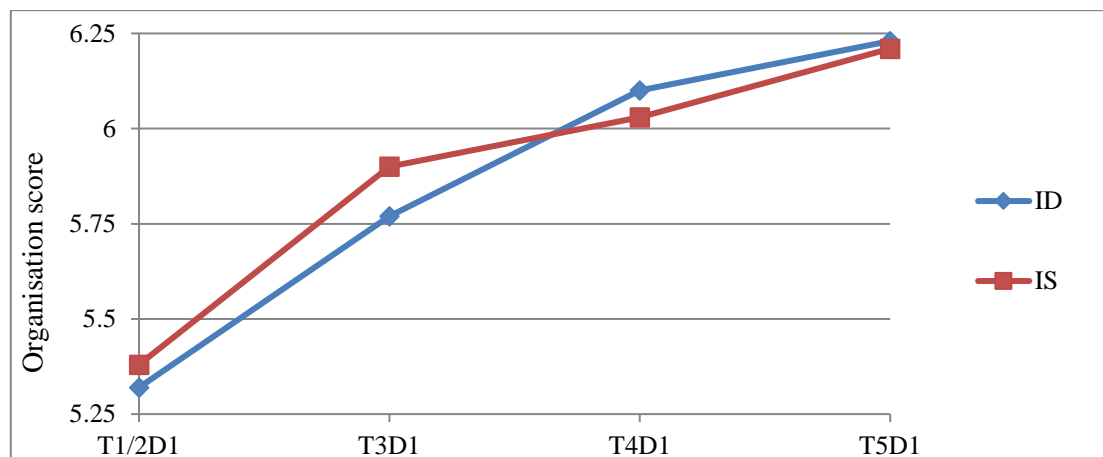
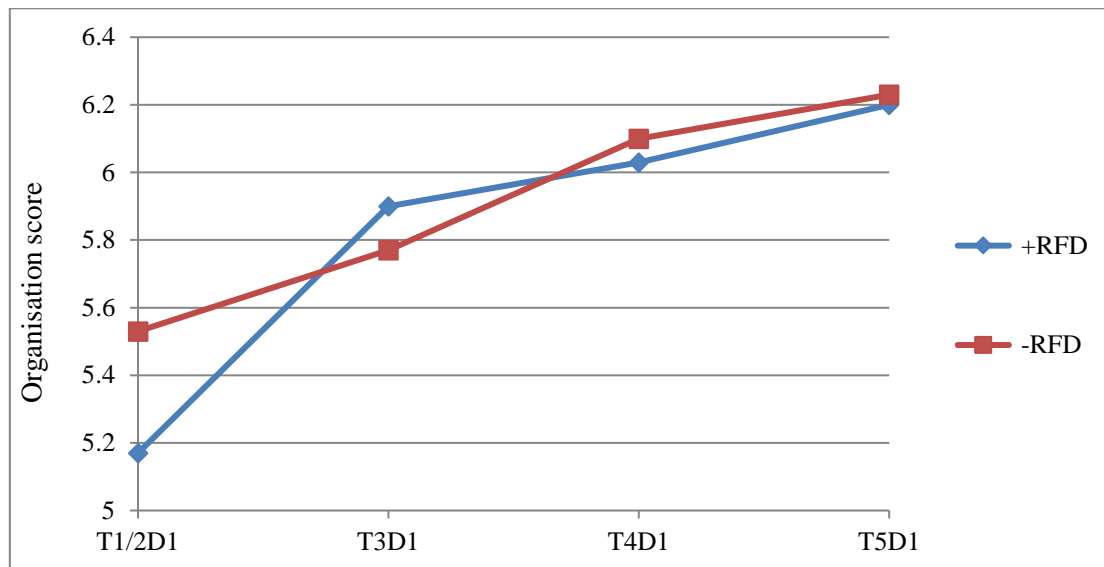


Figure 6.6 Mean first-draft organisation scores over Tasks 1/2–5 by \pm RFD



Based on the output of the ANOVA test, effects of task, WF type and \pm RFD on organisation quality in new writing are recorded in Table 6.19⁷. There was a significant main effect of task, $F(3, 168) = 17.56, p = .000$, partial $\eta^2 = .239$, but neither WF type nor \pm RFD was significant. None of the two-way and three-way interactions was significant. This outcome suggested that the four treatment groups, altogether, significantly improved across tasks, but the improvement did not vary with WF type or \pm RFD.

⁷ Box's test, Levene's test, and Mauchly's test found no violation of the assumptions.

Table 6.9 Effects of task, WF type, and \pm RFD on first-draft organisation quality in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	17.56***	.000	.239
WFT	.02	.886	.000
\pm RFD	.21	.647	.004
task \times WFT	.20	.895	.004
task \times \pm RFD	1.32	.271	.023
WFT \times \pm RFD	.25	.622	.004
task \times WFT \times \pm RFD	.24	.867	.004

Note. WFT = WF type. The *df* was 3 for task, the two-way interaction between task and WFT, the interaction between task and \pm RFD, and the three-way interaction. The *df* was 1 for WF type, \pm RFD, and their interaction effect.

*** $p < .01$

Within-subjects pairwise comparisons concerning WF types observed significant contrasts. Participants receiving ID condition significantly improved from Task 1/2 to Task 4 ($p = .000$, $d = 1.01$) and from Task 1/2 to Task 5 ($p = .000$, $d = 1.13$). Participants receiving IS significantly improved from Task 1/2 to Task 3 ($p = .041$, $d = 0.49$), from Task 1/2 to Task 4 ($p = .003$, $d = 0.71$), and from Task 1/2 to Task 5 ($p = .000$, $d = 0.89$). Within-subjects pairwise comparisons for the factor of \pm RFD also found significant contrasts. The +RFD students improved organisation quality from Task 1/2 and to Task 3 ($p = .001$, $d = 0.77$), from Task 1/2 and Task 4 ($p = .000$, $d = 1.03$), and from Task 1/2 to Task 5 ($p = .000$, $d = 1.15$). For the –RFD students, significant increase occurred between Task 1/2 and Task 4 ($p = .011$, $d = 0.69$) and between Task 1/2 to Task 5 ($p = .000$, $d = 0.86$). There was no significant between-subjects difference in any of the tasks with respect to feedback type or \pm RFD.

6.3.3 Summary of results for organisation quality

(1) Statistically significant contrasts concerning the five group conditions

Table 6.10 displays the significant within-subjects contrasts of the five treatment conditions across Tasks 1/2–5. The ID+RFD and the IS+RFD groups significantly improved organisation quality in new writing after mere one cycle of treatment; these two groups continued to produce significantly better organisation than what they produced before the treatment after a second cycle and a third cycle of treatment. Overall the respective effect sizes were large ($d = 0.83, 1.32, \text{ and } 1.23$ for ID+RFD; $d = 0.69, 0.80, \text{ and } 1.05$ for IS+RFD). The ID–RFD group made a significant improvement after three cycles of treatment with a large effect size ($d = 1.01$). By contrast, the IS–RFD and the control groups did not show significant improvement in new writing. In terms of between-subjects contrasts, no significant difference was found across the five groups in any of the tasks under examination.

Table 6.10 Significant within-subjects contrasts of the group conditions regarding first-draft organisation quality across Tasks 1/2–5

Group	First-draft organisation quality
ID+RFD	*T1/2 vs. T3 $d = 0.83$ **T1/2 vs. T4 $d = 1.32$ ***T1/2 vs. T5 $d = 1.23$
ID–RFD	**T1/2 vs. T5 $d = 1.01$
IS+RFD	*T1/2 vs. T3 $d = 0.69$ *T1/2 vs. T4 $d = 0.80$ ***T1/2 vs. T5 $d = 1.05$
IS–RFD	
Control	

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

In summary, of the five conditions the ID+RFD and the IS+RFD conditions were more effective in helping participants improve organisation quality in new writing, followed by the ID–RFD condition. The control condition was the least effective.

(2) Statistically significant contrasts concerning the factors of WF type and \pm RFD

Table 6.11 displays the significant within-subjects contrasts concerning the two feedback types and \pm RFD conditions. After two cycles of treatment, participants receiving ID produced organisation that was significantly better than what they produced before the treatment, and the effect size was large ($d = 1.01$). Such significant improvement recurred after three cycles of treatment ($d = 1.13$). Participants receiving IS wrote significantly better organisation than what they produced before the treatment when the treatment was provided merely once, and such significant improvement recurred after two cycles and after three cycles of treatment. The corresponding effect size increased from a medium one ($d = 0.49$) to a large one ($d = 0.89$). On the other hand, the +RFD students produced significant improvement from before the treatment when the treatment was given once, twice, and three times; the effect sizes were overall large ($d = 0.77, 1.03, \text{ and } 1.15$). The –RFD students produced significantly better organisation after the second treatment with a medium to large effect size ($d = 0.69$), and they significantly improved again after the third treatment with a large effect size ($d = 0.86$). There was no significant between-subjects difference in any of the tasks with respect to feedback type or \pm RFD.

Table 6.11 Significant within-subjects contrasts concerning WF types and \pm RFD conditions regarding organisation quality across Tasks 1/2–5

First-draft organisation quality	
ID	***T1/2 vs. T4 $d = 1.01$
	***T1/2 vs. T5 $d = 1.13$
IS	*T1/2 vs. T3 $d = 0.49$
	**T1/2 vs. T4 $d = 0.71$
	***T1/2 vs. T5 $d = 0.89$
+RFD	**T1/2 vs. T3 $d = 0.77$
	***T1/2 vs. T4 $d = 1.03$
	***T1/2 vs. T5 $d = 1.15$
-RFD	*T1/2 vs. T4 $d = 0.69$
	***T1/2 vs. T5 $d = 0.86$

*** $p < .001$

** $p < .01$

* $p < .05$

In summary, WF type did not play an evident role in participants' organisation quality in new pieces of writing. The ID feedback was in general similarly effective as the IS feedback. The IS feedback seemed to have an advantage when the treatment was provided once, but the ID feedback seemed to be more effective than the IS feedback in terms of effect size. On the other hand, the +RFD condition showed an advantage over the -RFD condition. Participants in this condition started to improve quality after one cycle of treatment while those in the -RFD condition achieved this after two cycles of treatments, and the former condition rendered a larger effect size.

6.4 Results for Grammatical Accuracy in New Writing

6.4.1 The effect of group conditions on grammatical accuracy in new writing

The descriptive results for grammatical accuracy were displayed in Table 6.12, and the means were illustrated graphically in Figure 6.7. As the graph shows, only the ID+RFD group exhibited a linear pattern of error ratio from Task 1/2 to Task 5.

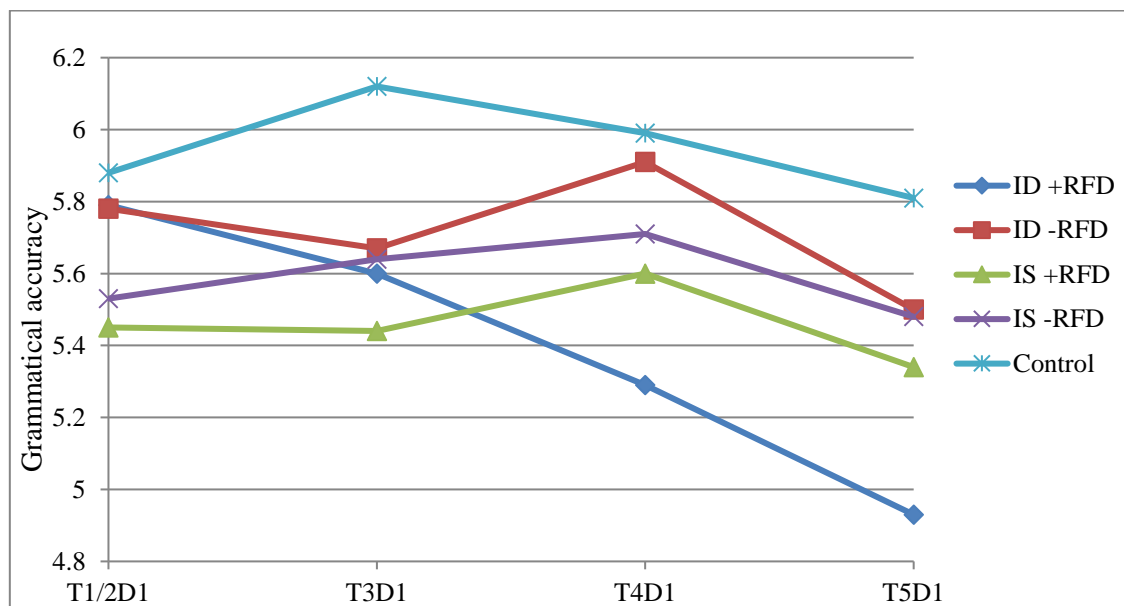
Table 6.12 Means and standard deviations for first-draft grammatical accuracy in task and group condition

Group	Task 1/2	Task 3	Task 4	Task 5
ID+RFD (<i>n</i> = 15)	5.79 (1.81)	5.60 (2.33)	5.29 (1.59)	4.93 ⁸ (1.29)
ID-RFD (<i>n</i> = 16)	5.78 (1.39)	5.67 (2.61)	5.91 (1.98)	5.50 (1.36)
IS+RFD (<i>n</i> = 15)	5.45 (1.92)	5.44 (2.50)	5.60 (2.11)	5.34 (1.38)
IS-RFD (<i>n</i> = 14)	5.53 (1.62)	5.64 (2.43)	5.71 (1.93)	5.48 (1.00)
Control (<i>n</i> = 17)	5.88 (0.96)	6.12 (1.82)	5.99 (2.15)	5.81 (1.17)

Note. Standard deviations are put in parentheses.

⁸ $z_{\text{skewness}} = 2.22$, $z_{\text{kurtosis}} = 2.04$

Figure 6.7 Mean first-draft grammatical accuracy scores over Tasks 1/2–5 by group condition



The output of the ANOVA to assess effects of task and group condition on grammatical accuracy in new writing is summarised in Table 6.13⁹. It can be seen that there was no main effect of task or group condition, and their interaction was not significant. Between-subjects pairwise comparisons revealed that the five groups produced in general similar first-draft grammatical accuracy in any of the tasks. Within-subjects pairwise comparisons found that for the ID+RFD condition the mean difference between T1/2 and T5 was significant ($p = .043$, $d = 0.55$), and that the other group conditions did not have any significant difference.

⁹ Box's test and Levene's test showed no violation of the assumptions. Mauchly's test showed a violation ($p = .000$) and the corresponding Greenhouse-Geisser estimate was .804 (i.e., greater than .75), so the Huynh-Feldt correction was used to report the effect of task and the interaction effect of task and group condition.

Table 6.13 Effects of task and group condition on first-draft grammatical accuracy in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	.78	.493	.011
group	.41	.798	.022
task \times group	.22	.995	.012

Note. The *df* was 2.64 for task, 4 for group, and 10.56 for the two-way interaction.

6.4.2 The effect of WF type and \pm RFD on grammatical accuracy in new writing

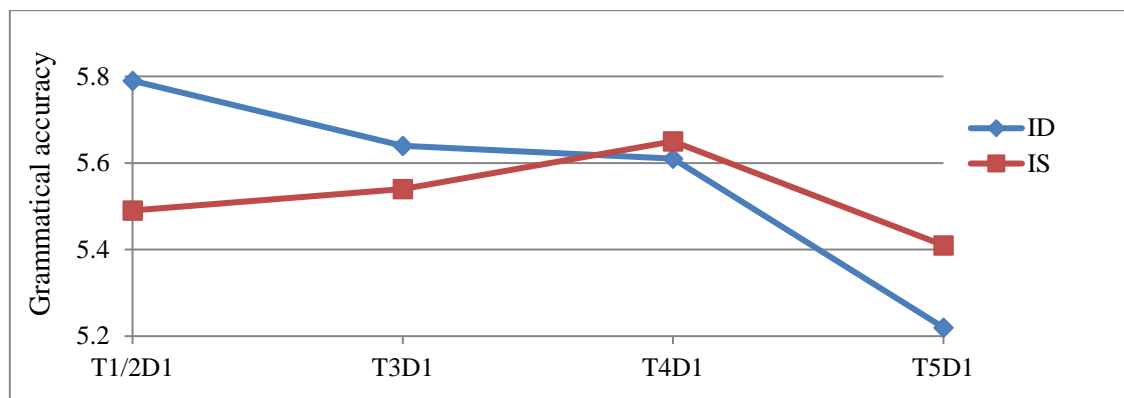
Table 6.14 includes the descriptive data on grammatical accuracy over tasks by the factors of WF type and \pm RFD. Figures 6.7 and 6.8 represent the means related to WF types and \pm RFD conditions, respectively. It is shown that students receiving the ID feedback reduced the error ratio continuously across tasks, while students receiving IS did not achieve that. In addition, the error ratio of the essays by the +RFD students showed some decrease across tasks, but the –RFD students did not decrease the error ratio continuously.

Table 6.14 Means and standard deviations for first-draft grammatical accuracy by task, WF type, and \pm RFD

	Task 1/2	Task 3	Task 4	Task 5
ID	5.79	5.64 ¹⁰	5.61	5.22 ¹¹
(n = 31)	(1.58)	(2.44)	(1.80)	(1.34)
IS	5.49	5.54	5.65	5.41
(n = 29)	(1.75)	(2.42)	(1.99)	(1.20)
+RFD	5.62	5.52	5.45	5.14
(n = 30)	(1.84)	(2.38)	(1.84)	(1.33)
-RFD	5.66	5.66	5.82	5.49
(n = 30)	(1.48)	(2.48)	(1.93)	(1.19)

Note. Standard deviations are put in parentheses.

Figure 6.8 Mean first-draft grammatical accuracy scores over Tasks 1/2–5 by WF type



¹⁰ $z_{\text{skewness}} = 2.04$

¹¹ $z_{\text{skewness}} = 2.34$

Figure 6.9 Mean first-draft grammatical accuracy scores over Tasks 1/2–5 by \pm RFD

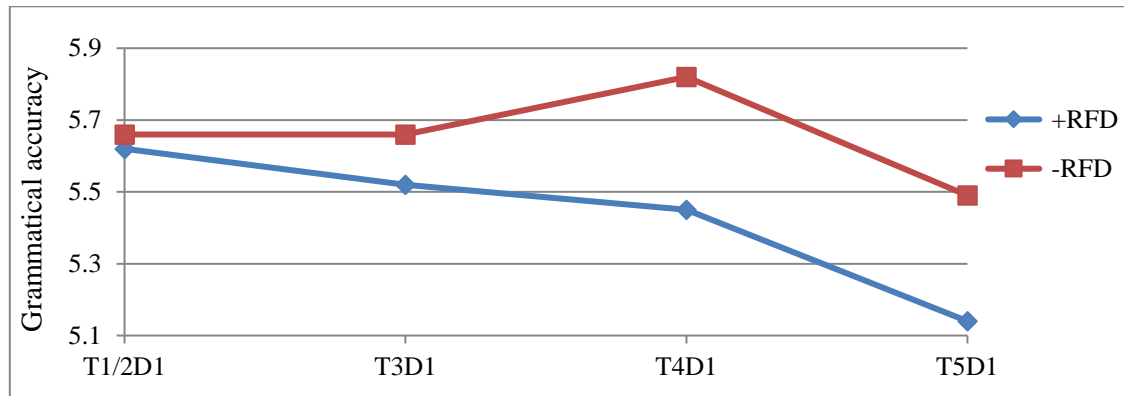


Table 6.15 presents the effects of WF type and \pm RFD on grammatical accuracy in new writing¹². There was no main effect of task, WF type, or \pm RFD, and the two-way and three-way interaction effects were neither significant.

Table 6.15 Effects of task, WF type, and \pm RFD on first-draft grammatical accuracy in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	.74	.507	.013
WFT	.01	.926	.000
\pm RFD	.35	.557	.006
task \times WFT	.35	.756	.006
task \times \pm RFD	.21	.860	.004
WFT \times \pm RFD	.05	.818	.001
task \times WFT \times \pm RFD	.22	.848	.004

Note. WFT = WF type. The *df* was 2.52 for task, the interaction between task and WFT, the interaction between task and \pm RFD, and the three-way interaction. The *df* was 1 for WF type, \pm RFD, and their interaction effect.

¹² Box's test and Levene's test showed no violation of the assumptions. Mauchly's test showed a violation ($p = .000$) and the corresponding Greenhouse-Geisser estimate was .763 (i.e., greater than .75), so the Huynh-Feldt correction was used to report the effect of task and its interaction with WF type and \pm RFD.

Within-subjects pairwise comparisons for WF type found that students receiving ID made a considerable, though nonsignificant, increase in grammatical accuracy from Task 1/2 to Task 5 ($p = .062$, $d = 0.39$). There was no significant contrast within the IS condition. Within-subjects pairwise comparisons for the factor of \pm RFD found no significant contrasts within either the +RFD condition or the -RFD condition. There was no significant between-subjects difference in any of the tasks with respect to feedback type or \pm RFD.

6.4.3 Summary of results for grammatical accuracy

(1) Statistically significant contrasts concerning the five group conditions

With respect to the significant within-subjects contrasts regarding the five group conditions across Tasks 1/2–5, only ID+RFD group increased grammatical accuracy significantly from before the treatment to after the treatment ($p = .043$, $d = 0.55$). This result indicated that the ID+RFD treatment needed to be repeated to increase student grammatical accuracy. In terms of between-subjects contrasts, no significant differences were found across the five group conditions in terms of first-draft grammatical accuracy in any of the tasks. In summary, of the five conditions the ID+RFD condition was relatively more effective in helping participants improve grammatical accuracy in new writing.

(2) Statistically significant contrasts concerning the factors of WF type and \pm RFD

No significant within-subjects or between-subjects difference was found concerning WF types or \pm RFD with respect to first-draft grammatical accuracy. However, given that the mean difference between Task 1/2 and Task 5 ($p = .062$, $d = 0.39$) for students receiving ID approached statistical significance, the factor of WF type seemed to play a role in grammatical accuracy in new writing, in favour of the ID feedback. The factor of \pm RFD did not exhibit a role in grammatical accuracy in new writing.

6.5 Results for Lexical Accuracy in New Writing

6.5.1 The effect of group conditions on lexical accuracy in new writing

Table 6.16 includes the descriptive results for lexical accuracy, and the means were illustrated in Figure 6.10. Based on the graph, from Task 1/2 to Task 4 the five groups showed some difference in the way their lexical accuracy changed over time.

Table 6.16 Means and standard deviations for first-draft lexical accuracy by task and group condition

Group	Task 1/2	Task 3	Task 4	Task 5
ID+RFD (<i>n</i> = 15)	2.80 (1.28)	2.91 (1.22)	3.08 (1.07)	2.82 (.93)
ID-RFD (<i>n</i> = 16)	2.81 (1.18)	2.87 (1.62)	2.93 (1.59)	2.72 (0.90)
IS+RFD (<i>n</i> = 15)	2.86 (0.82)	2.72 (1.08)	2.96 (1.76)	2.72 (0.99)
IS-RFD (<i>n</i> = 14)	2.68 (1.20)	2.93 (1.13)	2.90 (1.45)	2.79 (0.89)
Control (<i>n</i> = 17)	2.81 (0.67)	3.01 (1.06)	2.98 (1.05)	2.85 (0.91)

Note. Standard deviations are put in parentheses.

Figure 6.10 Mean first-draft lexical accuracy scores over Tasks 1/2–5 by group condition

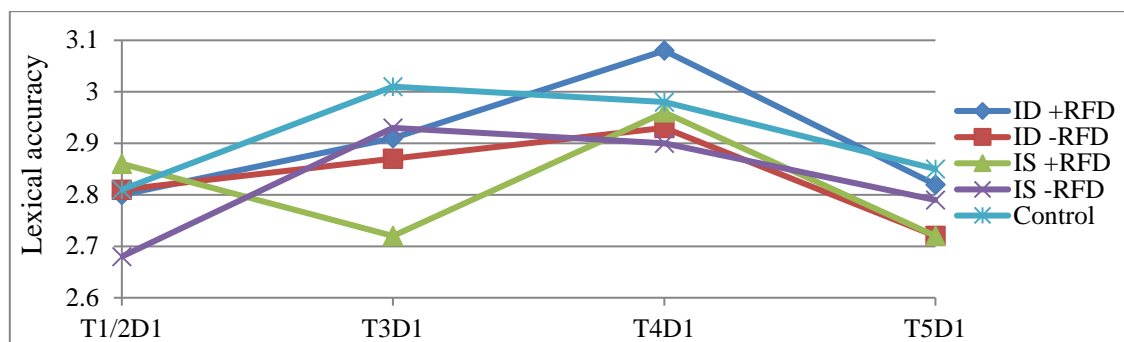


Table 6.17 summarises the effects of task and group condition on lexical accuracy in new writing¹³. There was no main effect of task or group condition, and the interaction effect was not significant. Between-subjects pairwise comparisons revealed that the five groups produced in general similar first-draft lexical accuracy in any of the tasks, and within-subjects pairwise comparisons found no significant difference for any group condition.

Table 6.17 Effects of task and group condition on first-draft lexical accuracy in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	.65	.565	.009
group	.05	.995	.003
task \times group	.09	1.000	.005

Note. The *df* was 2.69 for task, 4 for group condition, and 10.75 for the two-way interaction.

6.5.2 The effects of WF type and \pm RFD on lexical accuracy in new writing

Table 6.18 records the descriptive results for lexical accuracy in Tasks 1/2–5 by the factors of WF type and \pm RFD. According to Figure 6.11 which illustrates the means by WF type, students receiving either type of feedback followed a similar pattern of accuracy change over tasks. With regard to the mean lexical accuracy scores of the +RFD and the –RFD students, Figure 6.12 shows some difference in the way their accuracy scores changed over time.

¹³ Box’s test and Levene’s test showed no violation of the assumptions. Mauchly’s test showed a violation ($p = .000$) and the corresponding Greenhouse-Geisser estimate was .818 (i.e., greater than .75), so the Huynh-Feldt correction was used to report the effect of task and the interaction effect of task and condition.

Table 6.18 Means and standard deviations for first-draft lexical accuracy by task, WF type, and \pm RFD

	Task 1/2	Task 3	Task 4	Task 5
ID	2.81	2.88	3.01	2.77
(<i>n</i> = 31)	(1.21)	(1.42)	(1.34)	(0.90)
IS	2.77	2.82	2.93	2.75
(<i>n</i> = 29)	(1.00)	(1.09)	(1.59)	(0.93)
+RFD	2.83	2.81	3.02	2.77
(<i>n</i> = 30)	(1.06)	(1.14)	(1.43)	(0.95)
-RFD	2.75	2.90	2.92	2.75
(<i>n</i> = 30)	(1.17)	(1.39)	(1.50)	(0.88)

Note. Standard deviations are put in parentheses.

Figure 6.11 Mean first-draft lexical accuracy scores over Tasks 1/2–5 by WF type

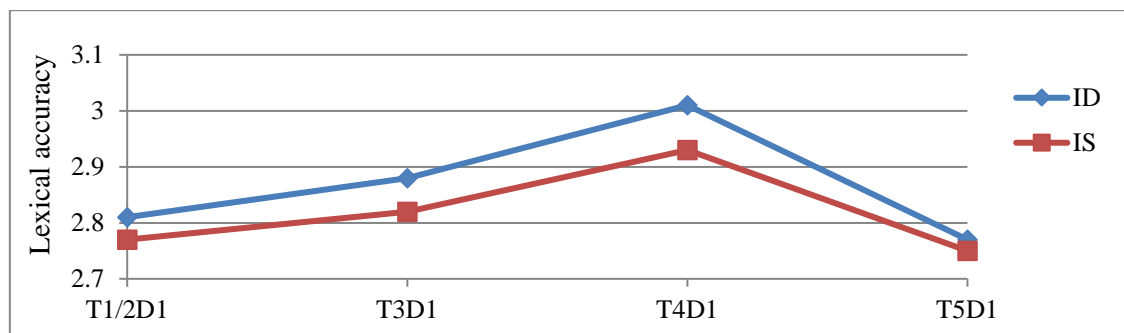


Figure 6.12 Mean first-draft lexical accuracy scores over Tasks 1/2–5 by \pm RFD

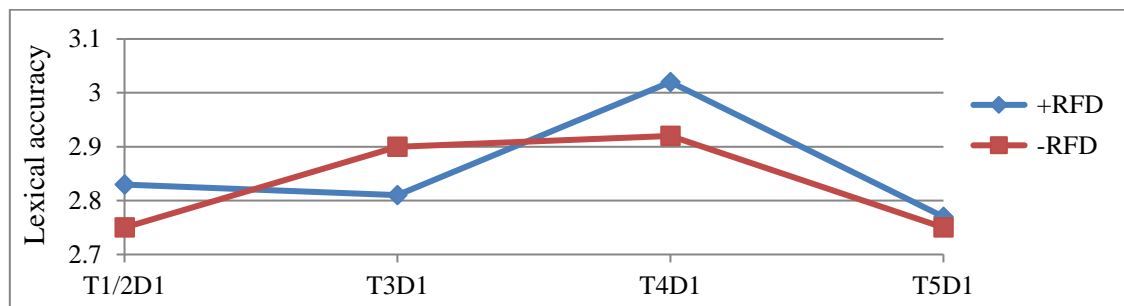


Table 6.19 displays the output of the ANOVA test¹⁴. It can be seen that there was no significant effect of task, WF type, or \pm RFD, and there was no significant interaction effects. Within-subjects pairwise comparisons for WF type and \pm RFD found no significant contrast, and between-subjects pairwise comparisons observed no significant difference in any of the tasks with respect to feedback type or \pm RFD.

Table 6.19 Effects of task, WF type, and \pm RFD on first-draft lexical accuracy in Tasks 1/2–5

	<i>F</i>	<i>p</i>	partial η^2
task	.51	.656	.009
WFT	.05	.829	.001
\pm RFD	.02	.893	.000
task \times WFT	.01	.997	.000
task \times \pm RFD	.12	.936	.002
WFT \times \pm RFD	.03	.862	.001
task \times WFT \times \pm RFD	.13	.925	.002

Note. WFT = WF type. The *df* was 2.68 for task, the two-way interaction between task and WFT, the interaction between task and \pm RFD, and the three-way interaction. The *df* was 1 for WF type, \pm RFD, and their interaction effect.

6.5.3 Summary of results for lexical accuracy

No significant within-subjects or between-subjects difference was found concerning the effect of group conditions on student lexical accuracy in new writing. Similarly, no significant within-subjects or between-subjects difference was shown concerning the effects of WF type and \pm RFD.

¹⁴ Box's test and Levene's test showed no violation of the assumptions. Mauchly's test showed a violation ($p = .000$) and the corresponding Greenhouse-Geisser estimate was .810 (i.e., greater than .75), so the Huynh-Feldt correction was used to report the effect of task and its interaction with WF type and \pm RFD.

6.6 Summary of Results Regarding the Effect of Treatment on Text Quality in New Writing

This section summarises the statistically significant results reported from section 6.2 to section 6.5, covering the results for all the four textual levels. It will start with comparing the effects of group conditions on first-draft text quality in a new task (see section 6.6.1) and next present the significant results focusing on the factors of WF type and \pm RFD (see section 6.6.2). Since no significant between-subjects differences were detected for any textual level, only within-subjects contrasts will be listed in each section. In the meantime, the mediation of textual level will be addressed.

6.6.1 The effect of group conditions and the mediation of textual level

According to Table 6.20 that summarises the significant within-subjects contrasts of the five groups, a distinct indication is that the group conditions differed from one another in their effects and that their effects were mediated by textual level. In general, the ID+RFD condition was the most effective as it helped students to improve text quality in new writing with respect to three textual levels. The ID-RFD and the IS+RFD conditions were only effective for organisation quality improvement over time, and the former seemed to need repeated treatment to take effect. The IS-RFD and the control conditions were the least effective, not bringing text quality improvement on any textual level.

From another perspective, organisation quality seemed to be more amenable to treatment than the other textual levels. Three of the treatment conditions enabled students to produce better organisation in new writing, and two of them took effect after the corresponding treatment was provided only once. By contrast, lexical accuracy seemed not subject to treatment; none of the treatment conditions rendered its increase in new writing.

Table 6.20 Significant within-subjects contrasts of the group conditions regarding first-draft text quality across Tasks 1/2–5

Group	Content quality	Organisation quality	Grammatical accuracy	Lexical accuracy
ID+RFD	*T1/2 vs. T5 (0.89)	*T1/2 vs. T3 (0.83) **T1/2 vs. T4 (1.32) ***T1/2 vs. T5 (1.23)	*T1/2 vs. T5 (0.55)	
ID–RFD		**T1/2 vs. T5 (1.01)		
IS+RFD		*T1/2 vs. T3 (0.69) *T1/2 vs. T4 (0.80) ***T1/2 vs. T5 (1.05)		
IS–RFD				
Control				

Note. Effect sizes (d) are put in parentheses. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

6.6.2 The effects of WF type and \pm RFD and the mediation of textual level

Tables 6.21 and 6.22 displayed the significant within-subjects contrasts in first-draft text quality with respect to feedback types and \pm RFD conditions. As no significant interaction effect was observed between WF type and \pm RFD, results for the two factors will be summarised in sequence.

With respect to the role of WF type, ID seemed more effective because it enabled students to raise both content and organisation quality in new writing while IS was only effective for organisation quality improvement. On the other hand, neither feedback method helped student to increase grammatical or lexical accuracy in new writing. In terms of the mediation of textual level, discourse levels, organisation quality in particular, were more amenable, while linguistic textual levels were not responsive to either type of feedback.

Table 6.21 Significant within-subjects contrasts concerning WF types regarding first-draft text quality across tasks 1/2–5

	ID	IS
Content quality	*T1/2 vs. T4 $d = 0.64$	
	**T1/2 vs. T5 $d = 0.69$	
Organisation quality	***T1/2 vs. T4 $d = 1.01$	*T1/2 vs. T3 $d = 0.49$
	***T1/2 vs. T5 $d = 1.13$	**T1/2 vs. T4 $d = 0.71$
		***T1/2 vs. T5 $d = 0.89$
Grammatical accuracy		
Lexical accuracy		

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

Table 6.22 Significant within-subjects contrasts concerning \pm RFD conditions regarding first-draft text quality across Tasks 1/2–5

	+RFD	-RFD
Content quality	**T1/2 vs. T5 $d = 0.76$	
Organisation quality	**T1/2 vs. T3 $d = 0.77$	*T1/2 vs. T4 $d = 0.69$
	***T1/2 vs. T4 $d = 1.03$	***T1/2 vs. T5 $d = 0.86$
	***T1/2 vs. T5 $d = 1.15$	
Grammatical accuracy		
Lexical accuracy		

Note. The blank cells indicate the absence of applicable results.

*** $p < .001$

** $p < .01$

* $p < .05$

Of \pm RFD conditions, the +RFD condition had an advantage in improving text quality in new writing on discourse levels when the relevant statistical data and effect sizes were compared. However, both conditions were ineffective for linguistic accuracy increase over time. Similar to the result in relation to the mediation of textual level on the effect of WF type, organisation quality was easier for students to be improved no matter whether they receive RFD or not, while linguistic accuracy was not responsive to either of \pm RFD conditions.

CHAPTER SEVEN

DISCUSSION

7.1 Overview

This chapter will discuss the results reported in the preceding three chapters. It will start with discussing the results concerning the immediate effect of treatment on revision. It will next discuss the results concerning the treatment effect on autonomous revision and finally address the results regarding the treatment effect on text quality in new writing. Throughout the discussion significant and interesting results will be highlighted, and the results will be compared with those from previous studies. Explanations for the results will also be provided by referring to theoretical propositions and relevant evidence in existing studies. For the discussion of the immediate effect of treatment on revision, the outcome of supplementary text analysis (see section 3.9.4.3) will be presented as evidence for interpreting the results.

In the following discussion “diagnostic feedback” and “prescriptive feedback” are sometimes used interchangeably with the terms ID and IS, respectively, in order to provide a direct sense of the nature of the two feedback types. In section 7.2 where the immediate effect of treatment on student revision is discussed, “the average effect size” of a certain treatment condition is used to refer to the average of its effect sizes for the two or three cases of significant improvement it brought during the three tasks in Stage 2.

7.2 The Immediate Effect of Treatment on Student Revision

With respect to the first set of research questions which addressed the immediate effect of treatment on student revision, effects of specific treatment methods and effects of the factors of WF type and \pm RFD were obtained. It was also found that those effects were subject to textual levels involved.

7.2.1 Overall effects of treatment methods in comparison to the control condition

Overall, each of the treatment methods (ID+RFD, ID-RFD, IS+RFD, IS-RFD) was more effective than the control condition in helping participants improve text quality in a revised draft on both discourse and linguistic levels. Given that feedback was a

component of each treatment method, the effectiveness of all the treatment methods showed that WF helped students identify the existence of the problems in their text and make corresponding improvement. The result is expected because provision of either diagnostic or prescriptive information facilitated students to reflect on their argument, the macro structure and connectedness, as well as the linguistic features in their writing. In comparison, students receiving no such feedback had difficulty in making equally effective improvement. They either lacked the necessary knowledge to modify their text problems or failed to activate the knowledge to implement the revising process (Adams et al., 2010; Butterfield et al., 1996). The following discussion will start with the treatment effect on revision on discourse levels and then present the treatment effect on revision on linguistic levels.

7.2.2 The effect of treatment for text quality on discourse levels

7.2.2.1 Overall effectiveness of treatment methods for both content and organisation

An important finding was that all the treatment methods enabled the participants to significantly improve their text in content and organisation, whereas the control condition did not render such an effect. This result differed from some previous findings which showed little effect of WF on content quality¹ (Ashwell, 2000; Fathman & Whalley, 1990; Lee, 1997; Sampson, 2012). This contradiction is probably due to the distinct features of content feedback provided in these studies. For example, the comment on content in Ashwell's (2000) study was generally vague other than being concrete, while Lee (1997) and Sampson (2012) only used an oversimplified form of code to indicate meaning problems. By contrast, content feedback in this study was more specific and usable, considering the information of expectations and rating criteria provided in the diagnostic feedback and the specific suggestions for operation in the prescriptive feedback. It can therefore be suggested that the effect of content feedback is contingent on the nature of the information provided within, such as whether it is understandable, informative, and applicable (Ferris, 1997; Jonsson, 2012). Although it has been proposed that addressing content problems is cognitively more demanding for students than correcting language errors because it entails evaluation beyond single sentences (Lee, 1997), specific comment or finely designed codes seem to enable

¹ Ashwell (2000) and Fathman and Whalley (1990) used the term "content" to include quality of organisation as well as quality of meaning.

students to overcome the difficulty and make successful revisions (Conrad & Goldstein, 1999; Ferris, 1997). Morra and Asis (2009) observed that EFL learners significantly reduced both linguistic errors and problems in content and organisation after receiving feedback with finely developed codes aimed at specific problems. In addition, the consistently poor revision result of the control group for content quality over Stage 2 did not support the conjecture that students could produce better content simply by rewriting (Ashwell, 2000), or at least, it was not the case for the EFL students in my study.

It is even possible that the discourse-level feedback in this study altered the task schema (Hayes, 1996; termed as “task definition” by Flower et al., 1986; for a review, see sections 2.4.2.3–2.4.2.4) of the participants by raising their awareness to the goal of rediscovering meaning (Murray, 1978; Wallace et al., 1996) in a revision task. In other words, it is likely that the feedback on content and organisation in this study took effect not only because it provided useful information related to discourse problems, but because it modified the participants’ perception of the revision task and activated their internal cognitive resources to process the aspects of argument and organisation. It has been observed that novice writers, especially L2 learners, tend to perceive revision as a surface-error correction task and not to recognise the need for reviewing content and structure (Barkaoui, 2016; Butler & Britt, 2011; Porte, 1997). Such a flawed task schema might also explain the ineffective revision on discourse levels by the control group in this study.

Another reason for the effectiveness of the feedback for content and organisation in this study might be its clear focus and its close correspondence to the rating criteria (for examples of diagnostic and prescriptive feedback on discourse levels, see Tables 3.9 and 3.10). First, the targeted problems were, to some extent, focused rather than being diffuse. As explained in the methodology chapter, the targeted problems regarding content fell into two broad categories (relevance to the topic and effective argumentation), and those regarding organisation also fell under two categories (macrostructure and connectedness). The focus of the feedback would lighten students’ attentional load and thus increase the chance of successful revisions. Second, the elements targeted by feedback corresponded to the criteria I used to measure discourse quality. For example, problems related to sufficient support and considerations for

counterargument that were addressed by feedback were also major components to determine the content quality of student writing. Therefore, the successful revisions based on the feedback might naturally lead to higher scores for the second draft.

It is also possible that repeated provision of such focused feedback over the three tasks in Stage 2 drew students' attention to the targeted aspects, and more importantly, consolidated those focused aspects in their working memory and even facilitated the intake into long-term memory (Butterfield et al., 1996; Hayes, 2004, 2012).

Apart from the above-mentioned result that all the treatment conditions brought significant revision effects on the discourse levels, closer inspections showed that the effect sizes of the treatment methods, taken together, showed a gradual upward trend over the treatment stage (see Table 4.26). This indicated that both feedback methods produced a consistent and accumulative effect in general. As most previous experimental studies included only one writing-revising task to assess the effect of feedback on revision, this finding is particularly important. It helps to confirm the effectiveness of teacher WF in EFL students' revision on discourse levels. From another perspective, an implication of this finding is that WF may have to be provided multiple times before an evident effect appears. Therefore, it may be hasty to refute the efficacy of WF if it is not obvious after a one-shot treatment.

7.2.2.2 Treatment being more effective for organisation

Although the treatment methods were effective for both content and organisation, organisation seemed to be somewhat more amenable to treatment as evidenced by the statistical results concerning the immediate effect of treatment on discourse-level improvement (see Table 4.26). For any of the four treatment methods, the average effect size for between-draft organisation quality improvement in Stage 2 tasks was somewhat larger than that for between-draft content quality improvement², d being 1.07 vs. 0.79 for the ID+RFD condition, 0.61 vs. 0.46 for ID-RFD, and 0.86 vs. 0.70 for IS+RFD.

² The IS-RFD condition brought significant improvement in content quality in two of the three Stage 2 tasks with an average medium effect size ($d = 0.55$); it brought two cases of significant improvement and one case of marginally significant improvement in organisation in Stage 2 tasks, the average effect size being medium ($d = 0.46$).

The larger effect for organisation quality might be because organisation was easier to improve, especially the aspect of macrostructure. Examination of the collected first drafts revealed some common problems in macrostructure, i.e., a lack of an introduction or a conclusion paragraph, and a failure to paragraph the body part. Interestingly, the most common changes in organisation in the revised drafts were found to be an addition of an introduction or a conclusion and improvement in paragraphing. It seemed that remedial actions on macrostructure were not difficult for the students once they received specific feedback on that aspect. It was also found that the students in the treatment conditions often made effective changes in cohesion, mainly additions of signal words, such as “first”, “in addition”, “in other words”, and “however”, to specify the links among sentences and ideas. It seemed that it was relatively easy for organisation quality to benefit from relevant feedback.

Similar findings have been reported by Conrad and Goldstein (1999), who observed that, compared with problems regarding coherence and cohesion, problems regarding explanation, analysis, and explicitness of argument were revised less effectively irrespective of feedback type. This is probably because some changes in organisation entail a rearrangement of existent concepts and ideas in a text while changes in content (or argumentation) involve production or removal of ideas and evidence. The activity of generating ideas may entail a higher level of cognitive effort (Faigley & Witte, 1981; Roca de Larios, Manchón, Murphy & Marín, 2008). In addition, it has been postulated that successful formulation of ideas largely depends on a writer’s long-term memory where information is retrieved (Hayes, 2012; Ong, 2013). Therefore, it would be difficult for a reviser to improve on the content level if relevant information is not available in his/her long-term memory.

7.2.2.3 The relative effects of the treatment methods

When the four treatment methods were compared with respect to the significant improvement they brought in discourse-level text quality, it was observed that the ID+RFD and the IS+RFD conditions were more effective than the ID–RFD and the IS–RFD conditions whether in the case for content or for organisation (for reference, see Table 4.26). Those differences are accounted for when we look at the effect of WF type and \pm RFD because they are the elements that constructed the treatment conditions.

7.2.2.4 The effect of treatment in relation to WF type

It was surprising that generally no differences were detected in the effects of the two feedback types on discourse-level text quality across drafts, irrespective of \pm RFD. Both feedback methods were effective for between-draft content quality improvement on a medium magnitude; the average effect size was respectively 0.61 and 0.60 for the diagnostic and the prescriptive feedback (see Table 4.28). For organisation quality improvement, the diagnostic feedback showed an advantage with an average large effect size ($d = 0.83$), compared to the average medium effect size ($d = 0.66$) of the prescriptive feedback (see Table 4.28).

The similar effectiveness of the two feedback methods in this study might be because both methods conformed to the reported features of useful feedback. Since few studies have employed an experimental design to compare the effect of descriptive feedback and prescriptive feedback on revision for discourse levels, as operationalised in this study, it is hard to discuss the results with specific reference to previous research. Most relevant studies approached this query by relating student revisions to the features of authentic teacher comment (e.g., Cho & MacArthur, 2010; Ferris, 1997). The findings have indicated the usefulness of explicit, direct comments and provision of understandable strategies to solve problems in comparison to mere indication of problems. Both the ID and the IS feedback in this study were explicit, with a similar length of approximately 13 words (if translated into English) for each comment item (see Tables 3.9 and 3.10). Moreover, both comment types were direct, stating the criteria or offering suggestions, and both were fairly understandable, using Chinese for technical terms.

Their difference lay in that the ID feedback offered general guidelines without relating them to student text while the IS feedback offered specific operations on student text but did not explain the relatively universal rationale or criteria behind. For example, to address the problem of lacking support to a proposed idea, the diagnostic feedback would be: Each proposed idea should be elaborated by examples, reasoning, or statistics to make it persuasive. By contrast, the corresponding prescriptive feedback would read: Add something to support this idea. You may take Tang poems as an example. Although direct corrections for linguistic errors tend to be most effective for student revision (Chandler, 2003; Van Beuningen et al, 2012), the effect of prescriptive comments on

discourse levels seems not so straightforward. Some studies found that students did not adopt teacher suggested solutions for content and structure because they perceived that teachers misunderstood them or exerted intrusion on their original idea development or that the suggestions were confusing due to a lack of explanations (Ferguson, 2011; Hyland, 2000; Poulos & Mahony, 2008; Taylor, 2011; Weaver, 2006).

Supplementary text analysis on the discrete revisions students made revealed that the IS feedback was less responded to than the ID feedback (77% vs. 68%), and students receiving ID made more feedback-related revisions (73% vs. 65%), as shown in Table 7.1. Therefore, the diagnostic feedback contributed to slightly more revisions than what the prescriptive feedback did. These data reflected that the prescriptive comments in this study were not fully adopted by the students. They also indicated that although the diagnostic feedback only offered explanations (i.e., criteria) for a marked problem without linking the problem to the criteria explicitly, university students might have the analytical ability to build the link and make corresponding changes. Since this study did not include a third type of feedback offering both diagnosis and solution, it is hard to determine if the third type of feedback yields the best result on revision. It would have been good to interview the students about the usefulness of these different types of feedback to them.

Table 7.1 Revisions on discourse levels for Task 3 by WF type

WF type	Total of revisions	Total of revisions due to comments	Total of comments	Total of comments acted upon
ID (<i>n</i> = 31)	178	130 (73%)	137	105 (77%)
IS (<i>n</i> = 29)	157	102 (65%)	121	82 (68%)

7.2.2.5 The effect of treatment in relation to \pm RFD

It was found that the factor of \pm RFD played a role in between-draft text quality change on the discourse levels, irrespective of WF type. The +RFD condition rendered an overall large effect for content and organisation improvement across drafts with the average effect size of 0.75 and 0.95, respectively. The corresponding effectiveness of the -RFD condition was on a medium magnitude with the effect size of 0.46 and 0.54

for content and organisation, respectively (for reference, see Table 4.28). Moreover, for the second draft of Task 3 students in the +RFD condition significantly outperformed those in the –RFD condition in both content and organisation quality. In other words, the WF was more effective for content and organisation improvement when students were directed to focus on discourse-level and linguistic-level revision in sequence.

This result suggested that the +RFD condition enabled students to focus on revising content and organisation and process the corresponding feedback more deeply. When processing discourse levels of text, the +RFD students did not need to distribute their limited attentional resources to linguistic aspects of their text. In this way, the amount of information to be computed was reduced than if they had to process language and meaning simultaneously; thus the +RFD condition was less cognitively demanding (Robinson, 2005, 2011; Skehan, 1998, 2015). This result was also in line with earlier research findings which indicated that idea development and idea transcription during the writing process competed for resources in working memory (Glynn et al., 1982; Hayes, 2012; Roca de Larios et al., 2008).

To corroborate this conjecture, extra text analysis was conducted to compute the number of discourse-/linguistic-level changes made by students according to their ±RFD condition. Data presented in Table 7.2 indicated that the +RFD condition triggered more changes on discourse levels than the –RFD condition (187 vs. 148) and more response to comments on discourse (85% vs. 59%). In addition, a larger proportion of the discourse-level changes by the +RFD students were attributable to comments (73% vs. 64%).

Table 7.2 Revisions on discourse levels for Task 3 by ±RFD

±RFD condition	Total of revisions	Total of revisions due to comments	Total of comments	Total of comments acted upon
+RFD (<i>n</i> = 30)	187	137 (73%)	130	111 (85%)
–RFD (<i>n</i> = 30)	148	95 (64%)	128	76 (59%)

However, this result in favour of the +RFD condition seemed to contradict findings of Fathman and Whalley (1990) and Ashwell (2000), who observed that giving content and

corrective feedback simultaneously did not differ from the “content feedback only” condition in the effect on content quality improvement. This incongruence is probably due to different research designs. Although the early two studies manipulated the focus of feedback, they did not actually manipulate student focus of revision. Therefore, students in the “content feedback only” condition might not devote focused attention to improving content, and consequently, the condition did not render a clear advantage therein. By contrast, this study manipulated the factor of revision focus by explicitly asking participants to focus on discourse levels first.

Another explanation for the incongruent research results may be the different methods of data analysis. This study detected a role of the \pm RFD factor by examining the index of effect size (Cohen’s *d*) and looking at between-subjects differences in addition to within-subjects differences. However, the early two studies only looked at within-subjects differences without reporting relevant effect sizes.

7.2.3 The effects of treatment methods for text quality on linguistic levels

7.2.3.1 Overall effectiveness of treatment for grammatical accuracy and overall ineffectiveness of treatment for lexical accuracy

Overall, the treatment methods were effective for grammatical accuracy increase but less effective for lexical accuracy increase in the revised text. For grammatical accuracy, the ID–RFD and the IS–RFD groups each made significant improvement in all the three tasks of Stage 2, and the ID+RFD and the IS+RFD groups each made significant increase in two of the three tasks³. In contrast, the control group significantly increased accuracy in only one of the tasks. With respect to lexical accuracy, the control and the ID+RFD groups made no significant increase in any of the three tasks while the other three treatment groups each made significant increase in only one of the tasks.

These results corroborated the propositions that grammatical and lexical errors react to corrective feedback differently (Ferris, 1999, 2006, 2010; Truscott, 1996, 2001). Specifically, they corresponded with Ferris’s argument that many grammatical features are “treatable” by corrective feedback and refuted Truscott’s assertion that corrective

³ For one of the two tasks where the IS+RFD group made significant improvement, the improvement approached statistical significance only ($p = .055$ for Task 2).

feedback could only work for discrete errors instead of grammatical features that belong to a complex system. According to Ferris (1999), corrective feedback, particularly metalinguistic feedback, is “only applicable when students’ errors occur in a patterned, rule-governed way” (p. 6) because in that case students can consult and apply rules. She further argues that no rules exist for discrete problems with word choice and sentence structure, so feedback that merely points out the problem is of little help. Her analysis may explain the results in favour of increases in grammatical accuracy across drafts (for a synthesis, see Ferris & Hedgcock, 2005).

In relation to the previous empirical research, as none of the error categorisation methods was identical to the method in this study, the results were not absolutely comparable. This study put problems related to a wide range of grammatical features into the category of grammatical errors and put word/phrase choice problems into lexical errors (see Table 3.8). The results seemed to correspond with those of Ferris and Roberts (2001), who found that three types of grammatical errors (verb, noun, article) were reduced to a larger extent than word choice in the revised text. The results differed from those of Van Beuningen et al. (2012), who observed positive effects of feedback for both grammatical and nongrammatical errors. It is notable, though, that Van Beuningen et al. lumped together word choice, spelling, punctuation, and capitalisation under the same category; their result was therefore not strong evidence for the amenability of lexical errors to feedback.

7.2.3.2 The different effects of the treatment methods

When the significant between-draft differences of the four treatment conditions were compared with respect to grammatical accuracy increase, an order was found in their effectiveness (for reference, see Table 4.26). The IS–RFD condition was the most effective, closely followed by the ID–RFD condition. These two conditions brought significant increase in any of the three tasks in Stage 2, and the relevant effect sizes were consistently large ($d = 1.14$ for IS–RFD and 0.99 for ID–RFD on the average). The ID+RFD and the IS+RFD conditions were less effective; they rendered significant increase in two of the tasks. Between them there was still a difference in effect sizes ($d = 0.72$ vs. 0.46 on the average), in favour of the ID+RFD condition. Such differences between the four treatment methods in their effect on grammatical increase across drafts

can be interpreted if we examine the factors of WF type and \pm RFD, which constructed the treatment conditions.

7.2.3.3 The effect of treatment in relation to WF type

The effects of the two feedback methods were similar irrespective of error type (grammatical or lexical). Statistical tests on WF type showed that both the ID and the IS feedback rendered significant between-draft grammatical accuracy increase in all the tasks of Stage 2, and that both rendered significant lexical accuracy improvement for two of the tasks. Table 7.3 lists the data related to linguistic-level revisions made by students with different types of feedback (noted as CF for corrective feedback). It can be seen that students receiving ID and students receiving IS made a similar number of revisions, and a similar percentage of the revisions they made were due to feedback (70% vs. 67%). The respective feedback points were acted upon to a similar extent (66% vs. 70%).

Table 7.3 Revisions on linguistic levels for Task 3 by WF type

WF type	Total of revisions	Total of revisions due to CF	Total of CF points	Total of CF points acted upon
ID ($n = 31$)	283	197 (70%)	270	177 (66%)
IS ($n = 29$)	301	209 (67%)	261	184 (70%)

It is somewhat surprising that IS (i.e., direct correction) did not differ from ID (i.e., metalinguistic explanations) in the effect on linguistic error reduction across drafts. There has been evidence of an advantage of direct correction over indirect feedback, be it underlining, codes, or underlining plus codes (Chandler, 2003; Frear, 2012). Such existent evidence has been expected because students can adopt the provided forms directly in their revision (Chandler, 2003; Frear, 2012; Van Beuningen et al., 2012). However, no comparison has been reported between the effects of direct correction and metalinguistic explanations on written accuracy of a wide array of linguistic features.

7.2.3.3.1 Relative effects of the two feedback methods on grammatical accuracy across drafts

The equal effects of the ID and the IS feedback methods on grammatical accuracy across drafts probably resulted from the nature of the metalinguistic information provided in ID. On one hand, the metalinguistic explanations, like mini-instruction, served the purpose of instruction in the way that Truscott (1999) had argued for treating complex grammatical errors. On the other hand, the explanations were more explicit and easy to understand for L2 students than alphabetic codes that represent grammatical terms (Ellis, 2009; Lee, 1997). The terms mentioned in the explanations were written in Chinese instead of English, demanding little extra effort to decipher. Then the students could apply the rules and made successful corrections (Ferris, 1999). Besides, for each original text a maximum of 10 linguistic errors were addressed and repeated errors were prioritised. As a result, the diagnostic feedback would not be overwhelming as was the coded feedback in Chandler's (2003) study that treated every single error.

Another reason for the benefit of the ID feedback comparable to IS might be that the participants had received systematic instruction on English grammar. As Chinese high schools teach basic grammatical features in accordance with *The Senior High School English Syllabus* (MoE, 2000), the participants were expected to have prior grammatical knowledge and consequently could benefit from the metalinguistic feedback.

7.2.3.3.2 Relative effects of the two feedback methods on lexical accuracy across drafts

Both feedback methods enabled students to significantly increase lexical accuracy across drafts in two of the three tasks in Stage 2 with an average small to medium effect size. This result is somewhat surprising because the above-mentioned merits of metalinguistic explanations for treating grammatical errors in participants' writing did not exist in the diagnostic feedback for lexical errors. As introduced in section 3.4.1, the diagnostic feedback I provided for inappropriate lexical choice was generally "wrong word choice", "improper phrase", "ambiguous", or "meaning unclear". Such diagnostic feedback was supposed to prompt students to change a word or phrase but was not expected to be as effective as direct correction for treating lexical choice problems. This result was not in accord with Ferris's (1999) proposition that direct correction works better for treating lexical choice problems which are discrete.

A possible explanation may be that students receiving the cue of improper wording would try to improve the wording by rewording or paraphrasing (cf. Diab, 2015). Sometimes their corrections were not ideal but did improve the original expression. The following are a few examples of lexical changes students made in response to diagnostic feedback.

Ex. 1

improper word choice

Draft 1: ... *just strengthen their concept that they can reap without sowing.*

Draft 2: ... *just strengthen their misconception that they can reap without sowing.*

Ex. 2

wrong phrase

Draft 1: *We should view this phenomenon in a different degree.*

Draft 2: *We should view this phenomenon in a different angle.*

Ex. 3

wrong word

Draft 1: *I don't know why some people think so. I just know they are fault.*

Draft 2: *I don't know why some people think so. I just know they are wrong.*

Ex. 4

improper word choice

Draft 1: *The ancient Chinese language is part of Chinese literacy.*

Draft 2: *The ancient Chinese language is part of Chinese literature.*

7.2.3.4 The effect of treatment in relation to \pm RFD

It was found that the factor of \pm RFD played a role in between-draft text quality change on linguistic levels, mainly in the change in grammatical accuracy, irrespective of WF type. With respect to grammatical accuracy increase, the $-$ RFD condition rendered a large effect size for all the three tasks (on average $d = 1.07$); the $+$ RFD condition produced a medium effect size for all the three tasks (on average $d = 0.52$; for reference, see Table 4.28). In addition, although the $-$ RFD and the $+$ RFD students did not differ in

grammatical accuracy for the first draft of Task 3, for the second draft the –RFD students outperformed the +RFD students on a medium magnitude ($d = 0.56$; see Table 4.29). Regarding lexical accuracy, the –RFD condition enabled students to make significant between-draft improvement for all the three tasks with an average small to medium effect size ($d = 0.41$) while the +RFD condition only rendered significant between-draft improvement for two tasks with an average small effect size ($d = 0.31$; for reference, see Table 4.28). These results suggested that the feedback was more effective for linguistic accuracy improvement when students were not directed to focus on discourse- and linguistic-level revision in sequence.

It was surprising that focusing on different text levels separately benefited revision on discourse levels (see section 7.2.2.5) but weakened the result for linguistic accuracy. A major reason might be that the +RFD students produced more discourse-level changes than the –RFD students (187 vs. 148; see Table 7.2), such as additions of a conclusion and examples, wherein they made new linguistic errors, and that these errors did not have the chance of receiving external feedback during Stage 2.

From another perspective, when the participants made discourse-level changes some corrective feedback points might be overlooked because the text that carried the feedback might have been replaced or deleted. As Table 7.4 shows, in Task 3 more corrective feedback points were acted upon in the –RFD condition than that in the +RFD condition (78% vs. 61%). In addition, the students might lack the ability to transfer the potential benefit of such “unusable” feedback into the monitoring of the unmarked part of their text. As a result, the potential effect of the corrective feedback was diluted in the +RFD condition. Even though the +RFD students focused on accuracy in the second half period of the revising session, they could not make so much use of corrective feedback as those –RFD students.

Table 7.4 Revisions on linguistic levels for Task 3 by \pm RFD

\pm RFD condition	Total of revisions	Total of revisions due to CF	Total of CF points	Total of CF points acted upon
+RFD ($n = 30$)	280	183 (65%)	269	165 (61%)
-RFD ($n = 30$)	304	227 (75%)	264	206 (78%)

The finding of the advantage of the -RFD condition indicated the students' intrinsic primary concern for linguistic accuracy during revision. Such a tendency has been reported with L2 English learners as well as inexperienced native revisers (e.g., Ashwell, 2000; Barkaoui, 2016; Sommers, 1982). Therefore, without being directed to focus on content and organisation first, the students might spontaneously invest more attention and effort in improving language and consequently make better use of the corrective feedback provided. In other words, during revision these students might also have a focus for cognitive resources investment or selective attention to task demands (Wickens, 2007). This focus did not result from external manipulation but from a preset priority to accuracy or an incomplete task schema for revision tasks (Butler & Britt, 2011; Hayes, 1996).

7.3 The Effect of Treatment on Student Autonomous Revision

The second set of research questions were intended to assess the effect of treatment on student autonomous revision. As mentioned in the literature review, many L1 and a few L2 studies experimented with interventions, mainly strategy instruction and prompts, to foster autonomous revision (alternatively termed as "revision skills" and "ability to revise"). However, very little research has looked at student autonomous revision in relation to WF methods or procedural guidance in terms of revision-focus direction. As an exception, Carifio et al. (2001) investigated the effects of written comments (containing both diagnostic and prescriptive information) on L1 students' autonomous revision of preset texts. They found that both the experimental and the control groups significantly improved their revising skills, but the former improved to a larger extent on effective major changes and macrostructure changes. As their measurement of treatment effect assessed discrete changes students made instead of text quality change

and their participants revised preset texts, their results were not readily comparable to this study. Therefore, the results concerning autonomous revision will be discussed with very limited reference to previous research.

In general, the inferential statistics found little treatment effect on autonomous revision. Following the pattern of discussion about the immediate treatment effect on revision, the following discussion will begin with the effect for discourse levels and then move on to the effect for linguistic levels.

7.3.1 The effect on autonomous revision concerning text quality on discourse levels

7.3.1.1 Overall effects of treatment methods

In terms of content and organisation quality, only the ID+RFD group produced some increase in the gain score from Stage 1 to Stage 3, though not significantly (see sections 5.2.3 and 5.3.3). The benefit of the ID+RFD condition will be discussed later when we look at the factors of WF type and \pm RFD. These results showed that, except the ID+RFD condition, the treatment methods failed to foster the students' revision skills in improving content and organisation.

Although Butler and Britt (2011) think that revision intervention is expected to generate a long-term effect if students can practise with the intervention for several times, the result is disappointing. It did not correspond with Carifio et al.'s (2001) positive findings. It was expected that the successively repeated treatment might have enabled students to internalise some information provided in the feedback, and that they could activate the internalised information to re-evaluate and improve their text when facing a new revision task. The ineffectiveness of the treatment methods (except ID+RFD) may be due to fatigue effect; the students might get bored of the tasks and would not work hard for the final task. Another possible explanation may be that three occasions of treatment were not sufficient.

7.3.1.2 The effect of treatment in relation to WF type

It was found that the factor of WF type did not make a difference in student autonomous revision on discourse levels. Neither feedback method rendered a significant increase in content or organisation gain scores from Stage 1 to Stage 3. It was expected that the

diagnostic feedback would be superior because its rating criteria-related information was assumed to be transferrable, but there was no such evidence. The ineffectiveness of either feedback method for autonomous revision will be discussed in section 7.3.3.

7.3.1.3 The effect of treatment in relation to \pm RFD

It was found that the factor of \pm RFD played a role in student autonomous revision for discourse levels, with an advantage for the +RFD condition. The +RFD students significantly increased their content gain score from Stage 1 to Stage 3 ($p = .038$, $d = 0.54$). Regarding organisation gain score, students in the two conditions produced similar gains in Stage 1, but the +RFD students made a larger gain than the -RFD students ($p = .053$, $d = 0.30$) in Stage 3.

The benefit of the +RFD condition is probably because during the treatment stage it allowed the students to process the feedback on content and organisation more deeply due to a reduced need to concurrently attend to feedback on discourse and language. As theories of attention (Robinson, 1995; Schmidt, 2001; Wickens, 2007) postulate, focused attention facilitates deep processing and leads to retention of material.

The repetition of the +RFD condition in Stage 2 might also contribute to its effect on student autonomous revision. The +RFD students might have perceived the revision-focus direction as a reasonable approach to a revision task or have unconsciously maintained the practice from the treatment stage to the autonomous revision task in the posttreatment stage. In other words, they might have internalised the revision-focus direction as part of their task schema (Hayes, 1996). Drawing on Butterfield et al.'s (1996) model, the influence of revision-focus direction on the students' working memory during the experiment might have been transferred to their long-term memory, either consciously or unconsciously. In consequence, they might have allocated more focused resources to revising content and organisation in the autonomous revision.

7.3.2 The effect on autonomous revision concerning linguistic accuracy

The effects of treatment methods on autonomous revision for linguistic accuracy were unclear. None of the five group conditions produced statistically significant effect concerning linguistic accuracy gains. Although the treatment methods helped students to increase accuracy, mainly grammatical accuracy, during the treatment stage, they did

not improve students' skills to increase accuracy when revising their new writing without the aid of feedback. A possible explanation may be fatigue effect; the students might get tired of the repeated revision sessions and would not put in as much effort toward the end of the study. Another reason may be related to the nature of the WF provided, which is to be elaborated in section 7.3.3.

7.3.2.1 The effect of treatment in relation to WF type and \pm RFD

The factor of WF type did not play a role in student autonomous revision regarding linguistic accuracy. Neither feedback type rendered any significant difference in gain scores across stages. The similar ineffectiveness of the two feedback methods indicated that neither the diagnostic nor the prescriptive information developed students' ability to detect and edit linguistic errors in a new revising task. This result was not very surprising because the feedback only targeted the errors appearing in student texts in Stage 2 while students might have made different errors for the writing topic of Stage 3. More importantly, the feedback did not include error-editing strategies, such as self-questioning and attending to one's own frequently made error categories. No effect was detected regarding the factor of \pm RFD, and neither +RFD nor -RFD gave rise to significant difference in the accuracy gain across stages, be it the grammatical accuracy gain or the lexical accuracy gain. Such a result might result from the above-mentioned ineffectiveness of the written feedback provided to the participants.

7.3.3 Ineffectiveness of feedback for autonomous revision

The results concerning the ineffectiveness of either the diagnostic or the prescriptive feedback for autonomous revision for any textual level suggested that WF to writing might not be conducive to developing students' revision skills. The WF in this study only addressed the local problems in students' specific drafts but did not contain guidelines regarding how to revise autonomously. Even though the diagnostic feedback contained some universal rules, it was related to good writing rather than effective revision strategies such as goal setting and procedures to carry out a revision task. In other words, the written feedback in this study might have directly affected students' working memory during the revising process but might have only indirectly affected students' long-term memory, in this case, cognition and metacognition related to revising strategies (Butterfield et al., 1996). WF was therefore found to be more

effective for assisting students in achieving the immediate goal of revising a text, as reported in Chapter Four (see sections 4.6 and 4.7), but not easily applicable to re-evaluation for another revising process.

It can therefore be assumed that for the purpose of fostering revision skills WF is not a promising tool, at least when it targets the written text alone or remains at the task level. WF to a specific draft may limit students' effort to improving the draft and impair the potential of acquiring revision strategies which can serve as students' internal feedback for future tasks. According to Hattie and Timperley's (2007) remark on feedback in a broad sense of education,

too much feedback only at the task level may encourage students to focus on the immediate goal and not the strategies to attain the goal. It can lead to more trial-and-error strategies and less cognitive effort to develop informal hypotheses about the relationship between the instructions, the feedback, and the intended learning. (p. 91)

Hence, writing teachers may consider giving feedback beyond student text, such as feedback about revision strategies and self-regulation (Hattie & Timperley, 2007). For example, written feedback may prompt students to reflect on the criteria for effective argumentation and check whether their writing has met those criteria. In addition, teachers may prompt students to reflect on the weaknesses in their written English and evaluate those aspects in their draft. Such feedback does not address specific problems but helps students to foster ability in self-evaluation.

In addition to written feedback, teachers can deploy other intervention methods such as revision strategy instructions (Sengupta, 2000; Song & Ferretti, 2013) and learning skills intervention (Hattie, Biggs & Purdie, 1996).

7.4 The Effect of Treatment on Text Quality in Student New Writing

The third set of research questions aimed to investigate the effect of treatment on text quality in student new writing. As reviewed in Chapter Two, very limited L1 literature touched upon the impact of WF on student writing development. In the L2 context, despite the burgeoning evidence on accuracy development in relation to CF, little research explored writing development regarding content and organisation in response to discourse-related feedback. The related literature will be interwoven into the

following discussion. In addition, the discussion sections will be structured by textual level.

7.4.1 The effect for content quality in new writing

Regarding content quality in new writing, only the ID+RFD group made a statistically significant improvement from Task 1/2 to Task 5, with a large effect size ($p = .014$, $d = 0.89$). This group's improvement may be explained by the advantage of the ID feedback and the +RFD condition.

7.4.1.1 The effect of WF type

Analysis of the factor of WF type showed that students receiving the ID feedback brought a significant improvement from Task 1/2 to Task 4 and from Task 1/2 to Task 5 with an average medium to large effect size ($d = 0.67$) while students receiving IS did not bring any statistical improvement. Parr and Timperley (2010) found that statement of writing aims or rating criteria (i.e., diagnostic information) and suggested actions to improve (i.e., prescriptive information) predicted student writing progress. However, they did not compare the relative effects of diagnostic and prescriptive elements in feedback. The result of this study indicated that diagnostic information was more effective than prescriptive information in helping students to produce better content in new writing.

One reason for the efficacy of ID may be that it engaged students in a more intense form of problem-solving. Students were pushed to link the diagnosis to their text, recognise the problem, and then exploit their own resources to generate the changes instead of adopting the solution provided. They had to make more cognitive endeavours to reduce the discrepancy between the current to the desired text features. This problem-solving process may be more motivational than following available solutions (Kluger & DeNisi, 1996), and the deep processing engaged may promote a larger learning potential (Wickens, 2007).

The benefit of ID may also result from its rating criteria-related feature. As introduced in section 3.4.1, when giving feedback on discourse levels I focused on problems in student writing that closely related to the elements in the criteria for rating argumentative essays. In terms of feedback on content, the feedback I had provided

centred around aspects such as a clear claim, sufficient evidence, and appropriate considerations for counterarguments, which were crucial elements in the criteria for rating the content students produced.

Song and Ferretti (2013) have found a positive effect of teaching rating criteria-based questions about argumentation on new writing in addition to the revising behaviour. They suggest that “standards for good writing acquired during the revising process should positively impact the quality of students’ writing” (p. 87). Furthermore, the criteria in the ID feedback seemed salient and important especially after being possibly restated over the three occasions of treatment. According to Wickens (2007), learners would pay selective attention to material of salience and value. The ID students therefore might absorb these criteria and apply them when writing anew. By contrast, the prescriptive feedback appeared more diffuse because it involved text-specific, concrete solutions although it actually had the same scope of targeted aspects as the diagnostic comments.

7.4.1.2 The effect of \pm RFD

Analysis of the factor of \pm RFD indicated that the +RFD condition rendered a significant improvement from Task 1/2 to Task 5 with a fairly large effect size ($p = .003$, $d = 0.76$) while the –RFD condition did not render any significant improvement. This suggested that +RFD promoted student writing proficiency in terms of content quality. The reason should be in accord with the mechanism that accounted for the +RFD condition’s effectiveness for discourse-level revision, as discussed earlier. Students’ focused attention to content feedback and resultant deep processing might have enabled students to convert the input into intake stored as part of their writing knowledge. It is notable that although the +RFD students taken together made a significant improvement, when taken separately only the ID+RFD group improved significantly. This is probably due to the advantage of diagnostic comments over prescriptive comments.

7.4.2 The effect for organisation quality in new writing

The results regarding the treatment effect on organisation quality in student new writing were more encouraging. Three treatment conditions were found to render significant improvement (see section 6.3.3). With respect to the ID+RFD and IS+RFD groups, the organisation quality in Tasks 3, 4, and 5 excelled that in Task 1/2, and overall the effect

size was large. The ID–RFD group’s organisation quality in Tasks 4 and 5 excelled that in Task 1/2, and the effect size was also large. The IS–RFD and the control groups did not improve significantly over time in new writing.

It seemed that both the ID and the IS feedback, especially when coupled with revision-focus direction, prompted students to acquire good writing practice concerning macrostructure, cohesion, and coherence in their writing. What is encouraging is that the treatment conditions took effect rather quickly after one or two occasions of treatment, and that the effect remained to the last task of the study.

A possible explanation may be that some features targeted by the feedback were salient and the actions to improve these features did not demand much effort. For example, appropriate paragraphing, addition of introduction and conclusion, and smooth transitions were some of the points emphasised in the feedback. It was found in the student texts that treatment groups could often make successful changes in these aspects. Therefore, students’ poor performance in these aspects in the initial first drafts (before receiving feedback) probably resulted from their ignorance or neglect of the requirements to organise their text appropriately but not from their lack of ability to meet these requirements. Once the importance of these requirements was conveyed through feedback, students would pay attention and improve relevant aspects even in the first draft of a new writing-revising task. An introspection of students’ first drafts detected many in-draft changes to improve organisation. This evidence confirmed the idea that some organisational improvements were not cognitively demanding; students could attain them once they realised the norms of desired organisational features. They could make in-draft remediation during the first drafting and did not need to delay that to the revising session.

However, analysis of organisation in this present study did not differentiate between the three elements of macrostructure, cohesion (linguistic devices to maintain connectedness), and coherence (connectedness in concepts and ideas, which is less tangible compared to cohesion). Separate analysis into these three elements might reveal different degrees of “treatability” of them.

7.4.2.1 The effect of WF type

The factor of WF type did not demonstrate a role in the effect on organisation quality in new writing. Both feedback types brought significant improvement. The IS feedback seemed to take effect sooner (after merely one occasion of treatment), but the ID feedback had a larger effect size on average than IS ($d = 1.07$ vs. 0.70 ; for reference, see Table 6.11).

The effectiveness of both feedback types may result from the above-mentioned nature of some of the organisational problems targeted, i.e., relatively easy to improve. Hence, the approach of feedback did not make a big difference as long as it pointed out the problems for students. The quicker effect of the prescriptive feedback (after merely one occasion of treatment) might be because students could transplant the suggested solutions to their new writing while the ID students might need additional effort and time to hypothesise solutions. The larger effect of the ID feedback, however, suggested that the rating criteria-related diagnosis led to a greater degree of self-monitoring during new writing.

7.4.2.2 The effect of \pm RFD

Analysis of the factor of \pm RFD revealed a quicker (after merely one occasion of treatment) and larger ($d = 0.98$ vs. 0.79 on the average) effect of the +RFD condition despite that students under both conditions significantly improved. The advantage of the +RFD condition was not surprising because it was likely to engage focused attention to discourse-level feedback and consequent deep processing in response to the feedback. The result that the –RFD students also made significant improvement may be attributed to the “treatable” nature of part of the organisational problems, as explained in the beginning of section 7.4.2.

7.4.3 The effect on grammatical accuracy in new writing

Of the five groups only the ID+RFD group significantly increased grammatical accuracy from Task 1/2 to Task 5 ($p = 0.043$, $d = 0.55$). The other groups improved only marginally based on descriptive data. This result seemed to contradict some positive evidence for the long-term effect of comprehensive direct or indirect feedback (Chandler, 2003; Sampson, 2012; Van Beuningen et al., 2012).

One reason for the different research results may lie in the research methods. For example, Chandler (2003) and Sampson (2012) did not differentiate between mechanical, grammatical, and lexical errors, so their results were based on CF's effect on overall accuracy. For another example, Sampson used arithmetic computation instead of inferential statistical tests to assess the treatment effect.

Another explanation for the long-term ineffectiveness of some treatment conditions might be that the participants in this study had already reached a high level of grammatical accuracy before the treatment started. The five groups' mean grammatical error rates in Task 1/2 ranged from 5.45% to 5.88%. However, the treatment groups in Van Beuningen et al.'s (2012) study with a similar starting grammatical error rate (ranging from 3.0% to 5.5%) still improved significantly after one-shot treatment. There might be hence some other reasons for the different results.

The students in Van Beuningen et al.'s study were in a second language context, and the target language was Dutch, while the students in this study were EFL learners. Probably more importantly, the treatment groups in this study received discourse-level comments in addition to CF throughout the three tasks in the treatment stage, which might divert them from focusing on improving linguistic accuracy only. According to a report by Chinese National Association of English Writing [NAEW] (2015), Chinese college students, first-year students in particular, tend to shift their priority from accuracy to meaning-making in EFL writing. This shift partly results from a different writing instruction approach in college from that in high school. Therefore, the emphasis on discourse in classroom writing instruction coupled with the discourse-level feedback may lead to an increased concern for content and organisation at the cost of reduced attention to accuracy.

Another reason for the little change in grammatical accuracy may be that some of the errors the students in this study made involved relatively complex features (Ferris, 1999), such as the adverbial clause and the attributive clause. Those complex features may be impervious to feedback, at least unfocused feedback; instead, they require a long and gradual period to be internalised into the learner's EFL system (Truscott, 1996, 2007). They are likely to be inaccurate especially when spontaneously produced (De Graaff & Housen, 2009).

In addition, different writing topics might have given rise to topic-related errors that had not been addressed in previous treatment. For example, in Task 5 a common error in student essays was using research as a countable noun (e.g., “a research”, “many researches”). Such an error had not occurred in Tasks 2–4 and had not had a chance of receiving feedback.

It seemed that the ID feedback combined with revision-focus direction rendered the best outcome for long-term grammar acquisition. This result may be understandable if we look at the effects of WF type and \pm RFD.

7.4.3.1 The effect of WF type

It was found that neither feedback type brought a significant grammatical accuracy increase in new writing; students receiving the ID feedback made some improvement from Task 1/2 to Task 5 that approached statistical significance ($p = .062$; $d = 0.39$). This result was in line with the Sampson (2012) findings about a larger benefit of coded feedback than direct corrections. However, in the study of Van Beuningen et al. (2012) direct feedback rather than coded feedback helped to generate grammatical accuracy increase in the posttests. It is notable that results of these studies are not completely comparable because of their different research designs, such as occasions of treatment, the form of diagnostic feedback (metalinguistic explanations vs. codes), and participants (university EFL students vs. Dutch-as-a-second-language pupils).

The result related to the ID feedback’s effect in this study suggested an advantage of metalinguistic explanations over direct corrections. The reason might be that the metalinguistic explanations promoted problem-solving and hypothesis testing, which is regarded as the benefit of indirect feedback approach for fostering learning (Bitchener & Knoch, 2008). In addition, the metalinguistic explanations might be superior to codes because of their explicitness. Even if some of them were not acted upon immediately in revision, a read of them may serve as a mini-instruction or review of certain grammatical features. The rules can thus leave a deeper impression in the students’ brain. They can be retrieved from their long-term memory for new writing or in-draft editing during new writing. By contrast, mere provision of grammatical forms will not be applied in new writing if students do not understand the nature of the problems (Ziv, 1984). Furthermore, the participants’ metalinguistic competency as mentioned earlier may

enable them to take advantage of the diagnostic information (Hyland & Hyland, 2006b; Sheen, 2007).

The effect of the ID feedback needs to be interpreted with caution, though, since its effect only approached statistical significance with a small effect size. Maybe the effect of the metalinguistic explanations would be larger if the treatment could be practised more frequently with more writing topics and diagnostic feedback could cover more grammatical features.

7.4.3.2 The effect of \pm RFD

The factor of \pm RFD did not make a difference in grammatical accuracy over time; no significant grammatical accuracy increase was observed under either condition. It was reported earlier that the $-$ RFD condition seemed more effective for student immediate revision in terms of grammatical accuracy increase (see section 4.4.3), and it was discussed that the $-$ RFD students probably gave priority to revising linguistic problems during the treatment stage and responded to CF more readily (see section 7.2.3.4). If they processed CF more deeply, why did not they produce higher accuracy in new writing?

An explanation may be that the $-$ RFD students did engage in more trial-and-error attempts (Hattie & Timperley, 2007) during the treatment stage but that they did not necessarily develop a higher level of metalinguistic reflection on their L2 system. It is possible that they focused on editing the diffuse errors in response to external feedback or internal feedback but did not reflect on their errors in a synthetic and inductive way. As a result, the specific L2 features that were noticed (Swain, 1995) and “registered in short-term memory as intake” might not be “further processed and integrated in long-term memory” (De Graaff & Housen, 2009).

7.4.4 The effect for lexical accuracy in new writing

None of the five groups significantly increased lexical accuracy in new writing. This result seemed to disagree with the studies of Ferris (2006) and Riazantseva (2012) that observed CF’s effectiveness in increasing lexical accuracy. Several reasons may account for the different results such as the research context and the length of treatment. Specifically, this study was conducted in an EFL context while the other two studies in

an ESL context. In addition, Ferris's and Riazantseva's CF treatment lasted for a semester in contrast to the three times in this study. Moreover, Riazantseva's feedback method was supposed to be more effective because it contained direct correction plus metalinguistic codes.

It seems that lexical accuracy, conceptualised as word choice here, is difficult to increase with short-term treatment at least in an EFL context. Students may solve specific word choice errors with the help of direct corrections from teachers (Ferris, 1999; Truscott, 1996), but the corrections can hardly be applied to a new context since wording errors are not patterned but miscellaneous (Ferris, 2006). Moreover, word choice problems may be more persistent for EFL learners because of limited exposure to the target language (NAEW, 2015) coupled with L1 interference (Chan, 2010).

7.4.4.1 The effects of WF type and \pm RFD

Analysis of the factors of WF type and \pm RFD showed no effect of either factor. Neither WF type rendered a significant lexical accuracy increase; neither \pm RFD condition did so. These results indicated the impervious nature of lexical errors in an EFL context. Neither direct correction nor a simplistic "wrong word choice" diagnosis could help students to improve word choice in a short period of three weeks. Probably due to the minimal effect of WF, how students attended to WF did not make a difference.

CHAPTER EIGHT

CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

8.1 Overview

This chapter will first summarise the major findings of the study. It will then explain how this study has contributed to knowledge in terms of theoretical implications, empirical insights, methodological contribution, and pedagogical implications. It will also discuss the limitations of the study and make recommendations for future research.

8.2 Major Findings

As introduced in Chapter One, this study aimed to investigate the role of WF and revision-focus direction in the effect of treatment on student revision and writing. It examined the effects of treatment methods, generated by a manipulation of the two factors, and analysed the role of WF type and \pm revision-focus direction. Major findings in relation to the three sets of research questions are presented as follows.

8.2.1 The immediate effect of treatment on revision

1. All the treatment methods (ID+RFD, ID-RFD, IS+RFD, IS-RFD) were effective in helping students improve text quality from the initial draft to the revised draft. Their effectiveness was relatively more evident for improvement in content, organisation, and grammatical accuracy and less evident for lexical accuracy.
2. WF type (ID vs. IS) did not play a role in the effect of treatment on revision on either discourse or linguistic levels.
3. The factor of \pm RFD made a difference in student revision result. The +RFD condition was more effective for revision on discourse levels, while the -RFD rendered better result for increases in grammatical accuracy.

8.2.2 The effect of treatment on autonomous revision

1. Overall, the treatment methods were not effective in promoting students' ability in autonomous revision. In terms of revision on discourse levels, only the

ID+RFD treatment method developed students' ability to make between-draft improvement autonomously, though the corresponding result did not reach statistical significance. In addition, none of the treatment methods led to better result for autonomous revision on linguistic levels.

2. The factor of WF type did not make a difference in student autonomous revision on either discourse or linguistic levels. Neither feedback method developed students' ability to autonomously improve text quality across drafts.
3. The factor of \pm RFD played a role in student autonomous revision on discourse levels, with an advantage for the +RFD condition. However, it made no difference in student autonomous revision on linguistic levels; neither the +RFD nor the -RFD conditions enhanced students' ability to improve accuracy across drafts.

8.2.3 The effect of treatment on new writing

8.2.3.1 Content quality in new writing

1. Regarding content quality in new writing, only the ID+RFD treatment method enabled students to make a statistically significant improvement over time.
2. The ID feedback was effective in helping students produce better content quality in new writing while the IS feedback was not.
3. The +RFD condition promoted student writing proficiency in terms of content quality while the -RFD condition did not.

8.2.3.2 Organisation quality in new writing

1. The ID+RFD, the IS+RFD, and the ID-RFD treatment methods promoted student writing quality in organisation over time.
2. The factor of WF type did not demonstrate a role in the effect on organisation quality in new writing; both feedback types brought significant improvement.
3. The +RFD condition exhibited a higher degree of effectiveness despite that both \pm RFD conditions helped students improve organisation quality in new writing.

8.2.3.3 Grammatical accuracy in new writing

1. Only the ID+RFD treatment method helped students increase grammatical accuracy in new writing.
2. The ID feedback led to some improvement over time, although neither feedback type brought a significant grammatical accuracy increase in new writing.
3. The factor of \pm RFD did not make a difference in student grammatical accuracy development; neither condition was effective.

8.2.3.4 Lexical accuracy in new writing

1. None of the treatment methods enabled students to increase lexical accuracy in new writing.
2. Neither WF type resulted in student lexical accuracy increase in new writing, and neither of the \pm RFD conditions did so.

8.3 Theoretical Implications

Theoretical implications of this study for the field of WF and L2 student revision include theoretical application and theoretical verification. The first aspect refers to the initiative in using revision theories to re-examine a crucial issue in L2 writing instruction, and the second aspect relates to the corroboration of revision theories with empirical evidence obtained from an L2 context.

8.3.1 The relevance of L1 revision theories to understanding the issue of teacher feedback and L2 revision

This study has demonstrated the possibility of exploiting revision theories to enrich field knowledge in teacher feedback and L2 writing. Previous research on teacher WF and L2 writing are either not theoretically grounded or based on SLA theories, mainly the social cultural approach and the interactionist approach (Bitchener & Storch, 2016; Ellis, 2010; Erlam, Ellis & Batstone, 2013; Polio, 2012). Despite their important contribution to the SLA/L2 writing interface, many SLA approaches are not unequivocally related to writing, and research adopting SLA approaches was primarily concerned with L2 learners' accuracy development through writing practice with the help of feedback.

However, English writing is more than a tool of grammar exercises, even for EFL learners (Wang, 2014; Zhang, 2013b). Therefore it is reasonable to adopt another theory that pertains more directly to writing to approach the issue of feedback, revision, and writing development. This study has found the potential of L1 revision theories in this regard.

First, the proposed essential processes for successful revision helped me to analyse the nature of teacher feedback according to the information it provides in relation to these essential processes. L1 revision theories posit that the core procedure of revision in writing entail three steps: detection of problems, diagnosis of problems, and solution to problems (Butterfield et al., 1996; Flower et al., 1986; Hayes, 1996; Scardamalia & Bereiter, 1983). Hence, successful revision of a problem requires corresponding information in the three aspects. The efficacy of teacher feedback can be analysed by examining whether it contains information related to detection, diagnosis, and solution of problems in student writing. Therefore, I designed two types of teacher feedback according to the information they supplied to the above-mentioned essential processes for revision. The ID feedback provided information related to problem detection and diagnosis, and IS contained information regarding problem detection and solution.

This approach to categorise teacher feedback seems more theoretically grounded and directly based on the reviser's thinking process. By contrast, early typologies of WF (e.g., direct, indirect, explicit, inexplicit) seem to start from the perspective of teachers' feedback methods; a notable drawback has been the inconsistent conceptualisations of WF types. For example, despite the same term, "indirect" feedback has been designed differently in studies. Although Ellis (2009) proposed a thorough typology of written CF, terms such as "direct" and "indirect" have still been conceptualised and operationalised in different manners.

Second, the claim about the cognitively demanding nature of the revising process postulated by Hayes (1996, 2012) and Butterfield et al. (1996) stimulated me to study revision not only in relation to teacher feedback but also to task conditions that would impact on cognitive effort and allocation of cognitive resources. This perspective has been overlooked in previous studies on feedback and L2 writing.

In addition, the idea of task schema in revision theories (Hayes, 1996; or “task definition” in Flower et al., 1986) was useful to interpret some of the research results. Task schema refers to how a writer approaches a revision task, involving the writer’s goals, criteria for good writing, and plan of conducting the revising process. It plays an essential role in the revising process and definitely mediates the effect of teacher feedback. If students’ task schema does not have a clear goal to improve certain aspects during revision, they are not likely to pay special attention to those aspects addressed by teacher WF; nor are they likely to activate relevant knowledge in their long-term memory. On the other hand, whether the writer’s criteria for good writing is in accordance with the teacher’s criteria illustrated in the feedback also influences how much and how well the writer would embrace teacher feedback (Conrad & Goldstein, 1999; Han & Hyland, 2015).

There are some other propositions in revision theories that were not investigated in this study but seem highly relevant to the field of feedback and L2 writing, and they bear latent connections with L2 learning queries which concern researchers and teachers. For example, task schema of students could be an important variable affecting revision process and outcome. Its underlying mechanism has close relations to L2 learning such as engagement with feedback and allocation of attention. For another example, the cognition-metacognition system (Butterfield et al., 1996) may illuminate and expand the understanding of learner factors such as language analytical ability. Furthermore, the contextual factors in Ellis’s (2010) framework for investigating CF seem to correspond to the concept of “environment” in Butterfield et al.’s (1996) revision model. In brief, L1 revision theories and L2 research on feedback and revision bear enlightening interfaces, which are worthy of inspection.

The aforementioned is the way of drawing on revision theories to enrich the understanding of feedback and revision. Nonetheless, revision theories can illuminate researchers on research perspectives as they informed me. A pool of interesting variables, such as topic knowledge, the readers of the text, and perceived importance of task, can be identified for empirical investigation into L2 student revision.

8.3.2 Corroborating L1 revision theories with L2 evidence

The findings obtained from the study have, in return, provided evidence to the propositions in revision theories. The finding that treatment groups achieved better

revision results than the control group has indicated that detection, diagnosis, and solution are essential for effective re-evaluation and revision (Flower et al, 1986; Hayes, 1996). Treatment groups, provided with diagnostic or prescriptive information, probably evaluated their text more deeply and sensed more discrepancies between their intended text and the produced text. They were likely to gain a better understanding of the sensed problems and figure out the strategies to solve them. As a result they improved their text more effectively than the control group.

Another important finding that revision-focus direction facilitated student revision on discourse levels has to some extent verified that revising is cognitively demanding, and that it is subject to the cognitive demand of the task and the cognitive resources available to the reviser (Allal et al., 2004). The research results have indicated that when EFL students did not concentrate their cognitive resources on revising at discourse levels, their resources were dispersed at both discourse and language which competed for their resources. In other words they could not manage to achieve both goals of improving discourse and language simultaneously.

The research results have also suggested that cognitive effort demanded by a revising task interplays with the reviser's task schema with respect to the focus of their effort investment. It was possible that students without revision-focus direction achieved better revision result at linguistic levels because they were more concerned with linguistic accuracy. Therefore those students invested a large part of their limited resources to improving language, and they engaged more with CF despite the provision of both CF and global feedback. In this sense task schema can monitor the allocation of resources and mediate the effect of cognitive demand of a revising task.

8.4 Empirical Insights

The findings obtained from this study have enriched the understanding of the role of teacher WF and revision-focus manipulation in student revision and writing development. The following presents the important contributions of this study in terms of empirical insights with due qualifications.

8.4.1 The immediate effects on revision

This research has generated clear evidence that teacher WF enabled Chinese university EFL students to improve text quality in content, organisation, and grammatical accuracy. This finding is well warranted given the consistent evidence observed on three treatment occasions. In view of the studies that approached discourse-level comments' efficacy in relation to their pragmatic and syntactic features (e.g., hedging, politeness, questioning), this study has indicated that the information teacher comments contain, rather than the way they are expressed, may be a more accurate predictor of their usefulness to tertiary students (Goldstein, 2006). Both criteria-related information and provision of solution facilitated students' revision in content and organisation. In regard to CF, the finding advances the understanding of the effect of comprehensive metalinguistic CF. Unfocused metalinguistic explanations proved to function as well as direct corrections for learners with some metalinguistic knowledge, although coded CF has been found inferior to direct corrections in previous studies. However, whether metalinguistic knowledge is a prerequisite for L2 learners' uptake from this type of CF is uncertain.

Another implication from this study is related to the mediation of textual level in the effect of teacher treatment on student writing. Organisation (macrostructure, cohesion, coherence) in argumentation was found to be more amenable to WF than content quality (overall persuasiveness), which, given the little evidence in this regard, lends important support to the finding of the Conrad and Goldstein (1999) study. However, whether it is the case for other genres of writing, which element of content quality (e.g., relevance to topic, support, or counterargument) is most difficult for students to improve, and which element of organisation quality is relatively difficult to improve remain unanswered. The finding that lexical accuracy (lexical choice) was less treatable compared to grammatical accuracy corresponds to a recent finding on Chinese college EFL students (NAEW, 2015). This piece of evidence offers insights into the issue of treatability of linguistic error types (Ferris, 1999, 2006; Truscott, 1996).

The probe into revision-focus manipulation confirms that revision is a cognitively demanding activity (Butterfield et al., 1996; Hayes, 1996). The relevant findings suggested that revising on different dimensions separately in sequence allowed L2 learners to pay focused attention to discourse-related WF and to revise discourse with focused cognitive resources. Such cognitive engagement facilitated the learning and

revising outcome in the corresponding aspects, just as attention to and cognitive processing with CF have been found to benefit acquisition of linguistic forms (Han & Hyland, 2015; Storch & Wigglesworth, 2010). However, the manipulation did not generate similarly positive revision effect for linguistic accuracy; plausible reasons were discussed in Chapter Seven. The design of another +RFD condition, directing students to focus on revising language prior to discourse, might generate different findings and offer further insights into the issue of cognitive load and effort competition during revision, although such a design seems counter-intuitive in terms of writing pedagogy (Bruton, 2009). On the other hand, revising discourse and revising language concurrently seemed to interfere with each other due to a competition of resources (Wickens, 2007); the result was to the disadvantage of discourse improvement probably because accuracy took priority for the EFL learners in this study. These findings complement the limited earlier evidence that meaning-level revision would be undermined by a heavy cognitive load (Chanquoy, 2001).

8.4.2 The effects on autonomous revision

The present study has demonstrated, for the first time, the overall ineffectiveness of WF in fostering autonomous revision when the feedback targeted specific problems in student writing. This confirms that a revision task demands not only knowledge of a better version but evaluative awareness and procedural knowledge as to what textual levels to work on and the sequence of working on them (i.e., task schema or task definition). As revision theories posit, revision is a reflective activity guided by goals, plans, and strategies (Flower et al., 1986; Hayes, 1996). Therefore, a student equipped with linguistic knowledge may not necessarily have the metalinguistic awareness to detect his/her errors, and a student who knows how to reorganise his/her ideas might fail to do that because the student does not have the goal to improve organisation during revision (Plumb et al., 1994).

At the same time, this study suggested that deep processing of discourse-level WF, diagnostic feedback in particular, contributed to the transfer of learning from an assisted revision task to an autonomous revision activity. The treatment groups under the +RFD condition, particularly the ID+RFD group, produced larger gain scores in content and organisation from the pre- to the post-treatment stage compared to the treatment groups under the –RFD condition. It is plausible to speculate that focused attention to discourse-

related criteria during the treatment stage helped the students to internalise those criteria, which they applied to govern the autonomous revision task.

Given that this study took the initiative in testing the effect of WF on Chinese EFL students' autonomous revising skills, the results need further verification. Maybe students with higher inductive reasoning ability or stronger motivation could take a step further beyond discrete feedback points, figure out strategies for a revising task in addition to knowledge of good writing, and move toward being self-sufficient revisers.

8.4.3 The effects on new writing

Being one of the few studies to include discourse levels and linguistic levels in the investigation of written feedback, the present study has extended the understanding of feedback efficacy in relation to textual level. The diagnostic feedback was found to be more effective than the prescriptive feedback for developing content quality and grammatical accuracy over time; the two feedback methods were equally useful for organisation quality but equally ineffective for lexical accuracy improvement. The mixed results suggested an interaction between feedback efficacy, feedback type, and the nature of textual levels.

It seems that for textual levels whose quality is governed, or can be explained, by rules or criteria, diagnostic feedback will yield a better result because of its transferability. For textual levels whose quality is governed by relatively simple rules or criteria, diagnostic and prescriptive feedback seem to be equally useful because learners are likely to extract certain rules from repeated instances of prescriptive feedback. Finally, for textual levels not governed by a set of rules prescriptive feedback might be a better treatment, but learners need to memorise the provided correct version in order to apply it in a new context; that is why memorisation has been found effective for EFL learners to master formulaic expressions (Ding, 2004; Wen & Johnson, 1997).

The evidence that +RFD produced a better effect than -RFD for discourse-level writing improvement demonstrates the role of attention allocation in student learning from WF regarding how to compose persuasive argument with smooth organisation. Available resources to process feedback in an attentive way, coupled with repeated treatment occasions, benefited student writing quality over time. This evidence confirmed Song

and Ferretti's (2013) finding of positive effect of teacher criteria-based questions about argumentation on new writing in addition to revising behaviour.

It is notable that the findings regarding the treatment effect on student new writing involved the factor of revision practice, i.e., both cognitive and behavioural engagement with WF (Ellis, 2010; Han & Hyland, 2015). Suppose students had not have the opportunity to revise with corresponding treatment but simply attended to feedback points under different revision-focus conditions, the results might have been different.

8.5 Methodological Contribution

The contributions of this research partly result from its innovations in research design. First, in view of the number of studies using a one-shot treatment, the three treatment occasions through three writing-revising tasks warranted the validity of the results to a larger degree. Second, the consideration of both discourse and linguistic quality and the dichotomy along each dimension (content and organisation; grammatical and lexical accuracy) yielded rich and complex findings. Moreover, the analysis of the factor of revision-focus demonstrated the role of allocation of resources in student revision and writing development.

In view of the plentiful studies which examined content feedback and CF separately, the most important insight offered by this study for methodology is the necessity of viewing student revising process and processing of feedback as a multi-tasking procedure where variables work in interaction rather than individually. Student response to feedback and their revising actions are likely to be in a dynamic interplay with the textual level they work on, their allocation of cognitive resources, and task conditions. Research should allow considerations for these multiple variables even if it carries a single focus (e.g., CF and accuracy), because students probably do not process their text on one track only.

8.6 Practical Implications

Practical implications drawn from the findings of this study are discussed in two aspects. One is how EFL teachers can be informed about effective feedback practice, and the other is how EFL teachers can facilitate learners to become self-sufficient revisers.

8.6.1 Informing teachers of effective feedback practice

It was introduced in the beginning of the thesis (Chapter One) that in Chinese EFL writing instruction teacher written feedback is still the most widely practiced feedback method with the coexistence of peer feedback and emerging computer-generated feedback. However, writing teachers have been confused about the efficient way of responding to student writing. This study provides some implications for promoting effective feedback methods.

It is crucial for writing instructors to recognise the interacting variables involved in student revising process, including feedback type, textual level to work on and student attentional focus. When they respond to student writing, they need to bear in mind these factors.

The first specific implication drawn from this study addresses the choice between diagnostic and prescriptive feedback. Teachers should realise the value of diagnostic feedback to students, be it criteria-based discourse-level comments or metalinguistic information. Diagnostic feedback proves to be helpful for both feedback-assisted revision and writing development of Chinese university EFL learners. It is, however, more time-consuming than direct provision of solutions, particularly for the case of treating linguistic errors. Therefore, teachers can select the representative or important problems to treat, and they can use recorded feedback to save time if convenient. Another alternative is to use finely designed codes, but the premise is to familiarise students with the codes. It should be noted that the efficacy of a feedback type is not absolute but contingent on related factors such as learners' metalinguistic knowledge, preference of comment type, and their changeable writing proficiency. Teachers should bear complex factors in mind and adjust their feedback practice strategically to achieve high efficacy.

The second implication for feedback practice derives from the findings about the role of revision-focus direction. Teachers might want to provide feedback on discourse and language separately and meanwhile prescribe revision-focus direction accordingly. That is to give discourse-related comments to students' first draft and ask students to attend to improving discourse, and then give CF to the second draft and ask students to focus on increasing accuracy (Sommers, 1982; Zamel, 1985). As the research findings

indicated, the +RFD produced contrary effects for discourse- and linguistic-levels. While +RFD had an advantage over –RFD for content and organisation improvement, the –RFD was superior for improving grammatical and lexical accuracy. The trade-off effect poses a dilemma as to whether teachers should direct students’ cognitive resources to different levels in sequence or not. Therefore, a better option may be to provide feedback on the two levels separately through a multiple-drafting procedure. The key point is that teachers explicitly ask students to focus on improving discourse when they are exposed to discourse-level feedback. Otherwise they might still be preoccupied with language problems due to an ill-informed revision task definition, and as a result students may not attend to teacher comments or do not process them deeply, so they can hardly benefit from them (Wickens, 2007).

The research findings can also raise teacher awareness of the role of limited capacity and attentional allocation in student revising process as well as student engagement with feedback. It might be unwise for teachers to give too many feedback points all at once on a single draft even if they fall into the same textual level. If teachers have to do so, they can highlight the high-stakes problems. A better method is to target particular goals in each feedback round.

With respect to the seemingly untreatable lexical choice errors, teachers need to look for alternative ways to help EFL students cope with lexical choice. In addition to recommending suitable wording, teachers can categorise lexical choice errors by analysing the reasons behind (Chan, 2010; Diab, 2015); such analysis may be more efficient by a discussion with students about representative inappropriate expressions than by teachers’ speculation on their own. Next teachers can provide diagnostic feedback targeting word choice difficulties of various sources. Such feedback may exert a positive effect similar to that of metalinguistic rules on grammatical errors.

8.6.2 Fostering self-sufficient revisers

The findings of this study have some implications for developing self-sufficient revisers. First, writing teachers can consider helping to modify Chinese students’ revision task schema, since evidence from this study indicated that students gave priority to revising language during a revising session. Such unbalanced priority may be moulded by the way they have been responded to, i.e., their teachers’ obsession with linguistic errors in

giving feedback; it may also be intentional ignorance of global revision due to students' laziness. Regardless of the reason, teachers can explicitly inform students of a sound approach to a revision task. A mini-lecture or a hand-out can largely serve the purpose. Teachers need to tell students English writing is more than language practice, and that revision in writing should have multiple goals including rediscovering meaning, rearranging structure, and refining style and language. Teachers should also prompt students to consider the proper sequence of working on these goals and how to achieve them with or without external feedback. Certainly these questions do not require unanimous or fixed answers, but students need to think about them to develop an individual schema for revision that can promote their L2 writing development. During regular writing and revising practice, teachers may occasionally ask students to reflect on their schematic approach to the revising activity to consolidate their awareness and modify their previous schema.

Second, writing teachers ought to give another form of intervention in student writing in addition to feedback, i.e., instructions on revision strategies and self-regulation (cf. Teng & Zhang, 2016). Being a neglected issue in the EFL context, autonomous revision deserves more attention since the ultimate objective of writing instruction is to foster self-sufficient writers and revisers (Bitchener & Ferris, 2012; Teng & Zhang, 2016). Teachers should seek effective methods to develop students' ability to revise their text autonomously. Based on the research findings, feedback on discrete problems in student writing does not seem to be effective for serving this purpose. Therefore, teachers need to provide extra assistance in this regard, such as procedural guidance with prompts and tutorials on how to revise, which have been found useful by researchers (Reynolds & Bonk, 1996; Song & Ferretti, 2013)

Finally, EFL students need to be aware of their limited resource capacity, and that working on discourse and language are likely to interfere with each other for most of them. With such awareness, students can strategically allocate their resources instead of trying to handle all kinds of problems concurrently. If provided with feedback, they can make strategic use of it based on their understanding of their resource capacity and task conditions. For instance, at the early stage they can pay selective attention to comments on their central idea and leave aside other comments; after making improvement in that

respect, they can respond to other aspects of teacher feedback. In this way, they can reduce the cognitive load and optimise the benefit from the feedback.

8.7 Limitations

This study has limitations in terms of scope and weakness. First, the small sample size, particularly the small size of each cell group ($14 \leq n \leq 17$), might have skewed the results to some extent. This weakness seems hard to avoid, however. Adopting a multiple-treatment experimental design, this study had difficulty in obtaining agreement to be conducted on intact classes. For the same reason, it was hard to recruit a large number of participants, and the recruited participants saw a rate of attrition. Another critical reason was my limited energy, which inhibited me from giving written feedback to a large body of texts within a couple of days.

Another limitation in scope is that it contained no complementary sources of data (e.g., think-aloud protocols, interviews and questionnaires) to triangulate the results based on the student written texts. In consequence, this study could only infer learners' cognitive processes based on performance-related measures. It lacks a more accurate representation of what was going on in the students' mind during the revising process. One particular issue is the lack of independent evidence for the validity of the revision focus direction. It could not be ascertained that students with revision-focus direction did follow the directions focusing on discourse and language separately in sequence. It was not impossible that they allocated relatively equal attention to both dimensions of their text. If that were the case, the results with respect to the differences between the \pm RFD students need to be explained by other reasons instead of the revision focus manipulation. In other words, the lack of evidence for student cognitive processing requires the effect of revision-focus direction to be interpreted cautiously.

The proficiency level of the participants should have been more rigorously assessed. I measured participants' proficiency by their linguistic accuracy exhibited in their first draft for Task 1. The imprecise assessment of participants' proficiency might have affected the random grouping, which could have been an extraneous variable affecting the results. Alternatively, an average of participants' scores at various textual levels over a set of writing tasks might have given a more reliable reflection of their writing proficiency.

Another weakness of this study lies in the analysis of between-draft revisions. I conducted the coding work independently without a corater, somewhat impairing the reliability of the corresponding result.

8.8 Recommendations for Further Research

This study has recommendations for future research in the area of written feedback, revision, and writing development. Some additional work can be undertaken to pursue the research perspectives of this study. First, more attempts are needed to look at both discourse- and accuracy-oriented feedback other than studying them separately. After all, it is common for teachers to give dual-level feedback on student writing. Second, another condition of +RFD can be included for comparison, i.e., directing students to focus on improving language earlier and to work on discourse later. As mentioned in section 8.4, such a design might lead to complementary findings concerning the role of attentional focus in EFL learners' revision performance and their learning from feedback. Furthermore, a delayed posttest can be added to assess the lasting effect of WF and revision-focus direction. Finally, the limited research on the moderation of learner differences in feedback effect tends to focus on learners' language-related ability; many other learner factors identified from revision theories, such as working memory, topic knowledge, and task schema, are important variables for research.

There are other important research ideas to pursue in the investigation of L2 written feedback, including a long neglected area: contextual factors (Ellis, 2010; Goldstein, 2001). It is, if resources are available, interesting to study the impact of macrocontextual factors such as the setting of target language learning (e.g., L1 and L2). It is, however, more informative to examine the efficacy of written feedback in relation to microcontextual factors, as Ellis (2010) pointed out. As this study manipulated student revision-focus, future research can investigate other task conditions, including writing genres and topics, the location and time limit for completing a writing or revising task, the medium of writing and giving feedback, and the timing of giving feedback (cf. Shintani & Aubrey, 2016). Further to that, whether and how the regular classroom instruction interacts with the feedback students usually receive is a research gap worth investigation.

In terms of research methodologies, more qualitative methods are in need to obtain substantial data given the prevalent experimental designs and quantitative methods in the field under discussion. Qualitative methods, such as think-aloud protocol, retrospective interview, and diary are particularly valuable to investigate into the role played by contextual variables and learner factors.

8.9 Conclusions

Drawing on L1 revision theories, this study has extended the understanding of the effects of written feedback and revision task conditions on EFL student revision and writing development in a Chinese university setting. The outcome of treatment proved to be subject to the interactive impacts of written feedback type, allocation of resources during revision, and the nature of textual levels. In addition to the contribution to the knowledge of written feedback and L2 writing, this study has implications for theoretical application and drawing up a new research agenda. It also has practical implications for writing instruction and learner development in autonomous revision and writing proficiency in the L2 field.

REFERENCES

- Adams, A. M., Simmons, F., Willis, C., & Pawling, R. (2010). Undergraduate students' ability to revise text effectively: Relationships with topic knowledge and working memory. *Journal of Research in Reading, 33*(1), 54–76.
- Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. Boston, Massachusetts: Kluwer Academic Publishers.
- Allal, L. (2000). Metacognitive regulation of writing in the classroom. In A. Camps, & M. Milian (Eds.), *Metalinguistic activity in learning to write* (pp. 145–166). Amsterdam, Netherlands: Amsterdam University Press.
- Allal, L., Chanquoy, L., & Largy, P. (Eds.). (2004). *Revision: Cognitive and instructional processes*. Boston, MA: Kluwer Academic Publishers.
- Ashwell, T. (2000). Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing, 9*(3), 227–257.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A*(1), 5–28.
- Barkaoui, K. (2016). What and when second-language learners revise when responding to timed writing tasks on the computer: The roles of task type, second language proficiency, and keyboarding skills. *The Modern Language Journal, 100*(1), 320–340.
- Barlett, E. J. (1982). Learning to revise: Some component processes. In M. Nystrand (Eds.), *What writers know: The language, process, and structure of written discourse* (pp. 345–363). New York, NY: Academic Press.
- Beach, R. (1979). The effects of between-draft teacher evaluation versus student self-evaluation on high school students' revising of rough drafts. *Research in the Teaching of English, 13*(2), 111–119.
- Beach, R., & Friedrich, T. (2006). Response to writing. In C. A. MacArthur, S. Graham., & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 222–234). New York, NY: Guilford.
- Beal, C. R. (1996). The role of comprehension monitoring in children's revision. *Educational Psychology Review, 8*(3), 219–238.

- Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*(2), 102–118.
- Bitchener, J. (2012). A reflection on ‘the language learning potential’ of written CF. *Journal of Second Language Writing, 21*(4), 348–363.
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York, NY: Routledge.
- Bitchener, J., & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*(3), 409–431.
- Bitchener, J., & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*(2), 193–214.
- Bitchener, J., & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*(4), 207–217.
- Bitchener, J., & Storch, N. (2016). *Written corrective feedback for L2 development*. Buffalo, NY: Multilingual Matters.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*(3), 191–205.
- Board of Foreign Language Teaching in Higher Education Institutions (2000). *Gaodeng xuexiao yingyu zhuanke yingyu jiaoxue dagang [Syllabus for teaching English majors in higher education institutions]*. Shanghai, China: Shanghai Foreign Language Education Press.
- Bridwell, L. S. (1980). Revising strategies in twelfth grade students’ transactional writing. *Research in the Teaching of English, 14*(3), 197–222.
- Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *Revising strategies in twelfth grade students’ transactional writing (11–18)*. London, England: Macmillan Education.
- Bruton, A. (2009). Improving accuracy is not the only reason for writing, and even if it were.... *System, 37*(4), 600–613.
- Butler, J. A., & Britt, M. A. (2011). Investigating instruction for improving revision of argumentative essays. *Written Communication, 28*(1), 70–96.

- Butterfield, E. C., Hacker, D. J., & Albertson, L. R. (1996). Environmental, cognitive, and metacognitive influences on text revision: Assessing the evidence. *Educational Psychology Review*, 8(3), 239–297.
- Carifio, J., Jackson, I., & Dagostino, L. (2001). Effects of diagnostic and prescriptive comments on the revising behaviors of community college students. *Community College Journal of Research and Practice*, 25(2), 109–122.
- Chan, A. Y. W. (2010). Toward a taxonomy of written errors: Investigation into the written errors of Hong Kong Cantonese ESL learners. *TESOL Quarterly*, 44(2), 295–319.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267–296.
- Chanquoy, L. (2001). How to make it easier for children to revise their writing: A study of text revision from 3rd to 5th grades. *British Journal of Educational Psychology*, 71(1), 15–41.
- Chanquoy, L. (2009). Revision processes. In R. Beard, D. Myhill, M. Nystrand, & J. Riley (Eds.), *The SAGE handbook of writing development* (pp. 80–98). London, England: SAGE.
- Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. *Written Communication*, 20(11), 99–118.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328–338.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- College English Test Committee. (n.d.). *Daxue yingyu kaoshi de zuoyong he yingxiang* [The function and influence of College English Test]. Retrieved from http://www.cet.edu.cn/cet_concept4.htm
- Connors, R. J., & Lunsford, A. A. (1993). Teachers' rhetorical comments on student papers. *College Composition and Communication*, 44(2), 200–223.
- Conrad, S. M., & Goldstein, L. M. (1999). ESL student revision after teacher-written comments: Texts, contexts, and individuals. *Journal of Second Language Writing*, 8(2), 147–179.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786.

- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Los Angeles, CA: SAGE.
- De Graaff, R., & Housen, A. (2009). Investigating the effects and effectiveness of L2 instruction. In M. H. Long, & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 726–755). Malden, MA: Wiley-Blackwell.
- Deng, L., Chen, Q., & Zhang, Y. (2014). *Developing Chinese EFL learners' generic competence: A genre-based & process genre approach*. New York, NY: Springer.
- Diab, N. M. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing*, 24(1), 16–34.
- Ding, Y. (2004). *Beisong yingyu kewen: Xiandai zhongguo gaodeng yuanxiao zhong chuantong de yuwen xuexi fangfa [Reciting English texts: A traditional method of learning Chinese adopted in modern tertiary institutes in China]*. Xi'an: China: Shanxi Normal University Press.
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14(2), 88–115.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36(3), 353–371.
- Ellis, R. (2009). A typology of written corrective feedback types. *English Language Teaching Journal*, 63(2), 97–107.
- Ellis, R. (2010). Epilogue: A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32(2), 335–349.
- Ellis, R. (2015). *Understanding second language acquisition* (2nd ed.). Oxford, England: Oxford University Press.
- Erlam, R., Ellis, R., & Batstone, R. (2013). Oral corrective feedback on L2 writing: Two approaches compared. *System*, 41(2), 257–268.
- Faigley L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32(4), 400–414.
- Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Eds.), *Second language writing: Research insights for the classroom* (pp. 178–190). New York, NY: Cambridge University Press.
- Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, 36(1), 51–62.

- Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do goals affect the structure of students' argumentative writing strategies? *Journal of Educational Psychology, 101*(3), 577–589.
- Ferretti, R. P., MacArthur, C. A., & Dowdy, N. S. (2000). The effects of an elaborated goal on the persuasive writing of students with learning disabilities and their normally achieving peers. *Journal of Educational Psychology, 92*(4), 694–702.
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing, 8*(1), 1–11.
- Ferris, D., Pezone, S., Tade, C., & Tinti, S. (1997). Teacher commentary on student writing: Descriptions and implications. *Journal of Second Language Writing, 6*(2), 155–182.
- Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly, 31*(2), 315–339.
- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland, & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge, England: Cambridge University Press.
- Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition, 32*(2), 181–201.
- Ferris, D. R. (2012). Written corrective feedback in second language acquisition and writing studies. *Language Teaching, 45*(4), 446–459.
- Ferris, D. R., & Hedgcock, J. S. (2005). *Teaching ESL composition: Purpose, process, and practice* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ferris, D. R., & Roberts, B. J. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*(3), 161–184.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, England: SAGE.
- Fitzgerald, J. (1987). Research on revision in writing. *Review of Educational Research, 57*(4), 481–506.
- Fleiss, J., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement, 33*(3), 613–619.

- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986). Detection, diagnosis, and the strategies. *College Composition and Communication*, 37(1), 16–55.
- Frear, D. J. (2012). *The effect of written corrective feedback and revision on intermediate Chinese learners' acquisition of English*. (Unpublished doctoral thesis). The University of Auckland, Auckland, New Zealand.
- Frey, N., & Fisher, D. (2013). A formative assessment system for writing improvement. *English Journal*, 103(1): 66–71.
- Fritz, E., & Ruegg, R. (2013). Rater sensitivity to lexical accuracy, sophistication and range when assessing writing. *Assessing Writing*, 18(2), 173–181.
- Galbraith, D. (1980). The effect of conflicting goals on writing: A case study. *Visible Language*, 14(4), 364–375.
- Galbraith, D. (1992). Conditions for discovery through writing. *Instructional Science*, 21, 45–72.
- Galbraith, D., & Torrance, M. (2004). Revision in the context of different drafting strategies. In L. Allal, L. Chanquoy, & P. Largy (Eds), *Revision: Cognitive and instructional processes* (pp. 63–85). Boston, MA: Kluwer Academic Publishers.
- Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ge, S., & Chen, X. (2007). Zhongguo EFL xuexizhe zidong zuowen pingfen tansuo [An investigation into automated essay scoring of the writings of Chinese EFL learners]. *Waiyu Jie [Foreign Language World]*, 5, 43–50.
- Glynn, S. M., Britton, B. K., Muth, K. D., & Dogan, N. (1982). Writing and revising persuasive documents: Cognitive demands. *Journal of Educational Psychology*, 74(4), 557–567.
- Goldstein, L. (2001). For Kyla: What does research say about responding to ESL writers. In T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 73–90). Mahwah, NJ: Lawrence Erlbaum Associates.
- Goldstein, L. (2006). Feedback and revision in second language writing: Helping learners become independent writers. In K. Hyland, & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185–205). New York, NY: Cambridge University Press.
- Goldstein, L. M. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, 13(1), 63–80.

- Goldstein, L., & Conrad, S. (1990). Student input and negotiation of meaning in ESL writing conferences. *TESOL Quarterly*, 24(3), 443–460.
- Graham, S., MacArthur, C., & Schwartz, S. (1995). Effects of goal setting and procedural facilitation on the revising behavior and writing performance of students with writing and learning problems. *Journal of Educational Psychology*, 87(2), 230–240.
- Graham, S. (1997). Executive control in the revising of students with learning and writing difficulties. *Journal of Educational Psychology*, 89(2), 223–234.
- Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16(1), 40–53.
- Guo, D. (2015). Daxue yingyu xiezuojiaoxue gaige huigu yu zhanwang [College English teaching reform: Looking back and looking ahead]. *Dangdai jiaoyu lilun yu shijian [Theory and Practice of Contemporary Education]*, 7(3), 111–112.
- Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental and cognitive psychology. *Behavioral and Brain Sciences*, 21(6), 803–831.
- Han, Y., & Hyland, F. (2015). Exploring learner engagement with written corrective feedback in a Chinese tertiary EFL classroom. *Journal of Second Language Writing*, 30, 31–44.
- Hanaoka, O., & Izumi, S. (2012). Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing*, 21, 332–347.
- Haswell, R. H. (1983). Minimal marking. *College English*, 45(6), 600–604.
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills intervention on student learning: A meta-analysis. *Review of Research in Education*, 66(2), 99–136.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hayes, J. R. (2004). What triggers revision? In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (pp. 9–20). Boston, MA: Kluwer Academic Publishers.

- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29(3), 369–388.
- Hayes, J. R. & Chenoweth, N. A. (2006). Is working memory involved in the transcribing and editing of texts? *Written Communication*, 23(2), 135–149.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg, & E. R. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach* (pp. 3–30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hayes, J. R., & Flower, L. S. (1983). Uncovering cognitive processes in writing: An introduction of protocol analysis. In P. Mosenthal, S. Walmsley, & L. Tamor (Eds.), *Research on writing: Principles and methods* (pp. 206–219). New York, NY: Longman.
- Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. *American Psychologist*, 41(10), 1106–1113.
- Hedgcock, J., & Lefkowitz, N. (1994). Feedback on feedback: Assessing learner receptivity in second language writing. *Journal of Second Language Writing*, 3(2), 141–163.
- Hillocks, G. (1982). The interaction of instruction, teacher comment, and revision in teaching the composing process. *Research in the Teaching of English*, 16(3), 261–278.
- Housen, A., Kuiken, F., & Vedder, I. (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Philadelphia, PA: John Benjamins.
- Hyland, F. (2000). ESL writers and feedback: Giving more autonomy to students. *Language Teaching Research*, 4(1), 33–54.
- Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, 31(2), 217–230.
- Hyland, K., & Hyland, F. (2006a). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101.
- Hyland, K., & Hyland, F. (2006b). Feedback in second language writing: Contexts and issues. Cambridge, England: Cambridge University Press.
- Jacobs, H. L., Zingraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.

- Jiang, H. (2014). Daxue yingyu xiezuojiao xue xianzhuang diaocha yu duice [Teaching college English writing: The status quo and some implications]. *Hubei Gongcheng Xueyuan Xuebao [Journal of Hubei Engineering University]*, 34(1), 64–67.
- Jonsson, A. (2012). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63–76.
- Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, 99(1), 1–18.
- Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition*, 15(3), 256–266.
- Kellogg, R. T. (1988). Attentional overload and writing performance: Effects on rough draft and outline strategies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14(2), 355–365.
- Kellogg, R. T. (1999). Components of working memory in text production. In M. Torrance, & G. C. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory in text production* (pp. 42–61). Amsterdam, Netherlands: Amsterdam University Press.
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, 75(3), 305–313.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, 20(2), 141–163.
- Lam, C. Y. P. (1991). *Revision processes of college ESL students: How teacher comments, discourse types, and writing tools shape revision* (Unpublished doctoral thesis). The University of Georgia, Athens, USA.
- Larigauderie, P., Gaonac'h, D. & Lacroix, N. (1998). Working memory and error detection in texts: What are the roles of the central executive and the phonological loop? *Applied Cognitive Psychology*, 12(5), 505–527.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.
- Lee, C. C., & Tan, S. C. (2010). Scaffolding writing using feedback in students' graphic organizers – novice writers' relevance of ideas and cognitive loads. *Educational Media International*, 47(2), 135–152.

- Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for college-level teaching. *System*, 25(4), 465–477.
- Lee, I. (2002). Teaching coherence to ESL students: A classroom inquiry. *Journal of Second Language Writing*, 11(2), 135–159.
- Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17(2), 69–85.
- Levy, C. M., & Marek, P. (1999). Testing components of Kellogg's multicomponent model of working memory in writing: The role of the phonological loop. In M. Torrance & G. Jeffrey (Eds.), *The cognitive demands of writing* (pp. 25–41). Amsterdam, Netherlands: Amsterdam University Press.
- Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, 33(4), 623–636.
- Liu, Q., & Brown, D. (2015). Methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *Journal of Second Language Writing*, 30(4), 66–81.
- Lowie, W., & Seton, B. (2013). *Essential statistics for applied linguistics*. New York, NY: Palgrave Macmillan.
- Lunsford, A. A., & Lunsford, K. J. (2008). 'Mistakes are a fact of life': A national comparative study. *College Composition and Communication*, 59(4), 781–806.
- McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing abilities. *Journal of Memory and Language*, 25(4), 431–444.
- McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, 8(3), 299–325.
- McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, 35(1), 13–23.
- McCutchen, D., Francis, M., Kerr, S. (1997). Revising for meaning: Effects of knowledge and strategy. *Journal of Educational Psychology*, 89(4), 667–676.
- Midgette, E., Haria, P., & MacArthur, C. (2008). The effects of content and audience awareness goals for revision on the persuasive essays of fifth- and eighth-grade students. *Reading and Writing*, 21(1-2), 131–151.
- Ministry of Education of the People's Republic of China. (2000). *Quanrizhi Putong Gaoji Zhongxue Yingyu Jiaoxue Dagang* (shiyan xiudingban) [*The Senior High*

- School English Syllabus* (revised version)]. Beijing, China: People's Education Press.
- Ministry of Education of the People's Republic of China. (2001). *Jiaoyubu guanyu jiji tuijin xiaoxue kaishe yingyu kecheng de zhidao yijian [Guidance of the Ministry of Education on promoting English curriculum in primary schools]*. Retrieved from http://www.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_711/200407/665.html
- Mirzaii, M., & Aliabadi, R. B. (2013). Direct and indirect written corrective feedback in the context of genre-based instruction on job application letter writing. *Journal of Writing Research, 5*(2), 191–213.
- Monahan, B. D. (1984). Revision strategies of basic and competent writers as they write for different audience. *Research in the Teaching of English, 18* (3), 288–304.
- Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing, 16*(2), 82–99.
- Morra, A. M., & Asis, M. I. (2009). The effect of audio and written teacher response on EFL student revision. *Journal of College Reading and Learning, 39*(2), 68-81.
- Mota de Cabrera, C. (2003). *Teaching, tutoring, and revision: The experiences of two freshmen ESL students in a rhetoric class* (Unpublished doctoral thesis). The University of Iowa, Iowa city, USA.
- Murray, D. M. (1978). Teach the motivating force of revision. *English Journal, 67*(7), 56–60.
- National Association of English Writing. (2015). *Baiwan tongti: Yingyu xiezuozu dashuju baogao [One million students composing on the same topic: Report on big data on English writing]*. Beijing, China: Beijing Language Intelligence Research.
- Nimehchisalem, V., & Mukundan, J. (2011). Determining the evaluative criteria of an argumentative writing scale. *English Language Teaching, 4*(1), 58–69.
- Olive, T., Kellogg, R. T., & Piolat, A. (2002). The triple task technique for studying the processes of writing. In T. Olive, & C. M. Levy (Eds.), *Contemporary tools and techniques for studying writing* (pp. 31–59). Dordrecht, Netherlands: Kluwer Academic Publishers.
- Ong, J. (2013). Discovery of ideas in second language writing task environment. *System, 41*(3), 529–542.

- Ong, J. (2014). How do planning time and task conditions affect metacognitive processes of L2 writers? *Journal of Second Language Writing*, 23(3), 17–30.
- Ong, J., & Zhang, L. J. (2013). Effects of the manipulation of cognitive process on EFL writers' text quality. *TESOL Quarterly*, 47(2), 375–398.
- 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(4), 218–233.
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing*, 15(2), 68–85.
- Peterson, S. S., & Portier, C. (2014). Grade one peer and teacher feedback on student writing. *Education*, 42(3), 237–257.
- Piolat, A., Roussey, J.-Y., Olive, T., & Amada, M. (2004). Processing time and cognitive effort in revision: Effects of error type and of working memory capacity. In L. Allal, L. Chanquoy, & P. Largy (Eds.), *Revision: Cognitive and instructional processes* (pp. 21–38). Boston, MA: Kluwer Academic Publishers.
- Plumb, C., Butterfield, E. C., Hacker, D. J., & Dunlosky, J. (1994). Error correction in text: Testing the processing-deficit and knowledge-deficit hypotheses. *Reading and Writing: An interdisciplinary Journal*, 6(4), 347–360.
- Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, 21(4), 375–389.
- Polio, C. G. (1999). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47(1), 101–143.
- Porte, G. K. (1997). The etiology of poor second language writing: The influence of perceived teacher preferences on second language revision strategies. *Journal of Second Language Writing*, 6(1), 61–78.
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, 13, 25–42.
- Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education*, 33(2), 143–154.
- Qin, J., & Karabacak, E. (2010). The analysis of Toulmin elements in Chinese EFL university argumentative writing. *System*, 38(3), 444–456.
- Reid, M. J. (1993). *Teaching ESL writing*. Englewood Cliffs, NJ: Regents/Prentice Hall.

- Reynolds, T. H., & Bonk, C. J. (1996). Facilitating college writers' revisions within a generative-evaluative computerized prompting framework. *Computers and Composition, 13*(1), 93–108.
- Riazantseva, A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System, 40*(3), 421–430.
- Robinson, P. (1995). Attention, memory and the 'noticing' hypothesis. *Language Learning, 45*(2), 283–331.
- Robinson, P. (2005). Cognitive complexity and task sequencing: A review of studies in a Componential Framework for second language task design. *International Review of Applied Linguistics, 43*(1), 1–32.
- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics, 45*(3), 193–213.
- Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Eds.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–38). Amsterdam, Netherlands: John Benjamins.
- Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behaviour in the allocation of time to writing processes. *Journal of Second Language Writing, 17*(1), 30–47.
- Roussey, J-Y., & Piolat, A. (2008). Critical reading during text revision. *European Journal of Cognitive Psychology, 20*(4), 765–792.
- Sachs, R., & Polio, C. (2007). Learners' use of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition, 29*(1), 67–100.
- Sampson, A. (2012). Coded and uncoded error feedback: Effects on error frequencies in adult Colombian EFL learners' writing. *System, 40*(4), 494–504.
- Scardamalia, M., & Bereiter, C. (1983). The development of evaluative, diagnostic, and remedial capabilities in children's composing. In M. Martlew (Eds.), *The psychology of written language: Developmental and educational perspectives* (pp. 67–95). New York, NY: John Wiley & Sons.
- Schmidt, R. (2001). Attention. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 3–30). Cambridge, England: Cambridge University Press.
- Sengupta, S. (2000). An investigation into the effects of revision strategy instruction on L2 secondary school learners. *System, 28*(1), 97–113.

- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41(2), 255–283.
- Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37(4), 556–569.
- Sheen, Y. (2011). *Corrective feedback, individual differences and second language learning*. New York, NY: Springer.
- Shintani, N., & Aubrey, S. (2016). The effectiveness of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. *The Modern Language Journal*, 100(1), 296–319.
- Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22(3), 286–306.
- Shintani, N., & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System*, 49, 110–119.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, England: Oxford University Press.
- Skehan, P. (2015). Limited attention capacity and cognition: Two hypotheses regarding second language performance on tasks. In M. Bygate (Eds.), *Domains and directions in the development of TBLT: A decade of plenaries from the international conference* (pp. 123–155). Amsterdam, Netherlands: John Benjamins.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Eds.), *Cognition and second language instruction* (pp. 183–205). Cambridge, England: Cambridge University Press.
- Sommers, N. (1980). Revision strategies of student writers and experienced adult writers. *College Composition and Communication*, 31(4), 378–388.
- Sommers, N. (1982). Responding to student writing. *College Composition and Communication*, 33(2), 148–156.
- Song, Y., & Ferretti, R. P. (2013). Teaching critical questions about argumentation through the revising process: Effects of strategy instruction on college students' argumentative essays. *Reading and Writing*, 26(1), 67–90.
- Sperling, M., & Freedman, S. W. (1987). A good girl writes like a good girl: Written responses to student writing. *Written Communication*, 9(9), 343–369.

- Stefanou, C., & Révész, A. (2015). Direct written corrective feedback, learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal*, 99(2), 263–282.
- Stern, L. A., & Solomon, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing*, 11(1), 22–41.
- Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition*, 32(2), 303–334.
- Tang, J., & Wu, Y. (2012). Xiezuozidong pingjia xitong zai daxue yingyu jiaoxue zhong de yingyong yanjiu [Integrating an automated writing assessment system in college English teaching]. *Waiyu yu waiyu jiaoxue [Foreign Languages and Their Teaching]*, 4, 53–59.
- Taylor, S. S. (2011). “I really don't know what he meant by that”: How well do engineering students understand teachers' comments on their writing? *Technical Communication Quarterly*, 20(2), 139–166.
- Teng, L. S., & Zhang, L. J. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal*. doi: 10.1111/modl.12339
- Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition*, 16(2), 183–203.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46(2), 327–369.
- Truscott, J. (1999). The case for “The case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8(2), 111–122.
- Truscott, J. (2001). Selecting errors for selective error correction. *Concentric: Studies in English Literature and Linguistics*, 27(2), 93–108.
- Truscott, J. (2004). Evidence and conjecture: A response to Chandler. *Journal of Second Language Writing*, 13(4), 337–343.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272.
- Truscott, J., & Hsu, A. Y.-P. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17(4), 292–305.

- Van Beuningen, C., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62(1), 1–41.
- Wallace, D. L., & Hayes, J. R. (1991). Redefining revision for freshmen. *Research in the Teaching of English*, 25(1), 54–64.
- Wallace, D. L., Hayes, J. R., Hatch, J. A., Miller, W., Moser, G., & Silk, C. M. (1996). Better revision in eight minutes? Prompting first-year college writers to revise globally. *Journal of Educational Psychology*, 88(4), 682–688.
- Wang, H. (2014). Da shuju shidai de daxue yingyu xiezuojiaoxue gaige [Reform in the teaching of college English writing in the big data era]. *Xiandai Yuancheng Jiaoyu Yanjiu [Modern Distance Education Research]*, 3, 66–72, 86.
- Wang, S., & Wang, H. (2011). Woguo gaoxiao daxue yingyu jiaoxue xianzhuang diaocha jidaxue yingyu gaige yu fazhan fangxiang [On the state of college English teaching in China and its future development]. *Zhongguo Waiyu [Foreign Languages in China]*, 8(5), 4–11, 17.
- Watson Todd, R., Khongput, S., & Darasawang, P. (2007). Coherence, cohesion and comments on students' academic essays. *Assessing Writing*, 12(1), 10–25.
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379–394.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge, England: Cambridge University Press.
- Weissberg, R. (2006). Scaffolded feedback: Tutorial conversations with advanced L2 writers. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 246–265). Cambridge, England: Cambridge University Press.
- Wen, Q., & Johnson, R. K. (1997). L2 learner variables and English achievement: A study of tertiary-level English majors in China. *Applied Linguistics*, 37(1), 27–48.
- Wen, Q. (2009). *Xuexizhe kekong yinsu yu yingyu chengji de guanxi [A study of L2 Learner variables and English achievement]*. Xi'an, China: Shanxi Normal University Press.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, 3(2), 159–177.
- Wickens, C. D. (2007). Attention to the second language. *International Review of Applied Linguistics in Language Teaching*, 45(3), 177–191.

- Wigglesworth, G., & Storch, N. (2012). What role for collaboration in writing and writing feedback. *Journal of Second Language Writing*, 21(4), 364–374.
- Williams, J. (2004). Tutoring and revision: Second language writers in the writing center. *Journal of Second Language Writing*, 13(3), 173–201.
- Wolfe, C. R., Britt, M. A., & Butler, J. A. (2009). Argumentation schema and the myside bias in written argumentation. *Written Communication*, 26(2), 183–209.
- Yan, R. (2010). Eryu xiezuo xiugai de geti chayi ji qi yu gongzuo jiyi de guanxi [The relationship between the individual differences in SLW revision strategies and working memory]. *Waiyu Jiaoxue [Foreign Language Education]*, 31(6), 55–59.
- Yang, C. (2016). *The reactivity and veridicality of concurrent verbal reports in second language writing*. Unpublished Doctoral thesis, The University of Auckland, Auckland, New Zealand.
- Yang, X., & Dai, Y. (2015). Jiyu pidaiwang de daxue yingyu zizhu xiezuo jiaoxue moshi shijian yanjiu [An empirical study on college English autonomous writing teaching model based on www.pigai.org]. *Waiyu Dianhua Jiaoxue [Computer-assisted Foreign Language Education]*, 2, 18–24.
- Yeh, S-W, & Lo, J-J. (2009). Using online annotations to support error correction and corrective feedback. *Computers & Education*, 52(4), 882–892.
- Yin, J. (1995). *Effects of revision strategy instruction on ESL college students' ability to improve the quality of their writing in English as a second language*. (Unpublished doctoral thesis). State University of New York, Buffalo, USA.
- Yu, L., Qi, P., & Guo, P. (2012). Yingyong shuzihua xiezuo pingtai ganyu daxue yingyu xiezuo jiaoxue de yanjiu [Teaching intervention in English writing on the basis of a digital writing system]. *Waiyu Dianhua Jiaoxue [Computer-assisted Foreign Language Education]*, 4, 19–23.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19(1), 79–101.
- Zeng, L., & Li, J. (2014). Xueshu yingyu xiezuo duomotai jiaoxue yanjiu yu yingyong [Multimodal approaches to the teaching of English academic writing]. *Beijing Keji Daxue Xuebao (Shehui Kexue Ban) [Journal of University of Science and Technology Beijing (Social Sciences Edition)]*, 30(6), 31–36.
- Zhang, L. (2013). Donghua daxue benkesheng shiyong zuowen pigaiwang pingyue yingyu zuowen de fankui yu sisuo [Feedback and reflections: A study on the use of www.pigai.org by Donghua University undergraduates]. *Yuwen Xuekan (Waiyu Jiaoyu Jiaoxue) [Journal of Linguistics (Foreign Language Teaching)]*, 8, 107–109.

- Zhang, L. J. (2010). A dynamic metacognitive systems account of Chinese university students' knowledge about EFL reading. *TESOL Quarterly*, 44 (2), 320–353.
- Zhang, L. J. (2013a). Perspectives on second and foreign language learning, teaching, and research: What are the foci? *Dangdai Waiyu Yanjiu [Contemporary Foreign Languages Studies]*, 20(12), 1–4.
- Zhang, L. J. (2013b). Second language writing as and for second language learning. *Journal of Second Language Writing*, 22(4), 446–447.
- Zhang, X. (2006). Daxue yingyu xiezuojiaoxue xianzhuang zhi diaocha [Teaching college English writing: The status quo]. *Waiyu Jie [Foreign Language World]*, 5, 28–32.
- Ziv, N. D. (1982, March). *What she thought I said: How students misperceive teachers' written comments*. Paper presented at Conference on College Composition and Communication, San Francisco, CA. ERIC Document 215 361.
- Ziv, N. D. (1984). The effect of teacher comments on the writing of four college freshmen. In R. Beach & L. S. Bridwell (Eds.), *New directions in composition research* (pp. 362–380). New York, NY: Guilford.

APPENDIX A: ETHICS DOCUMENTS

PARTICIPANT INFORMATION SHEET

(Dean)

Project title: The Effects of Teacher Written Feedback and Revision-focus Direction on Revision and Writing Improvement: A Study of Chinese University Student Writers of English-as-a-Foreign-Language (EFL)

Researcher introduction

My name is Hua Geng, a PhD candidate in the School of Curriculum and Pedagogy, Faculty of Education, the University of Auckland. I am conducting a research project as part of my doctoral thesis.

Project description and invitation

The objective of my research is to explore the effects of teacher written feedback and revision-focus direction on Chinese university students' revision and improvement in English writing. The findings are expected to shed light on EFL students' revision in writing and to provide suggestions for effective pedagogic intervention in EFL students' revision and writing. Because of the high quality of University's EFL teaching and your department's emphasis on EFL writing instruction, students in your department will be kindly invited to participate. Their approval and co-operation are the prerequisite to conducting this project. A detailed description of the research is explained as follows.

This study will comprise three stages: a treatment stage as well as pre- and post-treatment stages, all of which will be conducted out of class as something additional to the regular teaching. It will be conducted on 12 days over 6 weeks. In the pre-treatment stage the researcher will first evaluate participants' autonomous revision and English writing proficiency using a writing-revising task. The participants will also fill in a background questionnaire which will take a couple of minutes. In the treatment, the participants will be assigned into 4 treatment groups and 1 control group and then complete 4 writing-and-revision tasks. The researcher will give treatment groups written feedback on their first drafts and revision-focus manipulation (i.e., whether directing the participants' attention to content and organisation first and to language later), while the control group will not receive any feedback or revision-focus direction. Finally, in the posttreatment stage all the participants will write a new piece of argumentation and revise it with no intervention. The writing of each first draft will take around 35 minutes, and each revising session will also take about 35 minutes. In total, each participant will give nearly 7 hours to the entire research. At the completion of the research all the treatment and control groups will receive adequate feedback on their performance.

Participants' rights

Participants in this study are completely voluntary and entitled to withdraw either before or during the research, and they also have the right to withdraw part or all of

their data without giving any reason within up to 3 weeks after the research is completed. The collected data will only serve for academic and educational purposes. I want to ask for your assurance that students' participation or nonparticipation will not affect their grades or their relationship with the university.

Anonymity and confidentiality

In this study information of the university and students' names will be disguised. No identifiable information will be revealed to a third party. During the stage of data analysis, another teacher will assist me in analysing the written texts produced by participants, and she will be concealed from the information of the university and students.

Data management

The collected Consent Forms and hard copy data will be securely stored in a locked cabinet at the University of Auckland, and electronic data will be stored confidentially in the researcher's computer. After 6 years, all hard copy data will be shredded, and the digital information will be deleted. The collected data will be primarily presented in the researcher's PhD thesis and may also be used for future academic publications or conference presentations. If you would like to have a copy of the final research findings, please indicate this on the Consent Form.

If you have any questions or queries about my research or need to know more about this letter, please feel free to contact anyone in the following list.

<p>Hua Geng (Researcher) + 64 9 373 7599 ext. 48173 (New Zealand) + 86 13915981072 (China) hua.geng@auckland.ac.nz</p>	<p>Professor Lawrence Zhang (Supervisor) + 64 9 623 8899 ext. 48750 lj.zhang@auckland.ac.nz</p>
<p>Dr Rosemary Erlam (Cosupervisor, Department of Applied Language Studies and Linguistics) + 64 9 373 7599 ext. 87081 r.erlam@auckland.ac.nz</p>	<p>Professor Judy Parr (Head of School of Curriculum and Pedagogy) + 64 9 623 8899 ext. 88998 jm.parr@auckland.ac.nz</p>

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 ext. 87830/83761. Email: humanethics@auckland.ac.nz

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 22 May 2014 for (3) years, Reference Number 011857.

CONSENT FORM

(Dean)

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project title: The Effects of Teacher Written Feedback and Revision-focus Direction on Revision and Writing Improvement: A Study of Chinese University Student Writers of English-as-a-Foreign-Language (EFL)

Researcher: Hua Geng

I have read the Participant Information Sheet and understood the nature of the research and why I have been invited to participate. I have had the opportunity to ask questions and have them answered. I agree to assist this study and understand that my participation is voluntary.

- I agree to allow students in our English teaching programme to participate in this research. I agree that the researcher accesses students on site.
- I give my assurance that participation, nonparticipation or withdrawal of students will not affect their grades or relationship with the university.
- I understand that the research will be conducted out of class on 12 days over 6 weeks as something additional to the regular teaching.
- I understand that the research will involve writing-revising tasks, written feedback and revision-focus direction, and a questionnaire. Each writing session will take around 35 minutes, and so does each revising session. The questionnaire will take a couple of minutes.
- I understand that the treatment and control groups will receive different treatment, and that all of them will receive feedback on their performance at the completion of the research.
- I understand that students can withdraw their participation at any time before and during the research and withdraw their data within up to 3 weeks after the data collection is completed.
- I understand that the hard copy data and digital data will be stored separately and securely for a period of 6 years and then destroyed permanently.
- I understand that students' names will not be used in the final report on this project, so no one will know who the students are or their performance on the tasks.
- I understand that the data collected from the research will be used for the researcher's PhD thesis and may be used for academic publications and conference presentations.
- I understand that the information about the university will be disguised.
- I understand that a third party who has signed a confidentiality agreement will analyse the written texts produced by the participants.
- I understand that no identifiable information will be disclosed to a third party or the public.
- I wish to receive a copy of the research findings by email _____

Signature: _____

Name: _____ Date: _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS
ETHICS COMMITTEE ON 22 May 2014 for (3) years, Reference Number 011857.

PARTICIPANT INFORMATION SHEET

(Students)

Project title: The Effects of Teacher Written Feedback and Revision-focus Direction on Revision and Writing Improvement: A Study of Chinese University Student Writers of English-as-a-Foreign-Language (EFL)

Researcher introduction

My name is Hua Geng, a PhD candidate in the School of Curriculum and Pedagogy, Faculty of Education, the University of Auckland. I am conducting a research project as part of my doctoral thesis.

Project description and invitation

The objective of my research is to explore the effects of teacher written feedback and revision-focus direction on Chinese university students' revision and improvement in English writing. The findings are expected to shed light on EFL students' revision in writing and to provide suggestions for effective instruction in EFL students' revision and writing. As qualified research participants, you are kindly invited to participate. Your agreement and co-operation are the prerequisite to conducting this project. A detailed description of the research is explained as follows.

This study will comprise three stages: a treatment stage as well as pre- and post-treatment stages, all of which will be conducted out of class as something additional to the regular teaching. It will be conducted on 12 days over 6 weeks.

In the pretreatment stage the researcher will first learn about participants' English writing proficiency and their autonomous revision using a writing-revising task. The participants will also fill in a questionnaire with their background information, which will take a couple of minutes. This questionnaire, written in Chinese, is anonymous as participants are not required to write down their names or student numbers on it.

In the treatment, the participants will be assigned into 4 treatment groups and 1 control group and complete 4 writing-revising tasks. The researcher will give the treatment groups different types of written feedback and revision-focus manipulation (i.e., whether asking participants to revise content and organisation first and to revise language later), while the control group will not receive any feedback or revision-focus direction.

Finally, in the posttreatment stage all the participants will write a new piece of argumentation and revise it with no intervention. The writing of each first draft will take around 35 minutes, and each revising session will also take about 35 minutes. In total, each participant will give around 7 hours to the entire research. At the completion of the research all the treatment and control groups will receive adequate feedback on their performance.

Please consider the above procedure and decide whether you agree to participate in the writing tasks and to complete the questionnaire on your background. If you agree to participate in part or both of the above activities, please sign the Consent Form accordingly.

Participants' rights

Participants in this study are completely voluntary and entitled to withdraw either before or during the research, and they also have the right to withdraw part or all of their data without giving any reason within up to 3 weeks after the data collection is completed. The collected data will only serve for academic and educational purposes. The Dean has given his assurance that students' participation or nonparticipation will not affect their grades or their relationship with the university.

Anonymity and confidentiality

In this study information of the university and students' names will be disguised. No identifiable information will be revealed to a third party. During the stage of data analysis, another teacher will assist me in analysing participants' written texts, and she will be concealed from the information of the university and students.

Data management

The collected Consent Forms and hard copy data will be securely stored in a locked cabinet at the University of Auckland, and electronic data will be stored confidentially in the researcher's computer. After 6 years, all hard copy data will be shredded and the digital information will be deleted permanently. The collected data will be primarily presented in the researcher's PhD thesis and may also be used for future academic publications or conference presentations. If you would like to have a copy of the final research findings, please indicate this on the Consent Form.

If you have any questions or queries about my research or you need to know more about this letter, please feel free to contact anyone in the following list.

<p>Hua Geng (Researcher) + 64 9 373 7599 ext. 48173 (New Zealand) + 86 13915981072 (China) hua.geng@auckland.ac.nz</p>	<p>Professor Lawrence Zhang (Supervisor) + 64 9 623 8899 ext. 48750 lj.zhang@auckland.ac.nz</p>
<p>Dr Rosemary Erlam (Cosupervisor, Department of Applied Language Studies and Linguistics) + 64 9 373 7599 ext. 87081 r.erlam@auckland.ac.nz</p>	<p>Professor Judy Parr (Head of School of Curriculum and Pedagogy) + 64 9 623 8899 ext. 88998 jm.parr@auckland.ac.nz</p>

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 ext. 87830/83761. Email: humanethics@auckland.ac.nz

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 22 May 2014 for (3) years, Reference Number 011857.

CONSENT FORM

(Students)

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project title: The Effects of Teacher Written Feedback and Revision-focus Direction on Revision and Writing Improvement: A Study of Chinese University Student Writers of English-as-a-Foreign-Language (EFL)

Researcher: Hua Geng

I have read the Participant Information Sheet and understood the nature of the research and why I have been invited to participate. I have had the opportunity to ask questions and have them answered. I agree to participate in this study and understand that my participation is voluntary.

- I understand that my participation, nonparticipation or withdrawal will not affect my grades or relationship with the university.
- I understand that I can withdraw my participation at any time before and during the research and withdraw my data within up to 3 weeks after the data collection is completed.
- I understand that the research will be conducted out of class on 12 days over 6 weeks as practice additional to the regular teaching.
- I understand that the research will involve writing-revising tasks, written feedback and revision-focus direction, and a questionnaire. Each writing session will take around 35 minutes, and so does each revision session. The questionnaire will take a couple of minutes.
- I understand that the questionnaire is written in Chinese and intended to learn about my background information.
- I understand that the treatment groups and the control group will have different forms of participation, and that all of them will receive feedback on their performance at the completion of the research.
- I understand that the hard copy data and digital data will be stored separately and securely for a period of 6 years and then destroyed permanently.
- I understand that my name will not be used in the final report on this project, so no one will know who I am or my performance on the tasks.
- I understand that the data collected from the research will be used for the researcher's PhD thesis and may be used for academic publications and conference presentations.
- I understand that a third party who has signed a confidentiality agreement will analyse my written texts.
- I understand that no identifiable information of mine will be disclosed to a third party or the public.
- I wish to receive a copy of the research findings by email _____

(If you agree to participate in part or both activities in this project, please tick the boxes accordingly and sign.)

I agree to participate in the writing tasks.

I agree to complete the background questionnaire.

Signature: _____

Name: _____ Class Number: _____ Date: _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS
ETHICS COMMITTEE ON 22 May 2014 for (3) years, Reference Number 011857.

CONFIDENTIALITY AGREEMENT

(Written texts analysis assistant)

Project title: The Effects of Teacher Written Feedback and Revision-focus Direction on Revision and Writing Improvement: A Study of Chinese University Student Writers of English-as-a-Foreign-Language (EFL)

Researcher: Hua Geng

Supervisors: Professor Lawrence Zhang; Dr Rosemary Erlam

Analysis assistant:

I agree to assist the researcher in analysing the written texts of the participants for the above research project. I understand that the information contained within them is confidential and must not be disclosed to, or discussed with, anyone other than the researcher and her supervisor(s).

Name: _____

Signature: _____

Date: _____

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE ON 22 May 2014 for (3) years, Reference Number 011857.

APPENDIX B: INSTRUMENTS FOR COLLECTING DATA

Questionnaire

Age _____ Gender _____ Major _____

Approximate length of English learning _____ years

Where did you learn English?

Mainland China/Taiwan/Hong Kong/Other _____ (specify)

Had you ever lived in an English-speaking country or region? Yes/No

If yes, for how long? _____ months

Writing topics

1. When college graduates seek jobs, interest should be their top concern because one will not devote himself to his work unless he is doing what he really loves.
2. People should not give alms to beggars because many of them have the ability to make a living.
3. Some people argue that Chinese students should not be required to learn the ancient Chinese language because it has no practical use.
4. The use of social media like Facebook and Wechat should be discouraged because they do more harm than good to people.
5. Some people think that animal testing should be prohibited because it is too cruel to animals.

APPENDIX C: INSTRUMENTS FOR ANALYSING DATA

Ong's (2013, p. 541) rating scheme for quality of ideas in EFL students' essays

Scores	Features
1-3 point	The essay is overall not convincing. The development of ideas is not good. A stand is provided, with two or three main reasons to support the stand, but the reasons are not well- explained or elaborated with examples, reasons, or illustrations.
4-6 points	The essay is overall quite convincing. The development of ideas is quite good. A stand is provided, with at least three main reasons to support the stand. The reasons are well-supported and elaborated by examples, reasons, or illustrations.
7-9 points	The essay is overall very convincing. The development of ideas is very good. A stand is provided, with at least three main reasons to support the stand. The reasons are very well-supported and elaborated by examples, reasons, or illustrations. One or two counter-arguments are proposed with refutations.

Rating scheme for quality of content used in the main study

Argument	6-7 points	A claim is supported by at least three reasons. All the reasons are clearly stated and very well supported and elaborated by examples and reasons. The argument is very convincing.
	4-5 points	A claim is supported by at least two reasons. Some of the reasons are clearly stated and very well supported and elaborated by examples and reasons. The argument is quite convincing.
	2-3 points	A claim is supported by one or two reasons. Some of the reasons are fairly well supported and elaborated by examples and reasons. The argument is not very convincing.
	1 point	A claim is supported by only one reason, but the reason is not very well supported and elaborated with examples or reasons. The argument is not convincing.
Counter-argument	2 points	One or two counterarguments are proposed and well rebutted.

1 point	One or two counterarguments are proposed but not well rebutted.
---------	---

Rating scheme for quality of organisation use in the main study

Overall organisation, macrostructure	3 points	The essay is very well-organised. It has distinctly included an introduction, a body, and a conclusion. The introduction, the body, and the conclusion are all very well-organised.
	2 points	The essay is quite well-organised. It has basically included an introduction, a body, and a conclusion. The introduction, the body, and the conclusion are quite well-organised.
	1 point	The essay is overall not well-organised. It has not included all of the three elements - an introduction, a body, and a conclusion. The introduction, the body, and the conclusion are not all well-organised.

Cohesion	3 points	It has very logical sequencing. The writing flows. It has used cohesive devices very effectively.
	2 points	The sequencing is mainly logical. The writing generally flows. It has used cohesive devices quite effectively.
	1 point	It lacks logical sequencing. The writing is quite choppy. It has not used cohesive devices effectively.

Coherence	3 points	The propositions in the essay are very well connected. There is no break between the ideas.
	2 points	The propositions in the essay are mostly connected. There are one or two breaks between the ideas.
	1 point	The propositions in the essay are mainly disconnected. There are several breaks between the ideas.

Guidelines for coding linguistic errors

1. Count a repeated error each time it appears.

2. Count errors separately, even though they concern a single word. For example, a wrong word in a wrong verb tense is counted as a Wrong word error as well as a Verb error.
e.g., Finally he work it. (two errors; Finally he solved it.) / call me Miss Zhang respective (two errors; call me Miss Zhang respectfully)
3. Do not count an error that is incurred by another error, i.e., count them as one error.
e.g., Their living are no longer problems. (one error; Their living is no longer a problem)
4. Count a wrong phrase choice as a single error.
5. Count a nonidiomatic word choice as a lexical error.
e.g., alive things (living things)
6. Accept both British and American usages.
e.g., sympathise/sympathize; anymore/any more.
7. Do not count punctuation mistakes, except for those that result in fragments or run-ons, wrong form of the possessive case, or those related to restrictive/nonrestrictive relative clauses.
e.g., History is the memory of a nation, language is the carrier of history. (History is the memory of a nation. Language is the carrier of history) / Students mastering the ancient Chinese language is not necessary. (Students' mastering the ancient Chinese language is not necessary.)
8. Do not count extraneous commas, such as commas following "because" and "I think".
e.g., I think, social media should be used appropriately rather than being discouraged.
9. Do not count misspellings or wrongly derived words.
e.g., in the accient time / well-payed
10. Do not count oral/informal English, including words, abbreviations, and idioms.
e.g., 'cause, wanna, etc.
11. Do not count errors in proper nouns.
e.g., Kongzi (Confucius) / Nan Jing (Nanjing)
12. Do not count errors in capitalisation.
e.g., In my opinion, Animal testing should be forbidden.
13. Accept the use of "and", "but", "yet", and "so" at the beginning of a sentence.
14. Do not bother to consider whether an error results from a slip of the pen because it is hard to distinguish a slip of the pen from the otherwise.

15. Count not only a wrong word as a lexical error but also an expression that contains redundant vocabulary or lacks necessary vocabulary.
e.g., in a total word (in a word) / a piece of sentence (a sentence) / almost every student learns ancient Chinese language during his school life (... during his or her school life)
16. Think twice before judging a word/phrase as a wrong word. Try considering alternative intentions of the student writers.
17. If a preposition or pronoun error makes a sentence flawed in meaning but does not make it ungrammatical, count it as a wrong word instead of a pronoun or preposition error.
e.g., I will never agree on it. (I will never agree with it.)
18. Resort to the following references when necessary: 1) Longman Dictionary of Contemporary English, the online version: <http://www.idoceanline.com>; 2) BYU-BNC: BRITISH NATIONAL CORPUS <http://corpus.byu.edu/bnc/>

Yang's (2016) scheme for coding linguistic errors

1. Errors were counted separately, even though they concerned a single word. For example, spelling errors were counted together with other types of errors concerning the misspelt words.
 - a. I hardly read literitures ** and only write* essays on* writing class. (literature, wrote, in)
 - b. But from that time to now, my math grades has* never became* lower. (have, become)
 - c. The teachers are all modest and have a good knowledge of their major* field* (field of expertise, one error concerning the choice of the word major and the other concerning the modification).
 - d. My father negotiated** with my mother these days that he was* willing to travel around the world when they retired* (diction, tense, is, retire).
 - e. You said* "Working hard would delate** my disadvantage." (, offset)
 - f. [In middle school] I study** a lot of knowledge. (learned)
2. The missing comma before a coordinating conjunction (i.e., and, but, or, nor, for, so, yet) joining two independent clauses and that between a dependent clause placed at

the beginning of a sentence and the independent clause was not counted as an error unless confusion may be caused, neither was a comma that existed.

- a. My past friends start* their own life* but we always get together in the holiday.
(have started, lives)
 - b. I really like the knowledge I study** now and I hope they* can help me find jobs* in the future. (am learning, it, a job)
 - c. And I will try my best to help him/her for I treat his/her dream as my own dream.
 - d. Now, I* conscious about* that only if we do* hard in* our goals, can we success* in the future. (am, conscious, work/study, towards/for, succeed)
3. Both commas and missing commas after sentence-initial prepositional phrases, adverbs such as at present, surprisingly, thus, therefore, moreover, so, sometimes, now, then, today, evidently, of course, occasionally, soon, on the other hand, instead, and furthermore, and the conjunction but were accepted, unless confusion was caused.
- a. At that time I just realized that what he has* now* should be attributed to his endeavour* and perseverance but not * luck or talent. (had, then, spelling, ^his)
 - b. Moreover she will be still happy toward* her life and work, (preposition)
4. The missing comma following however, For example, What's more, and first, that in 4-digit or longer numbers, and that before a direct quotation was counted considering that confusion would be caused and that the use of a comma in such cases has been repeatedly emphasized in China's English classrooms.
- a. For example* my English writing is very terrible. (comma missing)
 - b. What's more* *He will never be ambitious. (comma missing, capitalization)
 - c. Now I an* in Xian* XXX University, 1000* miles away from my hometown.
(spelling, Xi'an, 1,000)
5. Comma errors related to restrictive/non-restrictive relative clauses should be counted.
6. Extraneous commas should also be considered errors. For example, Because followed by a comma should be counted as an error (but not but).
- a. *Because,* he used to drink wine and eat friend* food too much*. (fragment, extraneous comma, fried, order of the noun modifier)
 - b. But, she doesn't* like me. (didn't)
 - c. So I think,* the father's love is differ* from *mother's love, (extraneous comma, different, ^article)
 - d. Maybe,* I can get acqutance* of* more person*, (extraneous comma, spelling, with, persons/people)

- e. Maybe, * I'll be a millioniar*, but I'll never lost* mine* away*. (extraneous comma, spelling, lose, myself/my goal, redundant preposition)
 - f. I hope*, I can be a better person in the future. (extraneous comma)
7. A full stop that was used where a comma should have been used was counted as an error, with great discretion, so was a missing full stop at the end of a sentence.
 8. A comma that was used where a full stop should have been used was counted as an error, with great discretion.
 - a. My father was a polite person, and he seldom got angry,* He always educated me in a more soft* way, so I liked to stay with him at the time. (full stop, soft)
 - b. That * all of mine,* It's for you, also *to me. (no finite verb, full stop, for)
 - c. My father was born in He Nan* province,* He can say* * He Nan* language* until* now. (Henan, full stop, speak, ^the, Henan, dialect, even)
 9. Tense/reference errors were coded based on contexts. The seemingly inconsistent present tense usages in between or at the end of a series of past events were not considered errors if they were used to indicate how one feels about his/her past at present.
 - a. Every day time* he tried his best to raise* more money, but at night, we whole family would get together to watch TV, which thing* is* his favorite. Ohe*, I forget to say,* he also likes* to play chess since he was very young,* And he can play it very well. When he was free, I mean, relative* free.* He would play it with his old friends. (the next paragraph) Now my family are living better and better. (Every day/In the daytime, earn, which, was, Oh, extraneous comma/other bigger corrections, tense/other bigger corrections, full stop, relatively, comma)
 10. Accept both British and American usages (e.g., colourful, any more & anymore).
 11. Oral/Informal English, including words, abbreviations, and idioms ('cause, wanna, etc.), was not counted.
 - a. We always shotted* during class and shotted* louder after class.
 - b. And I think I will take good care of him, just like he looked after me when I was young.
 - c. I hung out with a girl but our love ended in failure.
 - d. Therefore, though unconsciously, I became kind of proud.
 - e. Days were unbearable when I stepped back into my city of memories, 'cause no time left* for homework. (was left)
 12. Article errors were counted.

13. Errors in proper nouns, especially errors in proper nouns or others nouns in the Chinese pinyinized forms, were counted (e.g., Shan Dong should be Shandong).
14. Chinese characters were counted as errors; blanks (left for words) were counted.
- a. I didn't let him down and was accepted by a famous university, which made him more * ,
15. Errors in capitalization were counted.
- a. Playing *chinese chess with old friends and reading books will be good choices to kill time.
16. Only the last of a series of run-on sentences was considered right.
- a. At class, the teachers give me lots of knowledge,* however, I can't understand what they said*. (say)
- b. I went to school*, he told me the news.
- c. I listning** to the* rock music everyday**, isn't it a happy college life? (listen, article, every day, run-on)
- d. Apart from studying, I hardly go to play basketball*, it accouts* to* * that now I have * relaxing ways,* I could go to a cinema or a KTV and so on. (run-on, misspelling, is due to, ^the fact, ^other, run-on)
- But the following was not considered a run-on sentence:
- a. We don't have a teacher, we don't have experience, but we all like acapella than* anything*. (more than, ^else)
17. Several cases where it was hard to judge full stops from commas in students' writings when such punctuations were followed by lower-case words were tolerated.
- a. But things changed. since my father left us to support my family.
18. A sentence fragment was coded as an error (the first capitalized letter of the fragment or the body sentence was not counted as an additional error).
- a. Our holiday of *past was truly have* fun. *Even though there were* a lot of holiday work waitting* for us. (^the, redundant have, fragment, was, spelling)
- b. Even though we struggling* through study day and night*. It's real life with * bitter and sweet taste. (struggled, fragment, ^a)
- An independent fragment was also counted as an error.
- a. Some times* *nervous. (Sometimes, no subject and finite verb)
- b. Gree* trees, * small house, my old father, * shiny sun and my father's sweaty* smile*. (Green, ^a, ^the, sweet, fragment)
19. A dangling modifier was counted as an error.

- a. *Having entered universities, our lives have been changed.
20. The number of errors in very rare cases was decided in reference to a correction that minimized the number.
- a. It* looks like* the same between* the past of me* and now, but *the attitude and thinking really make a difference*. (ambiguous reference, redundancy, wrong parallelism, my past, my (attitude), wrong phrasal usage, in reference to My life looks the same from the past to now, but my attitude and thinking are really different.)
21. Misuse or wrong choice of words or phrases was counted.
- a. Numerous books stuff * the whole bookshelf in his room.
- b. The thinking of the sages shocks* me deeply.
- c. Now I can see my future: reading books, enjoying a common job, having an ordinary wife and children, writing articles and books, * enjoying the thinking world I fabricate* for myself. (^and, diction)
- d. I will make* *future to repay my parents and have a fortunate family. (verb use, determiner, fortunate was tolerated)
- e. In class, I also confront* some strange students. (verb use)
- f. In my eyes, my father always* showed up* as an* soldier. (spelling, verb phrase use, article)
22. As may have been shown, some expressions that might not be idiomatic but were grammatical and intelligible were tolerated.
- a. There are some persons in the world thinking too much*! (persons tolerated, so profoundly)
- b. The changes are also in our bodies.
23. Errors related to collocations were counted once, and language points related to one error that should be correct in reference to the error were exempted.
- a. lay more attention on*
- b. He will become more wisdom*, more happy, and healthy. (wiser should have been used, but only one error was counted)
- c. Another different area* between college and high schools is the course* you study *which is more difficult. (aspect, courses, non-restrictive, is in the subordinate clause was no longer coded as an error)
24. Some redundancies (the italicized, in the following) were tolerated.
- a. In the past time, he had * opposite opinion. (an)

- b. He always keep* a good attitude* towards the things that are bad in his life.
(keeps, spelling)
 - c. It makes me feel very sad that *mother's parents died when she was 20 years old.
(^my)
25. Some China English forms (the italicized) were also tolerated.
- a. A fly without a head can present* my near* situation. (represent/describe, recent/present)
26. Though rarely seen, repeated errors were counted repeatedly.
- a. my love for him is becoming deeply** and deeply** day by day. (deeper, deeper, two errors counted for each, one for the misused adverb and the other for not using the comparative degree)
27. Misspellings were counted.
- a. And it's not hard for me to guess what his future life will be life*. (like)

Analysis of between-draft revisions

1. Guidelines for identifying between-draft revisions
 - a. Count a repeated change each time it appears.
 - b. Count a change even if it is incurred by another one. In other words, count them as separate changes. (e.g., Only in this way, will you devote yourself to your career. → You will devote yourself to your career only in this way.)
 - c. If a sentence boundary change also involves a change of a word or phrase, count them separately.
 - d. Count changed words separately if they can be divided into independent changes, even if they are adjacent. (e.g., their living places and costs → their habitation and income)
 - e. Count changed words as a single change if they cannot be divided into independent changes. (e.g., choose to be a housewife → reject opportunities at the workplace)
 - f. If a change involves more than one consecutive T-unit, count it once for each T-unit involved. For example, if a newly added conclusion paragraph contains 4 T-units, count them as 4 discourse changes.

- g. Do not count it if it is a change in punctuation or spelling; however, count it as a linguistic change if the change results in or corrects a grammatical error.
 - h. Do not count a change in capitalisation.
2. Guidelines for classifying between-draft revisions by textual level
- a. Count an addition, deletion, or replacement of a word, phrase, clause, or sentence as a linguistic change if it does not change the originally intended meaning in essence. In other words, do not count if the change merely results in a similar meaning, or makes the original meaning clearer or more accurate or specific, but does not lead to a different concept. (e.g., Girls are expected to be soft. Girls ~~are~~ expected to be gentle. / every student → every Chinese student / ... but not a must ~~but~~ not a must for today's women)
 - b. Count an addition or deletion of a word, phrase, clause, or sentence as a linguistic change if the meaning of the added or deleted part is predictable according to the context.
 - c. Count an addition, deletion, or replacement of a word, phrase, clause, or sentence as a discourse change if it changed the originally intended meaning in essence. In other words, count it if the concept is changed. (e.g., Girls are expected to be soft. Girls ~~are~~ expected to be pretty.)
 - d. Count an addition, deletion, or replacement of a word or phrase a discourse change if it changes the cohesion of the text.
 - e. Classify an addition, deletion, or replacement of a sentence or sentences into the discourse level if it changes the organisation of the text.
 - f. If a sentence is paraphrased into two sentences or if two sentences are combined into one, classify the change into the linguistic level.
 - g. Count reordering of sentences or paragraphs into the discourse level.
 - h. Count the addition or deletion of a thesis statement for the whole essay into the discourse level.
 - i. Count the addition or deletion of a topic sentence for a paragraph into the discourse level.
 - j. Count the addition or deletion of between-paragraph or between-sentence transitions into the discourse level.

APPENDIX D: SAMPLE DRAFTS AND FEEDBACK

The following are drafts of two students on the third writing topic (Some people argue that Chinese students should not be required to learn the ancient Chinese language because it has no practical use) and feedback on them. Student 1 was in the ID–RFD group condition, and Student 2 in the IS+RFD condition.

Student 1’s drafts and the feedback received (ID–RFD)

<div style="border: 1px solid black; padding: 5px; transform: rotate(-90deg); transform-origin: left top;">Intro is to introduce the topic and arouse the interest of readers.</div>	<p>Draft 1</p> <p>Some people may think that it is no use learning the ancient Chinese language. But In my opinion, it is necessary for us to learn and grasp the ancient Chinese language.</p> <p>There are many reasons supporting my opinion. First, ancient <i>wrong word</i> Chinese is a <u>culture</u> which is related to great people in the ancient China. Learning it not only <u>let</u> us know more about 5000 years’ culture the <u>old</u> China have, but also <u>help</u> us <u>attach to</u> great people closely. ①Second, learning ancient Chinese language is not easy. ②Grasping it can help us practice our mind so that we can learn math or physics <u>easier</u>. Last but not <i>An article is needed before a singular countable noun.</i> least, learning ^ ancient Chinese language can connect us to the former people and teach us some knowledge about their life. No doubt, we can learn many like experience from others’ life.</p> <p><i>wrong phrase</i> <u>In total</u>, I believe that learning ^ ancient Chinese language is useful to our study and daily life. Try to learn it, and you will learn much from it.</p>	<div style="border: 1px solid black; padding: 5px; transform: rotate(-90deg); transform-origin: left top;">The connection between two sentences should be logical. Your reasoning should be logical. Can language learning facilitate learning of math and physics? How?</div>
<div style="border: 1px solid black; padding: 5px; transform: rotate(-90deg); transform-origin: left top;">Each paragraph develops ONE idea only instead of several ideas.</div>		

Draft 2

Some people may think that it is no use learning the ancient Chinese language. They believe that there is no need to study what we don’t use today. But in my opinion, it’s necessary for us to learn and grasp ancient Chinese.

There are two reasons supporting my opinion. First, ancient Chinese is a cultural heritage related to great people like Kongzi in ancient China. Learning it not only lets us know more about their life but also helps us attach ourselves to them closely. Only in this way can we grasp their spirits more easily and regard them as a good example to study.

Second, learning ancient Chinese can help us know more about Chinese 5000 years' culture such as how some folk customs came into being, when the festival began and so on. No doubt, Chinese culture is profound and worth studying.

In one word, I believe that learning the ancient Chinese language is useful to our study and daily life. Try to learn it, and we will learn much from it.

Student 1's drafts and the feedback received (IS+RFD)

Draft 1

As

^ Time goes by, it seems that the ancient Chinese language is increasingly useless for us teenagers now. While many people point out that the ancient Chinese language has no practical use in modern society, I still treat it as an important course for students. The reasons are as ^{follows} followed.

Some will argue that we can read translated versions of ancient Chinese works. You need to rebut such ideas.

① First of all, learning the ancient Chinese language can help us ⁱⁿ when reading books of ancient China. Without ^{its} it's help we can hardly understand even a ~~piece~~ of sentence. ② Second, reading ⁱⁿ ^ the ancient Chinese language can improve our understanding level and respect to ancestors. ③ Last but not ~~the~~ least, learning ⁱⁿ to read or even write ^ the ancient Chinese language is one of the most useful ways to help us understand Chinese ancient culture.

Your body part has 3 ideas. Put them into 3 paragraphs. Develop each of them.

Explain idea ②. For example, we learn the unique vocabulary and sentence structures from abundant reading in ancient Chinese. Explain idea ③. For example, ancient writings, e.g. poems, recorded various aspects of ancient life.

For these reasons, what I believe without a shadow of doubt is

that only by learning the ancient Chinese language well can we
say it with *pride* proud that we are Chinese. I think it definitely that
Chinese students should be required to learn the ancient
Chinese language.

Draft 2

As time goes by, it seems that the ancient Chinese language is increasingly useless for us teenagers now. While many people point out that the ancient Chinese language has no practical use in modern society, I still treat it as an important course for students to learn. The reasons are as follows.

First of all, learning the ancient Chinese language can help us when we read books of ancient China. Without its help we can hardly understand even a single sentence. Ancient works can not be exactly transformed into the modern Chinese language. Translated versions sometimes contain translators' views, so it is better to read ancient books written in ancient Chinese.

Second, reading ancient language can improve our understanding level and respect to ancestors. For example, from reading Lao Zi's book we can get the wisdom about nature and society; from reading Xin Qiji's poems we can understand his honest to his country, etc.

Last but not least, learning to read or even write in the ancient Chinese language is one of the most useful ways to help us understand Chinese ancient culture, because Chinese history and culture are all recorded in the form of the ancient Chinese language.

For these reasons, what I believe without a shadow of doubt is that only by learning the ancient Chinese language well can we say it proudly that we are Chinese. I think definitely that Chinese students should be required to learn it.

APPENDIX E: SCORES IN TASK 1 AND TASK 5

Table D1 Mean first- and second-draft scores in Task 1 (*SDs* put in parentheses)

Group	Draft	Content	Organisation	Grammatical accuracy	Lexical accuracy
ID+RFD (<i>n</i> = 15)	D1	5.27 (1.16)	5.07 (0.70)	5.69 (1.77)	2.84 (1.13)
	D2	5.53 (1.13)	5.47 (0.92)	5.20 (1.61)	2.54 (1.00)
ID–RFD (<i>n</i> = 16)	D1	5.31 (0.87)	5.31 (0.79)	5.76 (1.71)	2.81 (1.31)
	D2	5.50 (1.03)	5.75 (0.86)	5.23 (1.45)	2.61 (0.88)
IS+RFD (<i>n</i> = 15)	D1	5.47 (0.83)	5.13 (0.83)	5.51 (1.75)	2.84 (0.93)
	D2	5.60 (1.06)	5.47 (0.83)	4.84 (1.47)	2.55 (1.32)
IS–RFD (<i>n</i> = 14)	D1	5.50 (1.02)	5.43 (0.85)	5.52 (1.69)	2.62 (1.35)
	D2	5.71 (0.83)	5.64 (0.93)	5.01 (1.07)	2.32 (0.97)
Control (<i>n</i> = 17)	D1	5.35 (1.00)	5.29 (0.77)	5.82 (1.06)	2.84 (1.17)
	D2	5.59 (0.94)	5.59 (0.71)	5.29 (0.86)	2.63 (1.12)

Table D2 Mean second-draft scores in Task 5 (*SDs* put in parentheses)

Group	Content	Organisation	Grammatical accuracy	Lexical accuracy
ID+RFD (<i>n</i> = 15)	6.67 (0.72)	6.87 (0.74)	4.23 (1.09)	2.57 (0.74)
ID–RFD (<i>n</i> = 16)	6.13 (1.20)	6.56 (0.89)	4.80 (1.01)	2.35 (0.72)
IS+RFD (<i>n</i> = 15)	6.40 (0.74)	6.67 (0.82)	4.71 (1.55)	2.39 (0.99)
IS–RFD (<i>n</i> = 14)	6.29 (0.91)	6.50 (1.02)	4.70 (1.04)	2.38 (0.98)
Control (<i>n</i> = 17)	5.65 (0.70)	5.71 (0.69)	5.36 (0.91)	2.64 (1.04)