



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Nishikawa, Mikako

Title:

Test-takers' Cognitive Processes While Synthesizing Multiple Texts and Graphs

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

**Test-takers' Cognitive Processes While Synthesizing
Multiple Texts and Graphs**

Mikako Nishikawa

School of Education
University of Bristol

September 2018

Test-takers' Cognitive Processes While Synthesizing Multiple Texts and Graphs

Abstract

Integrated writing is becoming more common than traditional topic-based writing in the field of second language assessment. This is partly because essay scores from integrated writing tests are believed to give a better prediction of how students perform in real-life academic writing tasks. There has been a rise in the number of publications on integrated writing concerning its construct validity, discourse types, the effects of borrowing from source texts, and test-taking strategies. These researchers have suggested the need for a careful consideration of the factors affecting the integrated writing process, especially when graphs are used. Incorporating information from graphs into integrated writing demands an additional cognitive skill set. Few researchers have addressed the effects of information from graphs in integrated writing, and even fewer studies have looked at the effects of the combination of text and graphs. The current study aimed to gain insights into the cognitive processes of L2 writers when both multiple source texts and graphs were used as prompts. It attempted to explain (1) the effects of source texts, (2) graph features, (3) the roles of reading in integrated writing, and (4) test-taking strategies during the integrated writing tasks. This study adopted an eye-tracking method to investigate the eye movements of the test-takers (N=38) that reflected their behaviours and decision-making processes. It used multiple approaches that combined eye-tracking, questionnaires and focus group discussions. The findings of the quantitative analysis suggest that language proficiency played a major role in fulfilling the task requirements. The first ten minutes of eye-movement recordings showed some crucial differences between upper intermediate and lower intermediate level participants. Qualitative analyses using the gaze-plots in timed segments, students' written outputs, and the questionnaires, helped understand test-takers' behaviours while they were making decisions during each stage of the integrated writing task. Finally, the implications for the test-developers, teachers and students were discussed.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

Acknowledgements

It would not have been possible to complete this thesis without the enormous help and advice I received from many people, to whom I wish to express my deep appreciation.

I would like to express my deepest gratitude to both my supervisors for their understanding and generous support, particularly Dr Guoxing Yu's invaluable expertise in language assessment and eye-tracking studies and Dr Liz Washbrook's wisdom in statistics. Without their incredible guidance, I would not have reached this amazing milestone of completing a thesis of my dreams.

To my examiners Dr. Tineke Brunfaut and Dr. Elena Hoicka, I would like to thank them both for giving me all the insightful feedback and suggestions to improve my dissertation via the Viva exams.

I would also like to thank Professor Stephen Bax of Applied Linguistics at the Open University who, although no longer with us, continues to inspire the community of applied linguistics which he served with dedication over the course of his career.

I would like to thank both Dr Nina Kazanina and Dr Talia Isaacs at the University of Bristol for their constructive feedback and critiques of this research from an early stage of writing, and both Dr Barry O'Sullivan and Dr Jamie Dunlea from the British Council for their invaluable input and advice at The Language Testing Forums in the United Kingdom.

Also, I would like to show my gratitude to Professor Akira Tajino, the Director of the International Academic Research and Resource Centre for Language Education at Kyoto University for his academic input and constructive feedback at The Japanese Association of College English Teachers Conferences in Japan.

My deep appreciation also goes to the devoted English teachers in Japanese high schools, namely Ms. Mika Iwasaki of Murasakino High School in Kyoto city, Ms. Miho Izumi of Kobe University Secondary School in Kobe city, Mr. Kenji Moriwaki of Kinran Senri Senior High School in Suita city, Mr. Kenzo Yoneda of Hagaromo Gakuen Senior High School in Takaishi City, as well as their incredible students, for their participation in this study.

In addition, I am indebted to my colleagues who supported me, including Mr. Shunken Shiozaki, Mr. Michael Todd Fouts, Mr. Daniel Joyce of the Eiken Foundation of Japan, Tokyo office, as well as Mr. Yoshihiko Taira of the Osaka office, for their generous support and encouragement. Also, I'd like to thank two anonymous raters at the University of Bristol, who helped to score the TEAP essays for further analysis.

Furthermore, many thanks are extended to the British Council for awarding this research funding to make it possible to understand students' language proficiency with the Aptis tests, to the Eiken Foundation of Japan for financing the leasing of the eye-tracking device to study their test product, TEAP, as well as to The International Research Foundation for English Language Education (TIRF) for awarding a Doctoral Dissertation Grant (DDG) in 2016. Given these research grants and funding, it was possible to attend international conferences and to

win two awards, namely the Best Poster Prize at the Language Testing Forum held at the University of Oxford in November of 2015, and the Award of Excellent Presentation as a Young Scholar at The Japan Association of College English Teachers' (JACET) 55th International Convention (Sapporo, Japan) 2016.

Finally, I would like to acknowledge my family; my mother Etsuko, my husband, Takehiro Nishikawa and my son Hiroki for their support and encouragement throughout this long journey.

Table

Abstract	i
Acknowledgements	iii
List of Abbreviations and Acronyms	x
List of Figures	xi
List of Tables	xiii
Chapter 1. Introduction	17
1.1 Background	17
1.2 Context of the Study	19
1.3 Purpose of the Study	22
1.4 Significance of the Study	24
1.5 Organization of the Thesis	26
Chapter 2. Literature Reviews	28
2.1 Introduction.....	28
2.2 Language Testing and Validation	29
2.3 Theory of Writing	31
2.3.1 A brief history of second language writing	31
2.3.2 Cognitive processes of writing	32
2.4 Socio-Cognitive Model for Writing.....	33
2.5 Integrated Writing for Assessment	37
2.5.1 Independent writing vs. integrated writing for assessment	38
2.5.2 Integrated writing with source texts for assessment.....	40
2.6 Graph-based Writing.....	43
2.7 Test-taking Strategies used for Integrated Writing	46
2.7.1 Overview of the use of test-taking strategies in L2 assessments.....	46
2.7.2 Test-taking strategies for language assessment.....	47
2.8 The TEAP integrated writing tasks.....	49
2.9 Conclusion	53
Chapter 3 Methodology of the Study	55
3.1 Research Questions	55
3.1.1 Research questions	55
3.2 Research Methods.....	56
3.2.1 Research design and justification of the method approach.	57
3.2.2. Rationale behind the use of eye-tracking technology as an instrument	60
3.2.3 Rationale behind the use of survey.....	63
3.2.4 Rationale behind the use of focus group discussion.....	65
3.3 Participants.....	66

3.3.1	Sampling of participants.....	66
3.3.2	Choosing participants for eye-tracking experiment	66
3.4	Research Instruments	67
3.4.1	(Phase 1) The Aptis test for describing English proficiency of participants	67
3.4.2	(Phase 2) Eye-tracking experiment using a Tobii Eye Tracker.....	68
3.4.3	[Phase 2] Analytical scores of the essays using the rating scales.....	70
3.4.3	(Phase 3) Cognitive Processing Questionnaire	73
3.4.4	(Phase 4) Focus group discussions.....	74
3.5	Methods of Data Analysis.....	75
3.5.1	Eye-movement data.....	75
3.5.2.	Statistical analysis	80
3.5.3	Cognitive processing questionnaires	83
3.5.4	Focus group discussions.....	83
3.6	Theoretical Perspectives and Researcher Positionality.....	85
3.7	Weakness and Strengths of the Chosen Approach.....	88
3.8	Ethics	88
3.8.1	Determining the stakeholders in the research.....	89
3.8.2	Collecting information	89
3.8.3	Maintaining confidentiality	90
3.8.4	Avoiding bias	90
3.9	Conclusions.....	91
Chapter 4. Quantitative Findings from Eye-movements.....		93
4.1	Introduction.....	93
4.2	Background of the Participants (Gender, Age).....	93
4.3	Participants' language proficiency (Aptis test results).....	94
4.3.1	Correlations between reading and writing skills	94
4.4	Eye-tracking movements for analysis	96
4.4.1	Wilcoxon signed-ranks tests in groups of AOIs.....	103
4.4.2.	Wilcoxon signed ranks tests in sub-sections of AOIs	107
4.4.3	Mann-Whitney U tests for groups of AOIs	110
4.4.4	Mann-Whitney U tests for sub-sections of AOIs	112
4.4.5	Multiple regression analysis for groups of AOIs	119
4.5	Summary of Aptis Test and Eye-tracking Results	121
4.6	TEAP Test and Eye-tracking Results	125
4.6.1	The Correlations between Rater 1 and Rater 2 ratings for the five criteria.....	125
4.6.2	Pearson Correlation test (Aptis scores and TEAP analytic scores).....	127
4.6.3	Summary of the TEAP essay ratings and eye-tracking results.....	135

Chapter 5. Findings of Cognitive Processing Questionnaires.....	137
5.1 Computer Literacy	138
5.2 Familiarity with Graphs	141
5.3 Test-taking strategies during different cognitive stages in the writing process	144
5.3.1. Preparing-to-Write	144
5.3.2. Translating and Writing	152
5.3.3. Monitoring & Revising	156
Chapter 6. Findings from Selected Cases	158
6.1 Preparing to Write.....	159
6.1.1 Task Representation	159
6.1.2 Macro-Planning.....	166
6.1.3. Summary of the Preparing to Write stage	171
6.2 Translating and Writing	172
6.2.1 Examples of students' writing.....	172
6.2.2 Test-taking strategies.....	176
6.3 Monitoring and Revising	178
6.3.1 Gaze-plots for the last five minutes.....	178
6.4 Comparing the performance of the same participant between Task 1 and Task 2.....	184
6.4.1 Evidence from the Gaze-plots of the two tests.....	184
6.4.2 Evidence from the written outputs of the two tests	188
6.5 Conclusions.....	194
Chapter 7. Findings from the Focus Group Discussions	197
7.1 Introduction.....	197
7.2 Analysis of the Focus Group Discussions.....	197
7.2.1 Focus group discussions as a methodology.....	197
7.2.2 Participants' inner thoughts revealed in focus group discussions	198
7.2.3 Summary of focus group discussions	213
Chapter 8. Discussion.....	215
8.1 Introduction.....	215
8.2 Research Question (1) Effects of Source Texts	218
8.2.1 Evidence from Eye-tracking data and analysis	219
8.2.2 Evidence from Surveys and Written outputs.....	220
8.2.3 Focus Group Discussions	221
8.3. Research Question (2) Effects of Information from Graphs and (3) Types of Graphs	221
8.3.1 Evidence from Eye-Tracking data and analysis	222
8.3.2 Evidence from Surveys and Written outputs.....	223
8.3.3 Focus Groups Discussions	224

8.4.	Research Question (4) Effects of Test-Taking Strategies	225
8.4.1	Evidence from Eye-Tracking data and analysis	225
8.4.2	Evidence from Surveys and Written Outputs	226
8.4.3	Focus Group Discussions	228
8.5	Research Question (5) Effects of Language Proficiencies	229
8.5.1	Evidence from Eye-Tracking data and analysis	229
8.5.2	Evidence from Surveys and Written outputs	230
8.5.3	Focus Groups Discussions	230
8.6	Conclusions	232
Chapter 9 Conclusions		234
9.1	Introduction	234
9.1.1	Implications from the lack of task familiarity and from the administration of the eye-tracking experiment	234
9.1.2	Implications for students with low reading proficiency	235
9.1.3	Implications for students with higher writing proficiency	236
9.1.4	Implications for teachers	237
9.1.5	Implications for test-developers	238
9.1.6	Implications for researchers	240
9.2	Limitations	241
9.2.1	Sampling	241
9.2.2	Research instruments	242
9.2.3	Research procedures and analysis	242
References		244
Appendixes		254
Appendix A	(Task 1) and Sample Essay (1)	254
Appendix B	(Task 2) and Sample Essay (2)	256
Appendix C	TEAP Writing test, Task B Scoring Rubrics	258
Appendix D	Survey on computer literacy, and graph familiarity	260
Appendix E	Questions for Focus Group Discussions	266
Appendix F	Feedback Sheet for Focus Group Discussions	267
Appendix G	Consent Form	269
Appendix H	Information Sheet	270
Appendix I	Information Sheet (Japanese)	272
Appendix J	GSoE Research Ethic Form	274
Appendix K	Focus Groups Discussions	280
Appendix L	Wilcoxon signed rank tests of differences in eye-movement measurements in Task 1 and Task 2 (All 40 minutes)	287

Appendix M	. Wilcoxon signed rank tests of differences in eye-movement measurements in Task 1 and Task 2 (First 10 minutes).....	292
Appendix N	Wilcoxon signed rank tests of differences in eye-movement measurements in Task 1 and Task 2 (Last 10 minutes)	297
Appendix O	Mann-Whitney tests of difference between participants with lower and upper intermediate skills in reading and writing (40 minutes).....	302
Appendix P	Mann-Whitney tests of difference between participants with lower and upper intermediate skills in reading and writing (First 10 minutes).....	311
Appendix Q	Mann-Whitney tests of difference between participants with lower and upper intermediate skills in reading and writing (Last 10 minutes)	321
Appendix R	Regression results for eye-movement variables in 40 minutes	328
Appendix S	Regression results for eye-movement variables in the first 10 minutes	334
Appendix T	Regression results for eye-movement variables in the last 10 minutes.....	340

List of Abbreviations and Acronyms

AOI	Areas of Interest
BMET	The Beijing Matriculation English Test
CAE	Certificate in Advanced English
CEFR	Common European Framework of Reference for Languages
CPE	Certificate of Proficiency in English
CTT	Classical Test Theory
EFL	English as a Foreign Language
FCE	First Certificate in English
G-Theory	Generalizability Theory
IELTS®	International English Language Testing System
KET®	Key English Test
L2	English as a Second Language
MEXT	Ministry of Education, Culture, Sports and Technology
NCT	The National Centre Test
PET®	Preliminary English Test
TEAP®	Test of English for Academic Purposes
TESOL	Teachers of English to Speakers of Other Languages
TOEFL®	Test of English as a Foreign Language

List of Figures

Figure 1. The Sample TEAP Writing Task B extracted from the Eiken Website.....	25
Figure 2. Weir’s Framework for a Socio-Cognitive Approach for Assessing the Validity of.....	35
Figure 3. Research Methods.....	58
Figure 5. Average Japanese high-school students according to CEFR scale cited in (MEXT, 2014)	68
Figure 6. TEAP Task B Sample Task 1	69
Figure 7. TEAP Task B Sample Task 2	69
Figure 8. Example of a Heat Map	77
Figure 9. Example of a Gaze Plot	77
Figure 10. Areas of Interests for Task 1.....	78
Figure 11. Areas of Interests for Task 2.....	78
Figure 12. Sub-sections of the Areas of Interest for TEAP Sample Task 1.....	79
Figure 13. Sub-sections of the Areas of Interest for TEAP Sample Task 2.....	79
Figure 14. Distribution of CEFR levels for Aptis Reading and Writing Scores	94
Figure 15. Questionnaire answers on Age of First Computer Access (Q1_5).....	138
Figure 16. Gaze-plot for ST030 [Task 1] ST030 (Aptis Reading C / Aptis Writing B2 / TEAP B1, N=1)	162
Figure 17. Gaze-plot for ST037 [Task 1] ST037 (Aptis Reading B1 / Aptis Writing C / TEAP B2, N=1)	163
Figure 18. Gaze-plot for ST053, First 5 min. [Task 1] ST053 (Aptis Reading B1/ Aptis Writing B / TEAP A2-Below, N=1)	164
Figure 19. Gaze-plot for ST033, First 5 min. [Task 1] ST033 (Aptis Reading B2 /Aptis Writing B2 /TEAP B1, N=1).....	165
Figure 20. Gaze-plot for ST030, Last 5 min. [Task 1]N=1.....	180
Figure 21. Gaze-plot for ST037, Last 5 minutes [Task 1]N=1	181
Figure 22. Gaze-plot for ST053, Last 5 min. [Task 1]N=1.....	182
Figure 23. Gaze-plot for ST033, Last 5 minutes [Task 1] N=1	183
Figure 24. ST033’s Gaze-Plots for the First 5 and the Last 5 minutes during Task 1	185
Figure 25. ST033’s Gaze-Plots for the First 5 and the Last 5 minutes during Task 2	185
Figure 26. ST053’s Gaze-Plots for the First 5 and the Last 5 minutes during Task 1	186
Figure 27. ST053’s Gaze-Plots for the First 5 and the Last 5 minutes during Task 2	186
Figure 28. ST030’s Gaze-Plots for the First 5 and the Last 5 minutes during Task 1	187

Figure 29. ST033's Gaze-Plots for the First 5 and the Last 5 minutes during Task 2 187

Figure 30. ST037's Gaze-Plots for the First 5 and the Last 5 minutes during Task 1 188

Figure 31. ST037's Gaze-Plots for the First 5 and the Last 5 minutes during Task 2 188

List of Tables

Table 1. IELTS® /TOEFL iBT® Score Ranking in the Asian Region	20
Table 2. Summary Table of Cognitive Processing across Main Suite Examinations	40
Table 3. Points of Advice and Examples of Common Mistakes in the TEAP Writing Test	52
Table 4. Justification of the Research Methods of the Explanatory Design	59
Table 5. List of questions related to the three stages of integrated writing.....	64
Table 6. Schedule for Eye-tracking Experiments and Survey	70
Table 7. TEAP Writing Task B, Construct and Rating Scales Adopted from Eiken Foundation of Japan website	72
Table 8. Eye-tracking metrics	76
Table 9. The Overview of Data Analysis.....	85
Table 10. Summary of Working Measurements	92
Table 11. Descriptive Statistics of the Aptis Tests	95
Table 12 Pearson Correlation Analysis for Aptis Reading, Writing and Grammar & Vocabulary	95
Table 13. Participants' Reading and Writing Proficiencies (N=40)	96
Table 14. Descriptive Statistics of Eye-tracking Metrics (N=40).....	98
Table 15. Wilcoxon signed-ranks Test Results (Task 1 and Task 2 in broader areas (groups) of AOIs for 40 minutes.....	104
Table 16. Descriptive Statistics of Time Spent on Gaining Information from Different Types of Graphs in Task 1 and Task 2.....	106
Table 17. Wilcoxon signed-rank Tests of Differences in Eye-movement Measurements for Different Graph Types (40 minutes)	106
Table 18. Rates of total fixation duration on sub-sections of AOIs across Task 1 and Task 2 (40 minutes).....	109
Table 19. Ranking of the Means for Wilcoxon signed-ranks Tests	110
Table 20. Significant Results of the Mann Whitney U Test to Compare the Groups of Upper and Lower Groups of Aptis Scores in Groups of AOIs	112
Table 21. Significant Results of the Mann Whitney U Test to Compare the Groups of Upper and Lower Groups of Aptis Scores in Sub-division of AOIs (all 40 minutes)	113
Table 22. Significant Results of the Mann Whitney U Test to Compare the Groups of Upper and Lower Groups of Aptis Scores in Sub-division of AOIs (First 10 minutes).....	114
Table 23. Eye-tracking measurements from the groups of AOIs: Mann-Whitney tests of differences between participants with lower and upper intermediate skills in reading and writing (Overall 40 minutes, the First 10 Minutes and the Last 10 minutes)	115

Table 24. Eye-tracking measurements from the sub-division of AOIs on Rate of Total Fixation Durations: Mann-Whitney tests of differences between participants with lower and upper intermediate skills in reading and writing (Overall 40 minutes, the First 10 Minutes and the Last 10 minutes).....	116
Table 25. Regression results for eye-movement variables for the full 40 minutes (groups of AOIs).....	119
Table 26. Regression results for eye-movement variables in the First 10 minutes (groups of AOIs).....	120
Table 27. Regression results for eye-movement variables in the Last 10 minutes (groups of AOIs).....	120
Table 28. Descriptive statistics of two raters of the TEAP essays	125
Table 29. Correlations between Rater 1 and Rater 2 ratings of the five criteria	125
Table 30. Inter-rater reliabilities between TEAP Task 1, Task 2 and the Average.....	126
Table 31. Descriptive statistics of Aptis writing, reading, TEAP scores	127
Table 32. Pearson Correlation Test on Aptis Reading, Writing and TEAP scores.....	127
Table 33. Aptis writing, reading, and TEAP scores in five criteria: correlations and descriptive statistics (N=40, Task 1 and Task 2).....	129
Table 34. Correlations between eye-tracking metrics and five criteria scores (40 minutes)	132
Table 35. A cognitive processing framework for L2 writers	137
Table 36. Question 1-1	138
Table 37. Question 1-2.....	139
Table 38. Question 1-3.....	139
Table 39. Question 1-4.....	140
Table 40. Question 1-6.....	140
Table 41. Question 1-7.....	141
Table 42. Question 2-1	142
Table 43. Question 2-2.....	142
Table 44. Question 2-3.....	143
Table 45. Question 2-4.....	143
Table 46. Question 3-1	145
Table 47. Question 3-2.....	145
Table 48. Question 3-3.....	145
Table 49. Question 4-1.....	146
Table 50. Question 4-2.....	146

Table 51. Question 7-1.....	147
Table 52. Question 7-2.....	147
Table 53. Question 5-1.....	148
Table 54. Question 5-2.....	148
Table 55. Question 5-3.....	148
Table 56. Question 5-4.....	149
Table 57. Question 5-5.....	149
Table 58. Question 6-1.....	150
Table 59. Question 6-2.....	151
Table 60. Question 6-3.....	151
Table 61. Question 8-1.....	152
Table 62. Question 8-2.....	153
Table 63. Question8-3.....	153
Table 64. Question 8-4.....	153
Table 65. Question 8-5.....	154
Table 66. Question 8-6.....	154
Table 67. Question 8-7.....	155
Table 68. Question 8-8.....	155
Table 69. Question 8-9.....	155
Table 70. Question 9-2.....	156
Table 71. Question 9-5.....	157
Table 72. Cases shown for in-depth analysis (Language Proficiencies).....	159
Table 73. Number of AOI Switches for 10 minutes of recording.....	167
Table 74. Skipped Paragraphs during the First 10 Minutes.....	168
Table 75. Saccades, Regression, Return, and Sweeps found in Paragraph 1 (Task1).....	169
Table 76. Questionnaire answers on the use of source texts.....	169
Table 77. Questionnaire answers on the use of information from graphs.....	170
Table 78. Questionnaire Answers on the Familiarity of Graph Types.....	171
Table 79. Questionnaire on Response Format and Translation for Writing.....	177
Table 80. Written outputs by ST053.....	189
Table 81. Written outputs by SST033.....	190

Table 82. Written outputs by ST030.....	192
Table 83. Written outputs by ST037.....	193
Table 84. Participants' language level in Focus Group Discussions	198
Table 85. Summary of feedback and comments in the Group Discussions on Question #4.....	202
Table 86. Summary of feedback and comments in the Group Discussion on Question #5.	204
Table 87. Summary of feedback and comments in the Group Discussion on Question #6.	206
Table 88. Summary of feedback and comments in the Group Discussion on Question #7.	207
Table 89. Summary of feedback and comments in the Group Discussion on Question #8.	208
Table 90. Summary of feedback and comments in the Group Discussion on Question #9.	210
Table 91. Summary of feedback and comments in the Group Discussion on Question #10.	211
Table 92. Executive Summary	218

Chapter 1. Introduction

1.1 Background

In the field of second language writing assessment, the types of writing tasks that are being tested are undergoing a revolution. Traditionally, writing skills have been tested by independent assessment tasks in which test-takers were asked to write an impromptu essay. Nowadays, integrated writing tasks in standardized tests are becoming a common way of assessing students' abilities in English as a second or a foreign language (EFL). Integrated writing here refers to writing from sources (e.g., texts, information from graphs, audio scripts), which uses multiple skills (i.e. reading, listening, writing) to synthesize information to be produced in a written form. The integrated writing tasks require test-takers to “summarize ideas coherently” by demonstrating “appropriate and meaningful use of orientations to source evidence” (Cumming, 2005. p. 34). In other words, the integrated writing tasks can inform examiners about test-takers' abilities “to construct knowledge effectively from and across relevant sources” (Cumming, 2013, p. 4). This is fundamentally different from “independent writing tasks” that are “decontextualized”, in the sense that test-takers only produce a written output based on their prior knowledge or experience (Zhu et al., 2016. p. 167). Indeed, source-based writing is one of the common integrated writing tasks in academic settings that is difficult even for English speakers (Cumming, Rebuffot & Ledwell, 1989; Shi, 2004; Delaney 2008; Gebril, 2010; Hyland 2005).

Over the past few decades, there have been an increasing number of publications attempting to validate the construction of the task types in the field of language assessments. According to Cumming et al.'s, (2016) analysis of the literature on first and second languages between 1993 and 2013, as many as 69 empirical studies have been published on integrated writing from

sources. Some of these will be reviewed more extensively in the following chapter in order to highlight the significance of the current study.

The rationale behind using integrated writing tasks instead of writing-only tasks is that integrated writing resembles the type of task required in a real-life academic writing assignment (Braine, 1989, Carson, 2000, Horowitz, 1986 cited in Plakans, 2008, Cumming, 2013, Yu et al., 2013). For example, “(s)tudents at schools, colleges, or universities are mainly asked to write in order to display their knowledge of ideas and information from reading, listening to lectures and interacting during courses as well as their abilities to analyze and communicate this material purposefully and coherently relevant to the fields of study” (Cumming, 2013. p. 2). Thus, integrated writing tasks are being used in language assessment for the purpose of assessing academic writing skills. For example, the TOEFL iBT® test includes an integrated task based on reading and listening texts. The IELTS® test also has an integrated task that involves reading into writing and using graph information as a prompt for its summarization task.

The current study attempts to use eye-tracking technology to explore and understand the cognitive processes of integrated writing and the reading-to-writing construct. It attempts to investigate what roles reading and writing play in the integrated writing test, given both multiple texts and graph information as prompts. Also, it explores what kind of test-taking strategies test-takers use in integrated writing. The study further intends to distinguish the characteristics of test-takers’ cognitive operations and performance between the higher-intermediate (CEFR B2 and above)¹ and lower intermediate levels (CEFR B1 and below). The

¹ CEFR, Common European Framework of Reference for Languages; the internationally recognized standard for describing language proficiencies (A1, A2, B1, B2, C1, C2 from the lowest to the highest).

research instruments for this study included Aptis® tests² to measure test-takers' abilities in reading and writing independently, and the Tobii TX300 Eye-tracker to track their eye-movements during the two TEAP tests.³ The TEAP Writing Task B rating scales were used to evaluate their essays, along with stimulated recall surveys (i.e., questionnaires for investigating cognitive-processing and test-taking strategies) to understand test-takers' writing processes and test-taking strategies. Finally, focus group discussions were used to learn about their test-taking experiences. The remainder of this chapter briefly introduces: the context of the study (1.2), its purpose (1.3), significance (1.4), and a description of the organization of the chapters of this doctoral thesis (1.5).

1.2 Context of the Study

Japan, which will host the 2020 Olympic and Paralympic Games, is under tremendous pressure to improve the English skills of the nation as a whole. Japan currently ranks among the least well performing countries in Asia in terms of scores for the Test of English as a Foreign Language (TOEFL®), although it does slightly better on the International English Language Testing System (IELTS®), as shown in Table 1. Japanese university admission policies are often blamed for this trend, since English teachers are put under pressure to prepare their students solely to pass university entrance exams (Underwood, 2010). For example, the National Centre Test (NCT) administers an English test once a year in January. The English element of the NCT exam consists of two parts: an 80-minute reading paper and a 30-minute

² The Aptis test assesses all four skills (listening, reading, writing, and speaking) and grammar and vocabulary components. The test is administrated by the British Council, testing English levels from A1 to C on the CEFR.

³ TEAP, The Test of English for Academic Purpose assesses all four skills to test the academic readiness for Japanese high-school students in the college admission purposes. The test is administrated by the Eiken Foundation of Japan, testing English levels from A2 to B2 on the CEFR

listening paper. Typical classroom instruction for English focuses on repetition and memorization of new words and grammar rules in textbooks so that students can improve their reading comprehension by translating from English to Japanese. The format and content of this test has been argued to have a considerable impact on both teachers' and students' attitudes towards learning English (e.g., Green, 2014; Nakamura, 2014; Underwood, 2010; Watanabe, 1996).

Table 1. IELTS® /TOEFL iBT® Score Ranking in the Asian Region

IELTS Ranking 2013	Country	Average (Overall Scores)	TOEFL iBT Ranking 2014	Country	Average (Overall Scores)
1	Malaysia Philippines	6.8	1	India	91
3	Hong Kong Sri Lanka	6.4	2	Pakistan	90
5	Indonesia	6.2	3	Malaysia Philippines	89
6	India Nepal Taiwan Vietnam	6.0	5	Bangladesh Indonesia South Korea Sri Lanka	84
10	Bangladesh South Korea Pakistan Thailand	5.9	9	Hong Kong	83
14	China Japan	5.7	10	Taiwan	80
16	Uzbekistan	5.5	11	Nepal Vietnam	79
			13	China Uzbekistan	77
			15	Thailand	74
			16	Japan	70

Sources: Test-taker performance (2013) in IELTS® Researcher Report and Test and Score Data Summary for TOEFL iBT® Tests (2014).

As far as the rankings for 2016 are concerned, Japan did better, with the average score going up to 82 from 70 in 2014, thus placing it 10th among the Asian countries listed above on TOEFL iBT®. Similarly, the average band-score has slightly risen to 5.9 from 5.8 in 2013, placing Japan 13th among the same list on the IELTS® in 2015. The recent trend remains very similar to the situation in earlier years, however, in that Japan is still ranked among the lowest among Asian countries in the results of language assessments in 2016.

The improved results could be attributed to the decision made by the Ministry of Education, Culture, Sports and Technology (MEXT) which has proposed a modification of the admissions policy for university so that English tests will include speaking and writing skills. A report issued by MEXT, entitled “The Five Proposals and Specific Measures for Developing Proficiency in English for International Competency”, calls for universities to accept external language certifications as proof of language proficiency for their admissions (MEXT, 2011). This report, however, did not give any guidelines or mandates for the construction of such tests. Given the recent trends towards the use of integrated tasks for assessing language proficiency levels in external language certifications, this new reform has had a significant impact on teachers and students. First, students now need to prepare for all four English skills, including speaking and writing. Even more challenging is the fact that test-takers need to apply a combination of English skills for integrated tasks. Lastly, teachers need to use a different pedagogical approach for the teaching of English in order to improve students’ performance in the tests.

In addition, recently, the number of universities that use external English exams for admission purposes has increased. The Test of English for Academic Purposes (TEAP®) is one of these external exams. The TEAP was developed by the Eiken Foundation of Japan in collaboration with Sophia University to measure the academic preparedness of high-school juniors and

seniors in Japan in terms of the four English skills at the university level (Weir, 2014). It was released to the public in 2014 for the purpose of university admissions in Japan. Between 2016 and 2017 the number of test-takers who have taken the writing component of the test has doubled from 5,466 to 10,839 (Eiken, 2016). The TEAP gives a measure of the test-taker's language proficiency in the academic contexts between the CEFR A2 and B2 levels in a way that is compatible with high school English guidelines in Japan. Test-takers can select which components they are tested on from among three options. The first option consists of reading and writing, the second consists of reading, listening and writing, while the final option includes all four skills. The TEAP writing tasks both involve integrated writing. Task A of the writing test is a summary writing task in which the writer must summarize the main ideas in a text (in approximately 70 words). Task B of the writing test uses multiple source texts and graphs as prompts and poses more cognitive challenges to test-takers than the Task A test.

The next decades will probably witness a considerable increase in the use of integrated writing tasks for the admission tests, such as the TEAP test. The amount of research that has addressed the impacts of the features of the source input (i.e., the impacts of texts and graphs on task performance, the roles of test-takers' language skills, as well as the cognitive processes involved in completing integrated writing tasks) remains relatively small, however. This is a field of study which demands new avenues of research to improve our understanding of the complexity of integrated writing in assessment contexts.

1.3 Purpose of the Study

The last few decades have seen an increasing number of publications on the use of integrated writing tasks for assessing L2 learners, some of which have contributed to identifying the construct validity of integrated writing (e.g., Chan, Wu & Weir, 2013; Zhu et al., 2016;

Cumming et al., 2005; Gebril, 2009; Gebril & Plakans, 2009; Weigle, 2002), discourse types of integrated writing (e.g., Delaney, 2008; Yu, 2009), the effects of source text borrowing (e.g., Shi, 2004; Weigle & Parker, 2012) and the effects of test-taking strategies (Yang, 2012; Yang & Plakans, 2012; Xu & Wu, 2012). Little research, however, has been conducted on the effect of graphic information within integrated writing tasks (e.g., Xi, 2010; Yu, Rea-Dickins & Kiely, 2011; Yang, 2012; Yu & Lin, 2014), despite the fact that this is one of the key elements for understanding the integrated writing processes of tasks with graphs. Graphically-based writing requires learners to demonstrate “dual abilities of comprehension of graph input and transformation of visual information into written discourse” (Yang, 2012. p. 174). Despite its complexity, there are only very few empirical studies on integrated writing that use both source texts and graph information as prompts, as is the case in the TEAP integrated writing test. The current study, therefore, aims to investigate the effects of source texts, graph features and the language proficiency of the test-takers on integrated writing.

Although many previous studies have focused on investigating the cognitive operations among native speakers (L1) during their normal reading, recently some applied linguistic researchers have begun to investigate the cognitive processes of proficient and less proficient test-takers in order to validate the value of eye-tracking methodology in the fields of second language testing. Examples include Bax (2013a) on the IELTS reading test and Brunfaut & McCray (2015) on the Aptis reading test. Very recently McCray & Brunfaut (2018) have revealed the gap-filling process, which is an exercise in which test-takers choose an appropriate word or phrase to fill in the blanks measured by using eye-trackers extensively. The current study also attempts to shed light on L2 learners' cognitive processes during reading-into-writing tasks with the use of an eye-tracking device. The eye-tracker collects data on eye-movements by tracking gaze patterns, from which visual attention can be examined for text comprehension and information processing, thereby helping to determine the cognitive processes of participants by their

information processing over a certain time segment. More details on eye-tracking will be reported in Chapter 3 (Research Methods). To the best of the author's knowledge, this work is one of the first studies to examine the effects of both multiple source texts and graph information using eye-tracking as its main data collection tool.

1.4 Significance of the Study

Despite many signs of progress and attempts to understand the impacts of integrated writing assessment, it has not been fully understood how some of the different features of the source inputs affect test-takers in integrated writing. The problem is partially explained by the many types of source inputs used for integrated writing. As mentioned earlier, the TOEFL iBT® requires test-takers to apply their reading and listening comprehension skills into writing. The IELTS®, meanwhile, requires test-takers to write a short descriptive essay based on visual information or data (i.e., tables, charts and graphs). Also, Task B within the writing module of the TEAP test uses both multiple texts and two types of graphs as prompts. A sample of the TEAP Writing Task B is shown in figure 1 below. To date, there have been few discussions of the key indicators of success in integrated writing tasks when multiple source texts and graph information are used as source inputs.

Task B

Your teacher has asked you to write an essay for class using the information below. Describe the situation concerning schools in Greenhill and summarize the main points about the solutions that have been suggested. In your conclusion, say which of the solutions you think would work the best based on the reasons given. You should write about 200 words.

Students' Problematic Behavior at Schools in Greenhill

Year	Number of incidents
2009	100
2010	150
2011	200
2012	250

Incidents at Greenhill schools in 2012

Category	Percentage
Bullying	31%
Fights	24%
Smoking	22%
Drug use	13%

Education News

There are several ways to deal with the problems in Greenhill schools. Mike Parker, the principal at North Greenhill High School, thinks that giving teachers more knowledge about background issues is the first step. Parker talked about his ideas at a recent meeting with parents. "We have to educate teachers better about problems like drug use and bullying," Parker said. He also pointed out that trouble at home can lead to behavior problems at school. "Our teachers need more information about this," he said. Parker wants to hold special classes for teachers this summer. The classes would help them understand the problems young people face.

However, Parker suggested that other steps are also necessary. For example, according to recent studies, students are more comfortable talking to counselors than to teachers. One reason is that students can talk to counselors in private. This allows students to discuss their feelings, thoughts, and personal experiences more easily. "Perhaps most importantly," said Parker, "counselors are trained to be good listeners." Many schools offer job counseling to students, but not advice about personal problems. Parker wants his school to hire more counselors as soon as possible.

LETTER TO THE EDITOR

Dear Editor,

The situation at Greenhill schools is serious and requires immediate action. I would like to offer some advice, based on my many years of experience as a junior high school teacher. First, the classes in our local schools are simply too large. Teachers cannot handle so many students at once. Research has shown that when schools reduce class sizes, student performance improves. Teachers can spend more time with each student, and students get to know their teachers better. This may not solve all of the current problems, but it is a good place to start.

I also believe that many teachers do not understand the challenges faced by today's students. The only way to solve this problem is to give teachers better training. For example, they could attend classes to learn about how student behavior is affected by stress at home. The classes could also provide information about new kinds of illegal drugs and what their dangers are. If teachers do not have this type of special training, they will not be able to help students. The problems in our schools will only get worse.

Finally, I want to thank all of our teachers for their hard work.

Sincerely,
Sarah Case

Figure 1. The Sample TEAP Writing Task B extracted from the Eiken Website
<http://media.eiken.or.jp/teap/writing/sampletest.pdf>

The outcomes of the study are intended to benefit test developers and teachers by offering a clearer understanding of students' cognitive processes when synthesizing texts and graph information in the process of producing essays. There are some existing studies on the washback effects of the TEAP writing test (e.g., Nakamura, 2014; Weir, 2014) as well as on the validity of the test through criterion-based approaches (e.g., Chan, Wu & Weir, 2013, Koizumi & Nakamura, 2016). This study, however, would be the first to explore cognitive processes of test takers while completing TEAP reading-into-writing Task B using the eye-

tracking method. In fact, the existing literature on the TEAP has used mostly questionnaire surveys and stimulated-recall interviews as research methods. Much of the previous literature investigating cognitive processes has depended on a conventional think-aloud method (See Section 2.5.1 and 2.5.2). The eye-tracking method represents an innovative alternative to a conventional think-aloud method, providing additional insights into the cognitive processes of integrated writing. By using an eye-tracking device this study combines these traditional approaches with eye-movement results, thus allowing test-takers' behaviours across different language proficiencies to be compared quantitatively with qualitative results from questionnaires and focus groups discussions. This mixed-methods approach was taken in order to reduce the risk of misinterpreting the eye-movement results. The findings could be used to help school teachers to use more scientific approaches to guide students in the learning of integrated writing skills.

1.5 Organization of the Thesis

This thesis is organized into three parts comprising nine chapters. The first part includes Chapter 1, which introduces the context in which this study was conducted and the main purpose and motivation to carry out this research, and Chapter 2, which summarizes literature that is relevant to seven areas: (i) language testing and validation, (ii) the theory of the cognitive processes of writing, (iii) the socio-cognitive model, (iv) integrated writing for assessment, (v) graph-sourced writing, (vi) test-taking strategies, and (vii) the TEAP integrated writing tasks.

The second part consists of Chapter 3 which introduces the research method, including the research questions and the four phases of the research: (i) identifying the English proficiency of participants, (ii) tracking participants' eye movements, (iii) questionnaires on test-taking

strategies, and (iv) focus group discussions; as well as (v) theoretical perspectives and the researcher's positioning, (vi) ethics and (vii) the limitations of the chosen approach.

In part three, Chapter 4 reports the quantitative findings using eye-tracking measures (e.g., Fixation Duration, Fixation Count, Visit Duration and Visit count) across a number of Areas of Interests (AOIs) over the different segments of recordings (i.e., all 40 minutes, the first 10 minutes and the last 10 minutes of recordings). The analysis of the test performance of the participants are reported using both Aptis scores, which the test-takers had taken before the eye-tracking experiments, as well as essay marking based on the rating criteria of the TEAP Writing Task B. Chapter 5 provides a quantitative analysis based on the cognitive processing questionnaire. The results are reported according to the three stages of the cognitive process of writing; namely (i) preparing-to-write, (ii) translation and writing, and (iii) monitoring and revising. Chapter 6 showcases some individual cases as examples to illustrate gaze-plots, keystrokes, AOI switches (i.e., the number of switches in defined areas of interests), saccades (i.e., rapid-eye-movement, defined in this study as forward or backward jumps while reading is in progress). Chapter 7 covers the results of the focus group discussions. Finally, Chapter 8 summarises and discusses the main findings in response to the research questions, and Chapter 9 is the conclusion, which includes some implications for test-developers, teachers, Japanese high-school students, and researchers in applied linguistics.

Chapter 2. Literature Reviews

2.1 Introduction

When studying the cognitive process of integrated writing, it is essential to understand how reading plays a role in writing. Over the past two decades, the relationship between reading and writing has been vigorously studied by researchers from different disciplinary backgrounds, including education, psychology and applied linguistics (Grabe, 2003). Some scholars have suggested that reading skills are highly correlated with the quality of writing, as the reading-into-writing process can be thought-provoking for writers when they plan, translate, edit and revise (Flower & Hayes, 1981). Some scholars, however, point to the complexity of reading-into-writing tasks, especially among second language learners, and suggest that careful consideration should be given not only to the role of reading in writing but also to other factors, including background knowledge of topics, motivational and social factors, and opportunities to use the target language (Grabe & Kaplan, 1996; Grabe, 2001, 2003).

In addition, the relationship between writing and speaking is important for understanding the cognitive process of writing. Weigle (2005) reports how writing is distinctively different from speaking in terms of its (1) textual features, (2) socio-cultural norms, (3) patterns of use, and (4) the cognitive processes involved in text production and comprehension through summarizing the work of others (e.g., Grabe & Kaplan, 1996). Among these differences, the most distinctive difference between writing and speaking lies in the cognitive processes involved. Speakers can use strategies such as pauses, turn-keeping signals, as well as the pitch and tone of their voices, in a spontaneous manner. In addition, writers need to be involved in the cognitive processes of planning, using existing knowledge, considering audience interests, addressing purposes, and staying cohesive and logical throughout the text in the absence of spontaneous feedback.

According to Weigle, writing involves such a “complex” mechanism with numerous constraints and considerations that the process of writing and creating written products results in quite a wide difference between skilled and unskilled writers (Weigle, 2005, p. 129). With this in mind, the following section will review some of the literature on relevant topics: (i) language testing and validation for writing (section 2.2), (ii) the theory of writing (section 2.3), (iii) socio-cognitive model (section 2.4), (iv) integrated writing for assessment and (section 2.5), (v) graph-sourced writing (section 2.6), (vi) test-taking strategies (section 2.7), and (vii) TEAP integrated writing tasks (section 2.8).

2.2 Language Testing and Validation

Large-scale language testing (such as TOEFL iBT®, IELTS®, and TEAP®) typically tests English skills beyond what students have learned in classrooms since these tests are proficiency tests rather than achievement tests. To assure that the test measurements are “consistent across different characteristics or facets of a testing situation, test developers usually design specifications for each test (Weigle, 2002. p.49)”. The specifications must be transparent and contain basic information such as a description of the test purposes, target population, the construct, as well as a description of the specific settings in which test-takers need to perform the language use (e.g., Bachman and Palmer, 1996). Validating tests consist of four essential elements, namely Validity and Reliability, Impact and Practicality (VRIP) (Bachman, 1990; Bachman and Palmer, 1996).

The traditional approach used to distinguish types of validity such as “content, predictive, concurrent, construct and even face validity” (Alderson & Banerjee, 2012. p.79). The belief that the validity should be treated as a unified concept emerged in the late 1980s (Messick 1989). Bachman and Palmer later developed the unified view of validity and argued that we consider different aspects of test development such as test usefulness as a part of the test

validity (Bachman and Palmer, 1996). Researchers also became more aware that language testing involves not just test-takers' ability to use the language but also their knowledge of language use. For example, Bachman (1991) argued that language ability and knowledge should also include metacognitive strategies. The discussion of validity further extended to the discussions of issues like computer familiarity (Kunnan, 1996) and policy and social considerations in language assessment (McNamara, 1998).

In this paper, we refer to three types of validities with which language test developers are often concerned: namely construct, content and criterion validities. Construct validity refers to the interpretation of the test scores, that is to say, whether a test successfully measures what it has intended. Content validity refers to "the adequacy and representativeness of the test content vis-à-vis the domain to which test results are intended to generalize" (Weigle, 2002. p. 50). Finally, criterion validity is calculated by finding the correlations between the test scores and other measures such as the teachers' judgments of the student's ability based on empirical evidence (Weigle, 2002). While researchers study different types of test validations, this study will specifically investigate different task features for writing by referring to the socio-cognitive framework developed by Weir (2005). Later in the review (section 2.4), Weir's socio-cognitive model, which applies a more unified approach to the validity of the tests, will be introduced as a theoretical basis for this study (Weir, 2005). This study is intended to explore how different features of the tasks within a test affect students' writing performance by understanding the cognitive processes of integrated writing with source texts and graph. The next section briefly reviews the history of studies on second language writing.

2.3 Theory of Writing

2.3.1 A brief history of second language writing

Traditionally, English language teaching for non-native speakers has focused on students' acquisition of four English language skills, namely reading, listening, writing and speaking. In studies of second language acquisition, writing was often neglected until the beginning of the 1960s, when countries like the United States began accepting a large number of immigrants and international students into higher education institutions (Matsuda, 2003). Writing instruction for second language (L2) learners started to receive wide attention as these higher educational institutions recognized the lack of academic preparedness among these students in their course studies, which led, in 1966, to the creation of a new organization for L2 specialists called TESOL (Teachers of English to Speakers of Other Languages). In the 1970s, experts in intensive English programmes began exploring ways to prepare L2 students in college courses as a "remedial" course, first by attempting the use of "free composition exercises", then the use of a "controlled composition" approach and finally the "*guided composition*" approach (Matsuda, 2003, pp. 19-20). According to Matsuda (2003), some critics, including Robert B. Kaplan (1966), argued that none of these approaches would be appropriate for teaching L2 learners the writing skills required for academic settings. For these critics, the problems for L2 writers usually went beyond issues of sentence structure, and they argued that more attention needed to be paid to learners' use of "rhetoric", including effective organizational structures for essays. These notions eventually led to the development of discourse analysis for L2 writing in the 1980s.

During the early 1980s, there was also a growing interest among researchers in the interaction between reading and writing. In the years that followed, a number of studies were conducted on speakers of English as a first language (L1) to examine the correlation between their reading

and writing abilities (Grabe, 2003). One of these studies, by Shanhan and Tierney (1990), reported that reading and writing abilities correlated between .50 and .70, implying that there was a 25 to 50 percent overlap in these abilities. In addition, researchers began exploring issues associated with learning to write from multiple texts. McGinley (1992) reported that better readers often tended to produce better writing products by synthesizing key information in a better organizational framework. Others, including Elley (1991) and Wagner & Stanovich (1996), reported that extensive reading resulted in better writing products.

Although these empirical studies supported a strong relationship between reading and writing among L1 writers, the same did not seem to apply for L2 writers. According to Grabe (2003), Flashive and Bailey (1993) studied the correlations between reading and writing among L2 learners and discovered that the relationship was only moderate, with an overlap of 12 percent. This study hinted at the complexity of L2 writing and reiterated that good L2 readers cannot automatically be assumed to be good writers. It was this notion which led the current study to investigate what affects reading-into-writing processes in L2 contexts.

2.3.2 Cognitive processes of writing

Since the study of second language writing in academic contexts is a relatively new one, there are not many theories prior to the 1970s that have described cognitive writing processes. Grabe and Kaplan (1996) report that research on second language writing began shifting its focus from “writing products” to “writing processes” in the early 1970s. This change is attributed to a remarkable case study by Janet Emig (1971, 1983) using “verbal protocol analysis” (Grabe & Kaplan, 1996. p. 90). This approach triggered a trend towards using think-aloud protocols to understand writing processes, including a famous cognitive model of the writing process (Flower & Hayes, 1981), which described the writing process in terms of the task environment, the writer’s long-term working memory, and a number of cognitive processes (Weigle, 2002).

There are three important processes to writing in the Hayes-Flowers (1981) writing model, namely planning, translating and reviewing. Most noticeably, their model implied that the writing process is a “recursive process,” rather than the linear process which traditional teaching instruction used to focus on.

In order further to understand the cognitive processes in writing, it is important to consider the sources of knowledge that writers draw upon in writing. There are two types of meta-cognitive processes strategies involved in writing. One is known as “knowledge telling” and the other as “knowledge transforming.” Weigle (2002) summarizes the difference between knowledge telling and knowledge transforming, as proposed by Scardamalia & Bereiter (1987). While in knowledge telling the features of writing are similar to a spontaneous conversation, with little planning and revision made, knowledge transforming involves much more skilful efforts, which often create new insights that the writer intends to communicate to their audience. In other words, writers construct meaning as they compose texts by interpreting texts that they read (Spivey, 1990). The writing strategies applied by skilled writers differ from those in unskilled writers in terms of how the writers successfully apply their knowledge telling and knowledge transformation skills to writing (e.g., Bereiter & Scardamalia, 1987; Grabe & Kaplan, 1996). Integrated writing requires more knowledge transforming skills than knowledge telling skills.

2.4 Socio-Cognitive Model for Writing

Weir’s early work on the socio-cognitive approach put an emphasis on both the use of language in performing a task as well as on the mental processing of the candidate (Weir 2005). In order to conceptualize the framework of the socio-cognitive approach, he introduced a chart such as the one shown in Figure 2. According to Weir, the cognitive validity of a writing task is

measured by the cognitive processing required for real-life performance in writing. In his 2007 publication, Weir reiterates the importance of understanding the current theoretical basis for the validation framework, such as the contextual and cognitive validities of a test. The implications of Weir's work for test developers are that it is important to develop more explicit descriptions by reviewing how context, cognitive processing and scoring interact with each other. The framework ensures both "the nature and quality of the test matches up to the requirements" of the tests (Shaw & Weir, 2007. pp. 4-5).

As figure 2 shows, the framework includes three main components, context validity, cognitive validity and scoring validity, in addition to test-taker characteristics and consequential validity. The context validity concerns whether the task demand is appropriate for the targeted test-takers in the social setting (age groups, previous knowledge, discourse modes, etc.). The scoring validity has to do with evaluation of the task performance. This study concerns mostly the cognitive validity, which consists of Macro-planning, Organisation, Micro-planning, Translation, Monitoring and Revising.

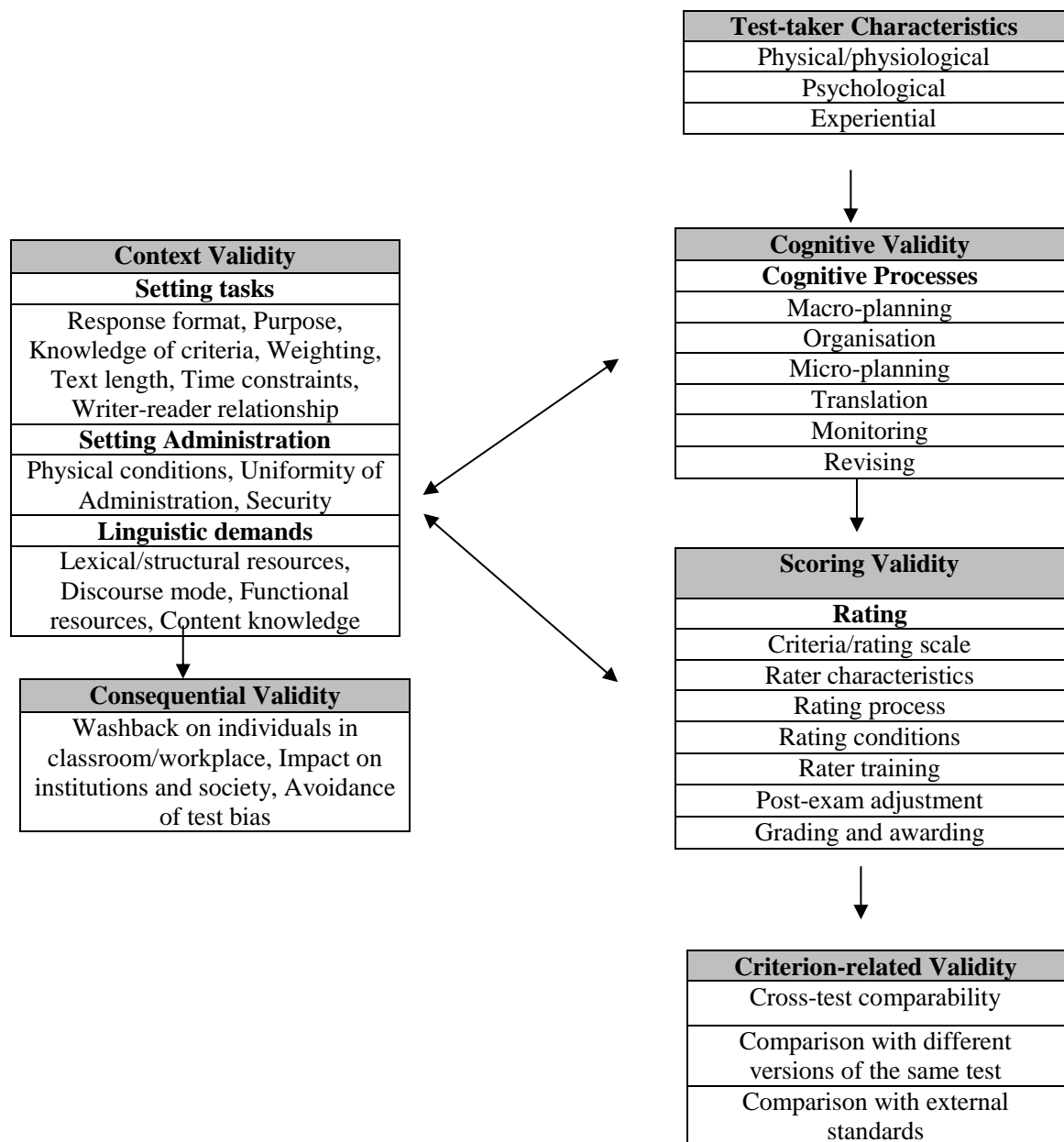


Figure 2. Weir's Framework for a Socio-Cognitive Approach for Assessing the Validity of Writing Performance (Weir, 2005b) cited in Shaw & Weir (2007)

Since the integrated writing Task B of the TEAP involves both reading and writing skills, it is also worth mentioning Weir's socio-cognitive processing model for reading. The cognitive processes of reading go beyond reading under test conditions; they are concerned about reading performance in real life. Readers often choose the types of reading which best serves their purpose given the context. In test situations, the type of reading depends on the types of questions that they are asked to answer. For example, careful readings at the local levels are applied for matching similar words or identifying appropriate lexical items. For comprehending main ideas, global readings may be applied. Other types of reading also include skimming and scanning and search reading (Khalifa & Weir, 2009).

Chan (2013) compared reading-into-writing processes between real-life academic tasks and reading-into-writing test tasks, in which she used the General English Proficiency Test and an in-house diagnostic task developed at the Center for Research in English Learning and Assessment in the University of Bedfordshire. Using Weir's socio-cognitive model (2005), she elucidated the similarities and differences in the cognitive processes of integrated writing between real-life academic writing tasks and the test tasks. She quantified the survey to show that there was a statistically significant difference between how high and low achieving groups behaved in their "task representation," "selecting relevant ideas," "organizing ideas," and "monitoring and revising" (Chan, 2013. p.213). Based on a correlational analysis, it was reported that the participants' scores in real-life academic tasks and reading-into-writing tasks had similarities in terms of their cognitive processes. The findings of the explanatory factor analysis also supported that the reading-into-writing tasks can predict higher or lower achievement in real-life academic tasks. Chan (2013) incorporated some of the previous literature on cognitive models to propose a cognitive validity framework for integrated writing which constituted of "conceptualization", "meaning and discourse construction" "organising" and "monitoring and revising at high and low levels" (Chan, 2013. p. 8). She further examined

the context validity and cognitive validity using the GEPT Writing test 1 (Chan, Wu & Weir, 2014) and criterion validity (Weir, Chan & Nakatsuhara, 2014).

The present study is built upon these studies in the hope of incorporating some additional aspects of the cognitive processes of integrated writing. Because the types of integrated writing task used in the present study involve multiple texts and graphs, the next section will review the latest research findings related to integrated writing for assessing L2 learners. This study particularly focuses on the role of participants' language ability as measured by the eye-tracking method which reflects Weir's understanding of the importance of test-taker characteristics as well as the cognitive processes used during the test.

2.5 Integrated Writing for Assessment

Until very recently, the writing proficiency of L2 learners has been assessed using independent writing tasks (e.g., Cumming et al., 2005; Weigle, 2002; Gebril, 2010; Weigle and Parker, 2012.) Nowadays, though, we find more integrated writing tasks being adopted by test developers because integrated writing tasks more closely resemble the tasks required in real academic settings (Plakans & Gebril, 2012. p. 19). While there are different types of integrated writing tasks, such as writing based on listening and reading, as seen in the Internet-based Test of English as Foreign Language (TOEFLiBT®), the present study mainly focuses on the reading-into-writing tasks most commonly used in the International English Language Testing System (IELTS®) and the Test of English for Academic Purposes (TEAP®). As an overview of reading-into-writing tasks, this section reviews the literature according to the following sub-themes: Independent vs. Integrated Writing for Assessment (2.5.1), and Integrated Writing with Source Texts for Assessment (2.5.2).

2.5.1 Independent writing vs. integrated writing for assessment

Over the years, independent writing tasks have been the norm for most ESL writing assessment. Gebril (2010) gave a good explanation of what impeded the introduction of integrated writing tasks for a long time before the TOEFL iBT® adopted it in 2005. He argued that test developers used to rely heavily on score reliabilities according to Classical Test Theory (CTT). According to Gebril, CTT did not allow for more than one error variable; therefore, it did not measure the impact of different sources of error, such as raters, tasks and scoring rubrics. Under this theory, integrated writing scores did not meet the score reliability that language testing companies would accept. By testing both the independent and integrated task types, Gebril was able to confirm that both independent and integrated writing task types had reliable univariate scores, implying that a combination of scores from both writing tasks would provide a more accurate picture of a student's writing ability than a single writing task (Gebril, 2010). The choice of statistical methods may not be the only reason why some scholars began exploring integrated writing tasks, however. The change to a more integrated writing approach probably also reflects changes in prevailing opinions about language use and pedagogy as I previously discussed in the Introduction as a purpose of the study (p. 18).

Over the past few decades, a number of studies have investigated the similarities and differences between independent and integrated writing (e.g., Cumming et al., 2005; Gebril, 2006; Lewkowicz, 1994; Watanabe, 2001). The areas of interest included not only construct validity and rater reliability, but also language usage in different writing tasks. For example, Plakans (2008) reported some of the advantages of integrated writing tasks by comparing the cognitive processes of test-takers. The aim of her study was to propose which approach gives a more accurate picture of a student's performance in English placement tests (Plakans, 2008). She studied ten non-native English L2 writers, who were tested using both independent and integrated writing tasks and used a think-aloud method to record their cognitive processes, as

well as stimulated recall interviews to uncover the writer's thinking. In the Plakans' study, the test-takers showed a more recursive approach during the process of integrated writing. This was because the integrated task required interpretation of the texts before the formulation of ideas. More importantly, the study suggests that integrated writing tasks showed "larger differences across writers based on their experience and interest in writing" (Plakans, 2008). This finding implies that integrated writing tasks are a better prediction of whether a learner will be a novice or an advanced writer in a real-life academic setting. Of course, independent writing also has its benefits, such as allowing test-takers to demonstrate their abilities to write an essay based on personal experience and knowledge. Integrated writing, however, seems to contain certain elements that are essential for real-life academic writing (McCulloch, 2013).

Some of the characteristics of the cognitive processing in types of essay writing were reported by Shaw & Weir (2007), in which they discussed a cognitive processing framework for L2 writing by referencing different kinds of tests, such as Key English Test (KET), Preliminary English Test (PET), First Certificate in English (FCE), and Certificate in Advanced English (CAE). The table explains what types of writing some of the tests targets. The table is organized by the different CEFR levels of language learners.

Table 2. Summary Table of Cognitive Processing across Main Suite Examinations

CEFR levels (Acronyms of the test product)	A2 (KET)	B1 (PET)	B2 (FCE)	C1 (CAE)
Types of essay writing	knowledge telling	knowledge telling	knowledge telling & knowledge transforming	knowledge transforming
Cognitive processing	limited planning encouraged, lower level monitoring and revision of vocabulary grammar and spelling	limited planning encouraged, monitoring and revision of vocabulary, grammar, and basic organization	analysis and evaluation, planning, monitoring and revision of style and content required	rhetorical and organizational analysis, evaluation, planning, monitoring and revision of style and content required

Source: Shaw & Weir (2007) Figure 3.2. Summary table of cognitive processing across Main Suite examinations, p.61.

This table will be used later as a reference for evaluating how the test-takers performed on the essays using the TEAP rating criteria.

2.5.2 Integrated writing with source texts for assessment

Leki & Carson highlighted the importance for L2 learners of having the writing skills to allow them to draw some ideas from source texts. They stated: “Writing without responsibility for the content of source texts misses the opportunity to engage L2 writing students in the kinds of interactions with text that promote linguistic and intellectual growth” (Leki & Carson, 1975, p. 39). This implies that what makes a reading-into-writing task so unique is the source texts, through which test-takers must demonstrate their abilities to comprehend and summarize relevant information, as well as to respond to ideas in a cohesive manner. This section reviews prior work concerning source texts for reading-into-writing skills.

There are some potential elements in source texts that may influence the quality of the written products, such as the effects of the use of the first language and target language for summarization (e.g., Yu, 2008), familiarity with the discourse types of a source text (e.g., Delaney, 2008; Yu, 2009) and the effects of source text borrowing (e.g., Cumming et al., 2005; Shi, 2004; Weigle & Parker, 2012). Regarding the first of these, Yu (2008) examined the effects of source texts by asking students to summarize the texts using both their first and second language, in order to give an indicator of their reading comprehension ability. In this study, he assessed the appropriateness of the source texts for reading-into-writing tasks. Yu (2009) investigated how source texts affect summary writing using three types of texts (narrative, expository and argumentative) that are similar in length and readability. Based on quantitative data, he validated the hypothesis that the different types of source texts had a significant influence on students' performance in summary essays. To further investigate what caused these effects, he employed post-summarization questionnaires and interviews, from which he found that the effects of the source texts were greater than the differences in students' language abilities (Yu, 2009).

Delaney (2008) also argues that the writing construct used for a summary and a response essay is quite different between L1 and L2 learners. Moreover, his study suggests that students' language proficiency can only modestly predict their performance in a responsive essay (Delaney, 2008). The findings of his study validated the complexity of the reading-into-writing construct, which is quite different from that of reading and writing as separate skills (Bachman, 2002).

Some studies have also investigated the effect of source text borrowing for integrated writing tasks in a language assessment context. For example, Weigle and Parker (2012) reported the extent to which students borrow the language of a source text by comparing two groups of students who have different language proficiency levels. Their aim was to study what was

impeding examiners from accurately measuring the language proficiency of students, such as the amount of text the students had borrowed. In their findings, there were only a few differences in the amount of, as well as the types of, borrowing between the two groups (Weigle and Parker, 2012). It is interesting to note that Weigle and Parker's findings contradict those of Cumming et al. (2005) who stated that textual borrowing differed according to the types of topics students were asked to write, and that students who had lower proficiency tended to borrow longer strings of words from the source texts.

As this section shows, integrated writing using source texts is apparently one of the most important skills for academic writing. It is worth noting, however, that such a fundamental skill of academic writing is difficult even for L1 learners (e.g., Cumming, Rebuffot & Ledwell, 1989; Delaney 2008; Hyland 2005; Shi, 2004). Shi (2004) reported various types of textual borrowing in essays among both L1 and L2 students at university level and suggested that plagiarism by L2 learners may be attributed not only to limitations in language proficiency in terms of being able to paraphrase the key concepts of the source texts, but also to cultural conventions (Shi, 2004; 2006). Hyland (2005) also examined academic plagiarism among L1 and L2 learners and found that "appropriate academic referencing" requires some degree of training. Also, in his view, there is "a link between good reading skills, knowledge of the topic and appropriate bounding and documentation of references in students' writing" (Hyland, 2005, p.70). In this study, the borrowing of text was studied in a real-life academic context, but textual borrowing under test conditions for integrated writing may be quite different from real-life academic conditions, in terms of the time allowed to refer to the text, as well as the lengths of the text(s) and the number of sources the writers use.

While most past studies focus on the relationship between the source texts and the quality of written products and language proficiency, the process of composing reading-into-writing tasks based upon source texts has been studied only recently. For example, Plakans (2008) compared

writing-only tasks with integrated tasks and proposed a model for reading-to-writing test tasks. Plakans & Gebril (2012) employed a mixed-methods approach using nine undergraduate students in order to understand L2 writers' processes while doing reading-into-writing tasks. Based on their results, they report that source texts were used mainly in three ways: to "shape writers' opinions, provide ideas on the topic, supporting writers' opinions" and finally "serving as a language source" (Plakans & Gebril, 2012, p.32).

2.6 Graph-based Writing

The last few decades have seen a rise in the number of publications on integrated writing tasks, concentrating variously on written products, task features or discourse types, and source texts. There have been only a few studies of the use of information from graphs in integrated writing, however; with most of these being focused on the process of the interpretation of the graphics in such tasks (e.g., Carswell, Emery and Lonon, 1993). Empirical studies on information attained from graphs in the field of test assessment for second language learners are also relatively new; although some of the recent L2 studies in the context of language testing include the effects of graph familiarity, as well as the use of information from graphs (Yu, Rea-Dickins & Kiely, 2011; Yang, 2012; Yu & Lin, 2014).

Yu, Rea-Dickins & Kiely (2011) studied the cognitive process of graphically-based writing, using the IELTS® Academic Writing Task 1 in order to investigate the factors that affect this process, including the types of graph, graphic knowledge and skills, writing proficiency and the training that test-takers receive. The IELTS Academic Writing Task 1 is a summary writing task in which candidates are "asked to describe some information from (a graph, chart, table, or diagram) and to present the description in their words" (Yu, Rea-Dickins, and Kiely, 2011, p. 4). Using the knowledge-based approach (Carpenter and Shah, 1998; Freedman and Shah,

2002), the study attempted to explain how the use of different graphs, their graphic knowledge and English writing abilities affected the test-takers. The study used think-aloud methods followed by surveys and interviews with eighteen IELTS candidates at one of the leading Chinese universities (Yu, Rea-Dickins, and Kiely, 2011, p. 8). Among the significant findings of this study were the following points. (1) The types of graphic information affect the written products, as witnessed in the use of different vocabulary. (2) Although it did not affect the written products, the survey results seemed to support the contention that the level of familiarity with graphic conventions does influence the way writers process the graphic information. (3) Based on findings from a think-aloud method, the writing abilities of the participants and their prior academic experiences did matter. (4) Finally, there was a strong correlation between performance before and after specific training on graphic-based writing, with the rating of the written output suggesting that the “coachability” of graphic-based writing has a significant effect (Yu, Rea-Dickins, and Kiely, 2012, pp 34-35). More recently, Yu & Lin (2014) investigated the cognitive processes of graphical-based writing using the GEPT advanced and the IELTS Academic Writing Task 1, describing a timescale in which to illustrate in depth the holistic process of summary writing using information from graphs. This study also attempts to use the three key stages of the cognitive process approach for graph-based writing proposed by Yu et al. (2011), namely “comprehending non-graphically presented task instructions”, “comprehending graphic information” and “re-producing graph comprehension in written discourse in English as a foreign language” as its theoretical basis.

Yang conducted another study, very similar to that of Yu & Lin above (2012; 2014). This study compared the cognitive processes for reading-based writing and graphic-based writing. Yang borrowed concepts relating to the process of graphic-based writing from previous studies, including ideas of “global” and “local” processing from Carswell, Emery and Lonon (1993), as well as “planning,” “formulating” and “editing” from Mickan, Slater and Gibson (2000).

Yang's hybrid model also has six major processes for the use of graphics in integrated writing tasks (i.e., macro-planning, organizing, micro-planning, translating, monitoring and revising), referring to Bridges (2010). In an attempt to investigate the similarities and differences of the processes writers employed in reading-based writing and graphic-based writing, Yang used a concurrent think-aloud method and stimulated recall interviews. She analysed these by coding them according to the processes used. One of her findings was that graphic-based writing posed more cognitive challenges to writers than reading-based writing. The findings seem to support earlier studies such as by Yu, He and Isaacs (2017) which was a follow-on study of Yu, Rea-Dickins and Kiely (2012).

This literature review on integrated writing for language assessment has raised awareness about some of the challenges surrounding integrated assessment. Previous works have been mostly limited to the ability of written outputs and language proficiencies to validate the construct of the integrated writing tasks, and/or the effects of source inputs such as texts or graphs. It has been shown that the past studies on integrated writing have revealed the impacts of different features of source texts. Such reviews included findings that different types of source texts had led to significant differences in the written outputs (e.g. Delaney 2008; Yu, 2009). Others pointed out about the effects of source text borrowing, including Weigle and Parker's (2012) contention that lower proficiency students tended to borrow longer strings of words from the source texts. Finally, the section reviewed studies on the effects of information from graphs (Yang, 2012; Yu, Rea-Dickins, and Kiely, 2012). Graphical knowledge was found to be just as important as having grammatical knowledge in graphically-based writing. There is still a need to explore other types of integrated writing, however, particularly ones that use multiple source texts and information from graphs as source inputs (e.g., the TEAP test). The use of both source texts and information from graphs mirrors academic literacy activities in the real world; indeed, the TEAP test was built upon Weir's "Language Testing and Validation; an evidence-based

approach” (2005) with “cognitively demanding topics and themes with a more sophisticated focus appropriate to academic settings, usually with wider social relevance” (Weir, 2014. p. 11). Thus, this thesis seeks to address the impacts of both source texts and information from graphs in order to shed more light on the cognitive processes of writing in a broader sense. Until now, however, only a few studies have adopted the eye-tracking approach for investigating multiple source texts and graphs.

2.7 Test-taking Strategies used for Integrated Writing

2.7.1 Overview of the use of test-taking strategies in L2 assessments

Test-taking strategies are part of the cognitive process of writing that affects students’ performance in integrated writing. Cohen (2006) argued that, until very recently, second language research for assessment and test developers did not fully incorporate “the kinds of strategies that respondents were drawing on as they completed language tests” (Cohen, 2006, p. 308.). According to Cohen, early works on test-taking strategies were mostly focused on conceptualizing different test-taking strategies by classifying them as: (1) language learner strategies; (2) test management strategies; and (3) test-wiseness strategies (Cohen, 2006. p.308.) According to Cohen, language learner strategies are students’ abilities to organize the features of the language that needs to be learned. Test-wiseness strategies reflect test-takers’ prior knowledge about the test format. Finally, test-taking strategies include test-takers’ use of language. Since the 1980s, more works have been published on test-taking strategies for test validation, and the use of think-aloud protocols, surveys and interviews to review strategy use have become common (Cohen, 2006). The following section will review some of the latest publications concerning test-taking strategies in integrated writing tasks for validation purposes using the think-aloud methods (e.g., Bridges, 2010; Xu & Wu, 2012).

2.7.2 Test-taking strategies for language assessment

Bridges (2010) examined the validity of IELTS academic writing task 1. She used a grounded theory approach to code data collected from the think-aloud protocols of four participants which revealed different test-taking strategies. She then triangulated the data through cognitive processes questionnaires, referring to the model developed by Shaw & Weir (2007), in order to understand the language use of participants in the different stages of macro-planning, organizing, micro-planning, translating, monitoring and revising (Field, 2004). The findings suggest that the cognitive processes required for the IELTS Task 1 are appropriate given the variety of cognitive processes employed by test-takers. In this study, it was evident that less time was spent on the organization of the task in IELTS Task 1 than the argumentative task in IELTS Task 2. Based on these findings, Bridges concludes that Task 1 of the IELTS measures what was intended.

Xu & Wu (2012) employed a think-aloud method and in-depth interviews to investigate the validity of two types of writing test for the Beijing Matriculation English Test (BMET), which use pictures as prompts, namely “Situational Writing” and “Interpretational Writing”. In situational writing tasks, the test-takers are asked to deliver a story based on the four pictures. During the course of their narrative writing, the test-takers are measured according to their abilities to convey the message coherently by use of discourse markers to show the links between the pictures. In interpretational writing, the test-takers are evaluated on their abilities to produce a cohesive, argumentative essay which consists of a picture description and theme representations. Xu & Wu (2012) developed a coding system in order to analyse the different test-taking strategies used by the students taking this test. The kind of test-wisness strategies they adopted were a) Intent consideration strategy, 2) Time-using strategy and 3) Copying strategy. Through a combination of a think-aloud method and interview and discourse analysis, the study attempted to evaluate the validity of the test by discussing whether the findings (e.g.,

the test-taking strategies employed by students) mirrored the outcomes that test developers had intended to measure. In doing so, they found that compensation strategies were used more for the “Interpretational Writing” than any other strategies (Xu & Wu, 2012, p. 179). The interviews confirmed that students took note of their teachers’ advice to avoid using words and phrases that they were unsure how to use; thus, students became overly concerned about the accuracy of language use and choosing “the best” theme for the essay (Xu & Wu, 2012, p. 185).

Yang & Plakans (2012) examined second language writers’ strategies in reading-listening-writing test tasks by developing the strategy inventory for integrated writing. Its aim was to explore mental and behavioural activities before, during and after writing. The study examined six factors including monitoring, test-wiseness, organization, connecting, evaluating and selecting. Then, the integrated writing scoring rubrics were used to compare how these strategies were applied during the process of writing essays. The study provided empirical evidence that test-takers rely on selected strategies. It was found that the use of a discourse synthesis strategy (i.e. selecting, connecting and organizing) had a positive impact on the test performance while the test-wiseness strategy had a negative impact on the test performance, implying that abilities to execute “regulation skills for management reading, listening, and writing interactions” could have impacts on integrated writing (Yang & Plakans, 2012. p. 80).

Yang (2012) also conducted another study to explore the relationships between test-taking strategies and test performance in a graphical writing task. The graphical writing strategy inventory was developed through six-point Likert questionnaires to help understand whether test-takers had completed the writing tasks making use of skills such as global processing, evaluating, linking, local processing, selecting and planning. By using the graphical writing scoring rubrics, the study investigated the relationship between the use of a graphical writing strategy and graphical writing test performance. The study concluded that while the use of graphical writing strategies by L2 writers had some positive relationships to test performance,

familiarity with graphs and the use of a test-wiseness strategy should also be carefully reviewed. The study also suggested the importance of building graphical literacy skills such as graphical familiarity and lexical knowledge on the comprehension of graphs.

The findings of the above studies exemplified that test-taking strategies play an important role in the understanding of test-takers' integrated writing processes. To inquire into the kind of strategies used for integrated writing tasks, researchers have used inventories in the forms of a Likert-scale questionnaire. As shown in Yang and Plakans (2009) for example, it would be helpful to form questions on the different stages of writing (i.e., before, during and after). Also, drawing from this literature, the questionnaire in the current study should include graph familiarity and the lexical understanding of graphs, as suggested in Yang (2012) (See Chapter 5 on the Cognitive Processing Questionnaires, the name coined after Chan, Wu & Weir's study in 2013).

2.8 The TEAP integrated writing tasks

The final section of the literature review includes the specific type of integrated writing tasks used in this study. The present study explored cognitive processes of integrated writing with multiple source texts and graph information using the Test of English for Academic Purposes, or TEAP, as an instrument. TEAP was developed by the Eiken Foundation of Japan in collaboration with Sophia University, which is one of the leading private universities in Japan. In addition, Professor Cyril Weir and his collaborators at the Centre for Research in English Language Learning and Assessment (CRELA) at the University of Bedfordshire in the United Kingdom have contributed by validating the design of the TEAP Writing Test tasks and have published an associated research report (Weir, 2014). The main purpose of his study was to evaluate the context and cognitive validity of the test, as it is a high-stakes test taken by Japanese high-school students who wish to enter Japanese universities.

The study examined both the inter-rater validity and the scoring validity of the rating of the TEAP writing test. As far as the inter-rater agreement is concerned, the study found that all raters had shown their ability to assign scores “in a suitable uniform manner...with adequate levels” (Weir, 2014, p.19). In terms of the cognitive validity of the TEAP writing test, Weir examined the features of the writing scripts by analysing a total of 112 scripts using the Coh-Metrix program. From this, the study confirmed that raters were able to distinguish successfully between the A2 and B1 level and that the scripts at B1 level had used a diverse range of lexis, as well as syntactic complexity. The distinction between B1 and A2 levels is critical since B1 is the proficiency level recommended by MEXT for high-school graduates (Weir, 2014).

In order to assess the socio-cognitive validity of the TEAP writing test, Weir proposed that the TEAP test adopt the socio-cognitive process model (Weir, 2005). The real-life cognitive processes reflected in the TEAP writing tasks are “(1) Task representation, (2) Macro-planning, (3) Reading source texts, (4) Selecting, (5) Connecting, (6) Organizing, (7) Micro-planning, (8) Translating, and (9) Monitoring and Revising” (Weir, 2014, p. 8).

The TEAP writing test has two sections, both of which are integrated writing tasks. Task A requires test-takers to write a summary of a single text (approximately 70 words), and Task B evaluates their ability to synthesize information from multiple sources, including texts and graphs (approximately 200 words). Test-takers are allotted 70 minutes to complete both Task A and B for the TEAP writing test. The Task B writing task poses more cognitive challenges than the Task A writing task, since in Task B test-takers are required to synthesize information from multiple source texts and two types of information from graphs that are given as prompts.

In both tasks, five criteria are used for evaluation: (1) Main ideas, (2) Coherence, (3) Cohesion, (4) Lexical range and accuracy and (5) Grammatical range and accuracy. Weir’s study reported a score distribution for Task A and Task B in which fewer students were graded at B1 for Task

B than for Task A, confirming that more students had difficulty with Task B than Task A, as the test designs intended (Weir, 2014). The scoring rubrics for the TEAP writing Task B are provided in Appendix C.

Another past study is worth mentioning in relation to the TEAP writing task. By referring to the same essay scripts that were from Weir's study in 2014, Nishikawa (2015) published *The TEAP Writing Teachers' Manual*, which described the most commonly-seen linguistic features of students' written scripts. The guidebook was published in order to provide high-school teachers with useful information on how to give meaningful feedback in class. Before its publication, Nishikawa (2015) first conducted a survey which studied seventeen high-school English teachers, in order to understand what challenges the teachers face in teaching academic writing (Nishikawa, January 2015, unpublished survey for a needs analysis.) There were three significant findings from this survey: (1) only one of the teachers had taken an academic writing assessment test themselves. Thus, it can be said that the teachers in the survey lacked experience of academic writing, (2) the teachers surveyed mostly taught their students to compose a single sentence using some target words and phrases. Here, it is important to note that this task mostly involved translating a sentence from Japanese into the English language, and (3) the teachers' feedback mostly focused on correcting grammatical errors and inappropriate use of language. Table 3 below summarizes some advice given in the guidebook titled "*TEAP Writing Teachers' Manual*" (Nishikawa, 2015), highlighting some of the language features of students' writing in integrated writing tasks for the TEAP writing test.

Table 3. Points of Advice and Examples of Common Mistakes in the TEAP Writing Test

TEAP WRITING TASK A	Points of Advice	Examples of Common Mistakes
<p>[Prompt] p. 6 Summative Essay</p> <p><i>Your teacher has asked you to read the following passage and summarize what the writer says about pets at work. Your summary should be one paragraph of about 70 words.</i></p>	(1) Understand task instruction correctly	p. 7 [Example 1] -stating one’s own opinion using the first-person pronoun and phrases such as “I suggest” and “I understand.”
	(2) Identify key information from the source text	p. 8 [Example 2] -copying entirely from the third paragraph of the source text.
	(3) Paraphrase the source texts using your own words	p. 8 [Example 3] -borrowing extensively from the source texts without citing the sources.
	(4) Express the ideas in cohesive and coherent manner	p. 9 [Example 4] -using the same word and phrase repetitively -heavily depending upon similar sentence structures to those in the source texts
	(5) Review other commonly found mistakes with grammar	-using constructions in academic writing (can’t don’t, won’t) that need to be written in full

TEAP WRITING TASK B	Points of Advice	Examples of Common Mistakes
<p>[Prompt] p. 10 Integrated Essay (Problem-Solution)</p> <p><i>Your teacher has asked you to write an essay for the class using the information below. Describe the situation concerning schools in Greenhill and summarize the main points about the solutions that have been suggested. In your conclusion, say which of the solutions you think would work the best based on the reasons given. You should write about 200 words.</i></p>	(1) Understand task instruction correctly	p. 11 [Example 1] -stating one’s own opinion using the first-person pronoun and phrases such as “I think” and “I want to” rather than describing the main points of the source texts
	(2) Identify key information and paraphrase the source texts using your own words	pp. 11-12 [Example 1] -not describing information from non-verbal inputs (graphs) -borrowing extensively from the source texts without citing where the ideas come from
	(3) Think about what constitutes a paragraph and make an outline of the paragraphs to write an argumentative essay	pp. 11-12 [Example 1 & 2] -composing a paragraph without a clear topic sentence, supporting sentences and a concluding sentence in the paragraph
	(4) Make the paragraph more coherent and cohesive	p. 11-12 [Example 1 & 2] -using the same word and phrase repetitively -heavily depending upon similar sentence structures to those in the source texts

Note. Adapted from the guidebook, “*TEAP Writing Teachers’ Manual*” (Nishikawa, M. 2015)

The manual provides an overall picture of the language features commonly shared by test-takers in the TEAP test. The process of creating this guidebook was helpful for understanding overall trends within the written products of test-takers when a complex task such as integrated writing is required.

2.9 Conclusion

The chapter first reviewed how writing research has shifted from written outputs towards writing processes as a way of understanding the construct between independent and integrated writing tasks. As noted in this chapter, integrated tasks are becoming more common in L2 language assessments since they allow test-takers to demonstrate the ability to synthesize sources in the process of writing a cohesive composition. Many issues remain unanswered, however, including relating to the basic question of whether integrated writing accurately measures what the test developers have intended to measure given a combination of skills. In addition, its complex nature poses many challenges, even in real academic settings. For example, some of the literature has addressed concerns about the process of borrowing from the source texts.

Many of the current studies, however, were conducted with native speakers, rather than in the L2 context. Furthermore, only a few studies have looked at both graph-based writing and source texts as prompts. In these the level of familiarity with the graph types determined the level of comprehension of the graphs; raising some concerns among researchers as to whether this might affect one's ability to interpret information from graphs for writing. The previous literature, therefore, has mostly focused on either graph-based writing or source-based writing, not both.

Despite the increased number of publications on integrated writing processes, much of the previous literature has used traditional think-aloud protocols and interview data. This thesis

takes a new look at what constitutes integrated writing processes by exploring the cognitive processes of integrated writing using a new approach with a combination of eye-tracking methods, cognitive processing questionnaires and focus group discussions. The next chapter explains the research questions and the research methods used for the current study.

Chapter 3 Methodology of the Study

3.1 Research Questions

As illustrated in the literature review, there have been few studies in an L2 context on the impact on integrated writing tasks of using multiple texts and information from graphs. This study addresses an overarching research aim of exploring the key variables that affect the cognitive process of reading-into-writing tasks. Five research questions were explored in order to understand the cognitive process of integrated writing tasks among the L2 writers.

3.1.1 Research questions

The main research question is:

What are Japanese EFL test-takers' cognitive processes while completing the TEAP reading into-writing Task B?

The question is supported by additional subsidiary research questions:

1. To what extent do test-takers incorporate information from the multiple texts and the graphs?
2. To what extent do features of the graphs (e.g. line graph vs. bar graph) affect the cognitive processes of integrated writing tasks?
3. What role does language proficiency play in integrated writing tasks?
4. What kinds of test-taking strategies are used for integrated writing tasks?

3.2 Research Methods

Unlike with experimental research, which would tightly control the elements that would affect the sensitivity of measurement, this applied research explored what happens during integrated writing tasks with a targeted audience of high-school students (n=38).

This study collected both quantitative and qualitative data in four stages. In the first phase, participants' reading and writing test scores were independently measured by means of a standardized test, called the Aptis test, which was designed to assess English levels in all four skills. This process was intended to enable exploration of the role of reading and writing proficiency in integrated writing tasks.

In the second phase, eye-movements were recorded in order to explore the cognitive processes of integrated writing with different graph features. An eye-tracker, model TX300 with a frame rate of 300 Hz, and a screen resolution of 1,920 by 1,080 was used for the study. The lease was made for the duration of the data collection (one month, February 2016). Two sample tasks from the TEAP Writing Task B (Task 1 and Task 2, in Appendix A and B respectively) were developed. These were similar in the length of their text length but contained different graphical features and these were tested with high-school students at four different sites.

In the third phase, cognitive-processing questionnaires were used to explore the test-taking strategies adopted by the students using Weir's socio-cognitive framework. The cognitive-processing questionnaire explored participants' cognitive processes in terms of task representation, macro-planning, organizing, micro-planning, translating and reviewing and monitoring.

In the final stage, the participants shared their experiences of the test in focus group discussions. Together, these phases were intended to help gain a deeper understanding of the cognitive

processes of integrated writing when multiple source texts and information from graphs were given as prompts.

3.2.1 Research design and justification of the method approach.

This study employed a mixed method explanatory research design, which is considered to be “most useful when the researcher wants to assess the trends and relationships with quantitative data but also be able to explain the mechanism or reasons behind the resultant trends” (Creswell, 2013, p. 82). According to Creswell, an explanatory research design typically has three phases. The first and second phases involve collecting and analysing quantitative data to guide the third phase of qualitative analysis. In the third phase, researchers attempt to explain the findings of the quantitative data in more depth using qualitative methods. Accordingly, the present study consists of four stages to follow the explanatory research design framework: (1) An initial phase to characterize the participants, (2) Understanding the cognitive process of integrated writing tasks using eye-tracking, (3) Cognitive-processing questionnaires, and (4) Focus group discussions.

The main reason for using the sequential mixed methods design was that while the quantitative data might only give a general understanding of the problem, the qualitative data can help refine and explain the statistical results by exploring participants' views (Ivankova, Creswell & Stick, 2006). The first phase of quantitative data helps determine the aspects that are attributing the participants' language abilities with the Aptsis scores. Also, by analysing the eye-tracking data in the second phase, it helps identify which eye-tracking variables might be a better indicator of successful and less successful writers (e.g., Bax, 2013a, 2013b; Brunfaut & McCray, 2015, 2018; Yu, He & Issacs, 2017). The scores can be aggregated and be analysed quantitatively from the survey to help explain the results by examining qualitative data with focus group discussions (e.g., Kitzinger, 1995; Rabiee, 2004). The processes involved in each stage are

shown in Figure 3 below and Table 4 summarizes the methodological justification linked to existing literature.

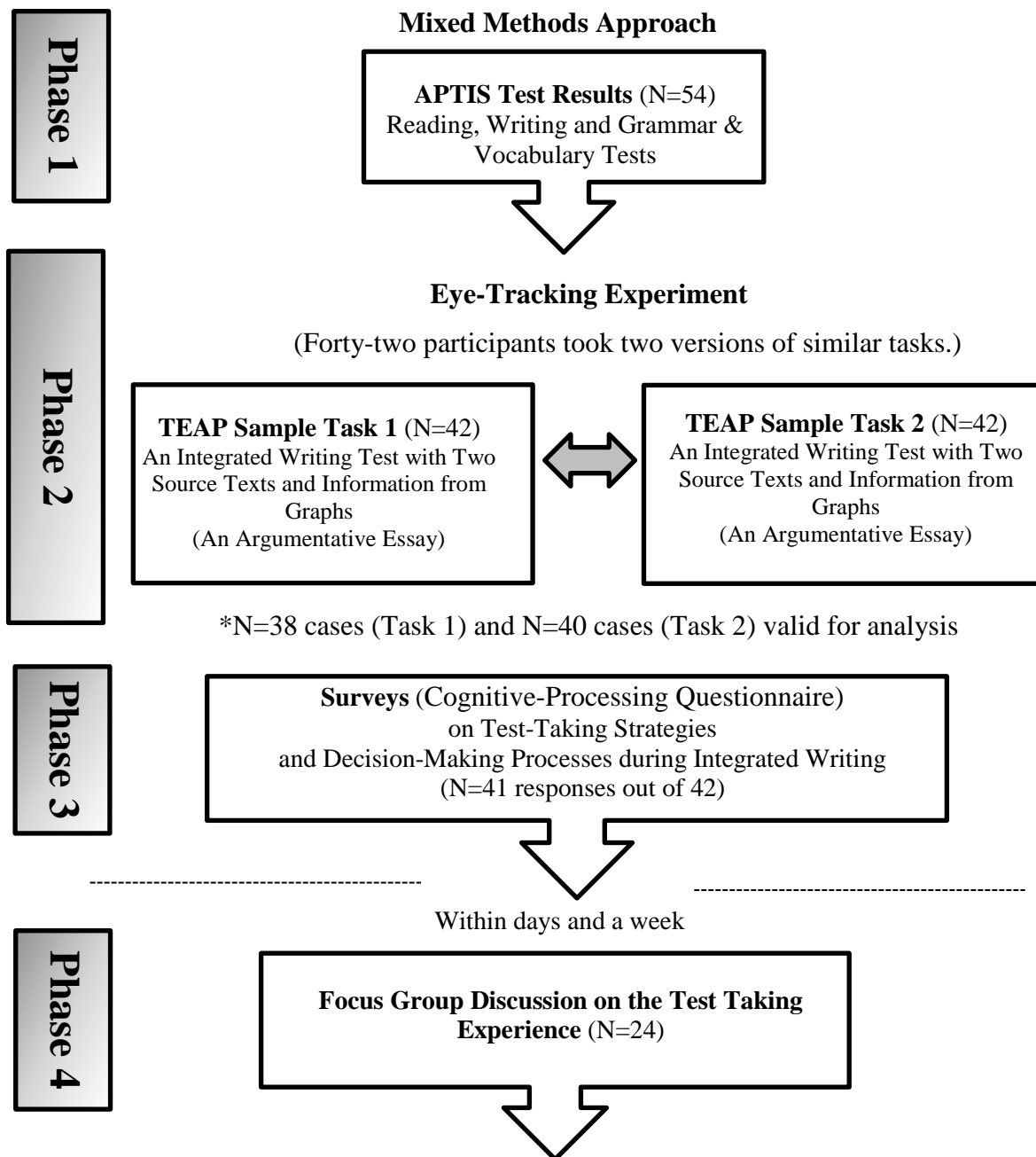


Figure 3. Research Methods

Table 4. Justification of the Research Methods of the Explanatory Design

Stages	Research Instruments	Justification	Relevant Literature
Phase 1 & 2 Quantitative Method	-Describe participants' language proficiency levels using Aptis test	-Identify the characteristics of the? quantitative sample	
Phase 2 Quantitative Method	-Data collection from Eye-tracker during the integrated writing tasks	-Analyse the quantitative data to answer the quantitative research questions and facilitate the selection of participants for the second phase	Use of eye-tracking method for assessing L2 learners' cognitive processing (Bax, 2013a, 2013b; Brunfaut & McCray, 2015, 2018; Yu, He & Issacs, 2017)
Phase 3 Quantitative Method	-Surveys (Cognitive-processing questionnaires) on Test-taking strategies	-Help explain the quantitative results by finding group differences and significant results from Phase 1 and 2	Advantage: -Helps to interpret the results from the eye-tracking experiment Disadvantage: -skipped questions, accessibility and dishonesty Relevant literature: -Developing questionnaires for finding the effects of graph-source writing (Yu, Rea-Dickins & Kiely, 2012); test-taking strategies (Phakiti, 2003; Bridges, 2010; Chen, 2012; Xu & Wu, 2012; Heo, Stoffa & Kush, 2012; Yang 2012) Developing questionnaires using the cognitive processing model (Weir, 2005; Chan, 2013)
Phase 4 Qualitative Method	-Focus Group Discussions	-Purposefully select a qualitative sample that can help explain the quantitative results -Collect open-ended data with protocols informed by the quantitative results	Advantages: -Allows to obtain data in natural occurring environment (e.g., Kitzinger, 1995) -Encourage participants to give honest feedback that are difficult in a face-to-face interview (e.g., Rabiee, 2004) Disadvantages: -Group dynamics could overshadow the voice of silent participants -Inappropriate when a sensitive personal issue was to be discussed (e.g., Folch-Lyon & Trost, 1981)

3.2.2. Rationale behind the use of eye-tracking technology as an instrument

As discussed in the literature review in Chapter 2, traditionally, research to investigate the underlying constructs in language assessment used to rely on think-aloud protocols or stimulated recall interviews. Indeed, think-aloud protocols were developed based upon Flower and Hayes' (1981) theory that the writing process is an interactive, goal-oriented activity where expert and novice writers may produce different levels of writing. While this theory has many advantages, some research has argued that the practice of translating the internal processes in this method takes place retrospectively and that its findings may therefore not be reliable. One reason is that the findings from this approach may be influenced by the capacity of the short-term working memory, which allows students to talk and process information simultaneously (Plakans, 2009.) Also, some studies report that participants unintentionally self-correct their behaviour while thinking aloud (e.g., Bridges, 2010; Green, 1998; Plakans, 2009). By recording actual rather than remembered or self-articulated behaviour eye-tracking can circumvent these problems and this study, therefore, attempts to use an eye-tracking device to help understand the cognitive processes involved in integrated writing.

Indeed, eye-tracking technology is increasingly becoming available to linguistic professionals. For example, Suvorov used an eye-tracking device to compare context and content videos when studying the construct of second language acquisition in listening (Suvorov, 2015). Likewise, Bax (2013a, 2013b) investigated the cognitive elements of the IELTS reading test using an eye-tracking device. Although there have been some studies to gain insights into readers' cognitive processes using eye-tracking technology, most of the research in the past focused on native English speakers during "default" reading (Bax, 2013. p.444.) Bax's eye-tracking studies were among the first attempts to explore the cognitive behaviours of L2 learners during reading and writing in the field of language assessment.

According to Bax and Weir, eye-movements have been regarded as good indicators of cognitive processes up to lexical levels for default reading (Bax & Weir, 2012). The nature of language testing, for example, that requires the use of different parts of texts to locate the key information, makes it possible to use eye-tracking technology to investigate “readers’ higher-order, post-lexical processing behaviour” (Bax, 2013, p. 446). Bax (2013) used the *Tobii T60* model to test 38 participants’ eye movements in order to discover the distinctive traits for successful and unsuccessful readers by measuring cognitive and metacognitive processes during reading. The study validated the reliability of the test items, measured by the eye-movements of test-takers, with different English proficiency levels, when answering different test items. The study also showed that successful readers focused on expeditious reading, which is defined as “quick, selective and efficient reading to access desired information in a text” (Khalifa & Weir, 2009. P. 46). Selective reading of a text often involves skimming and scanning as test-takers attempt to match descriptions of main ideas with paragraphs. Bax was able to locate the signs of expeditious reading by tracking what parts of the text readers viewed, measured by the total fixation duration, fixation counts, total visit duration, and visit count to the part of the text needed for a correct answer. He also used the evidence from recording eye-movements during his stimulated-recall interviews with the participants. In his findings, he was able to identify some notable differences between successful and unsuccessful readers at lexical and grammatical levels (e.g., matching words and synonyms in the text), as well as in expeditious reading and meta-cognitive awareness (e.g., locating the part of the text needed for a correct answer quickly).

Another recent study that employed a mixed method of eye-tracking and stimulated recall was conducted by Brunfaut and McCray using the Aptis reading test (2015). The study investigated the construct validity of Aptis components by analysing the task processing of 25 test-takers in a reading test. They claimed that the combination of eye-tracking data and some stimulated-

recall interviews provided crucial information about whether each component of the Aptis reading test measures what it was intended to measure (Brunfaut & McCray, 2015.) By examining the construct validity of each element of the Aptis test, Brunfaut and McCray investigated whether the differences in cognitive processes were affected by types of task.

Brunfaut & McCray (2015) validated each component of the Aptis test by making a list of hypotheses based on the eye-tracking measure in respect to global processing, text processing and task processing in relation to reading tasks at different CEFR levels. Variables used from eye-recording included total number of fixations, total fixation time, number of saccades (forward jumps), median length of saccades, number of regressions (backward movements), median length of regression and number of Area of Interest (AOI) switches. According to some empirical studies (Blanchard, Pollatsek & Rayner, 1989; Jacobson & Dodwell, 1979 cited in Brunfaut & McCray, 2015), when fixation lengths increase, the saccade lengths tend to decrease (Frenck-Mestre, 2005, p. 176). Thus, it is useful to measure fixations in particular AOIs in order to identify where test-takers' have more difficulties in comprehension. Also, regression indicates that the readers are double-checking the parts of the texts although some of it could imply some breakdown of comprehension. According to Rayner (1998), cited in Brunfaut & McCray in 2015, regression in more than ten-character spaces might be "an attempt to remedy the situation" (Brunfaut & McCray, 2015, p. 9).

In both studies, the eye-tracking device has proven to be an effective tool to describe some of the characteristics of test-takers in reading processes. Data from Tobii Technology made it possible to conduct experimental research without having to struggle with complex mathematical functions to visualize eye-movement and fixations.

3.2.3 Rationale behind the use of survey

One advantage of a survey is that it helps us to interpret the results from the eye-tracking experiment. The quantitative data collected by eye-tracking software helps to make inferences about eye movement and decision-making. Based on the assumption that there could be a reason why they were looking at particular points in the sources, a participant's fixation rate in a certain area of interest and their eye movement could be an indicator that they were processing the writing. It could, for example, be an indication that the test-takers were struggling to process the information or trying to incorporate the information into their writing, although it does not mean that they would necessarily be successful in synthesizing the information into writing. As discussed earlier, the eye-tracking data alone would not explain why they decided to do what they did during the process. In other words, the data requires further explanation to help understand the cognitive patterns displayed in the eye-tracking data. A survey was therefore also conducted to investigate what strategies participants used and what decisions they made in respect to their writing. Disadvantage of surveys may include skipped questions, accessibility and dishonesty. Such risks can be minimized by careful planning of the survey and by facilitating a focus group discussion. The focus group discussions were particularly useful in addressing the issues of skipped questions and dishonesty. The discussions helped to reveal which questions they might have misunderstood. Also, the discussion in the open and friendly environment helped more honest feedback.

In this study, the questionnaire consisted of nine sections, which included 41 items (See Appendix D). Likert scale questions (1-5) were created to generate numerical ratings, as well as yes-or-no questions to generate binary numbers. The questionnaire covered range of topics including computer literacy and graph familiarity (Yu, Rea-Dickins & Kiely, 2012), test-taking strategies (Phakiti, 2003; Bridges, 2010; Chen, 2012; Xu & Wu, 2012; Heo, Stoffa & Kush, 2012; Yang 2012) and the socio-cognitive processing of writing (Weir, 2005; Chan, 2013) as

shown in Table 4. The Research Instrument Section (3.4.3) discusses the relevant literature in detail.

Table 5. List of questions related to the three stages of integrated writing

Preparing-to-Write	Related survey questions	Question Number
Task Representation	Did you understand the instructions on how to write your essay? How easy or difficult did you find it to fulfil the task requirement? Which part of the task requirements did you find most challenging?	(Q3_1) (Q3_2) (Q3_3)
Macro-Planning [Response format]	Did you identify the purpose of the essay? Did you think about which solutions would work the best? Did you make an outline BEFORE writing your essay? Did you decide how many paragraphs there should be in your essay?	(Q4_1) (Q4_2) (Q7_1) (Q7_2)
Micro-Planning [Graph Information]	How easy or difficult was it for you to read the titles of the graphs? How easy or difficult was it for you to read the values on the graphs? How easy or difficult was it for you to read the units on the x- and y-axis on the graphs? How easy or difficult was it for you to interpret the information in the graphs? How easy or difficult was it for you to summarize the main trends of the graphs in English?	(Q5_1) (Q5_2) (Q5_3) (Q5_4) (Q5_5)
[Text Information]	Did you re-read the parts where you thought it was important to include in your essay? How did you decide which information to include from the texts? How did you connect the ideas from the texts?	(Q6_1) (Q6_2) (Q6_3)
Translation-writing		
[Fulfilling the task requirements]	How well do you think you have described the situation? How well do you think you have summarized the main points? How well do you think you have stated which solution would work best? How well do you think you have given the reasons for the choice made?	(Q8_1) (Q8_2) (Q8_3) (Q8_4)
[Translating-use of language]	Did you think what verb tense form to be used before writing? When did you decide which verb to use? Which tense did you mostly use? Did you think which pronoun form should be used for writing? Which pronoun form did you use mostly use?	(Q8_5) (Q8_6) (Q8_7) (Q8_8) (Q8_9)
Monitoring & Revising		
	I checked if my sentences were grammatically correct. I checked if my spelling was correct. I checked if I had connected the ideas from the graphs I checked if I put my ideas in a logical order. I checked if I had fulfilled the task requirements by going back to the instruction. I tried my best to avoid repeating the same word or expressions in the essay. I checked if my essay was an appropriate length. I used some sentences and phrases prepared in advance to be used in the essay.	(Q9_1) (Q9_2) (Q9_3) (Q9_4) (Q9_5) (Q9_6) (Q9_7) (Q9_8)

3.2.4 Rationale behind the use of focus group discussion

Furthermore, the focus group was used as a methodology because it is a format that encourages an open conversation about test-taking experiences. Focus group discussions allow "people to explore and clarify their views in ways that would be less easily accessible on a one to one interview" (Kitzinger, 1995. p.299). According to Kitzinger, they work particularly well when participants can engage themselves in open-ended questions to explore the issues on their own and using language with which they are comfortable. Focus groups have both benefits and drawbacks compared to conventional methods of interviewing participants. Benefits include the fact that focus group discussions can obtain data in a naturally occurring environment, which encourages participants to give honest feedback that is otherwise difficult to collect even in a face-to-face interview (e.g., Rabiee, 2004). In other words, focus group discussions allow students to see how their peers had a very similar or different experience or opinion on a topic without feeling different in sharing their views and experiences.

According to Folch-Lyon & Trost (1981), focus group discussions are not without drawbacks. For example, discussion groups could overshadow the voice of silent participants, depending on the group dynamics. They may be inappropriate when some sensitive personal or controversial issues need to be discussed (e.g., personal difficulties, illness, or one's sexual orientation). When discussing sensitive issues, focus group discussions may not be appropriate; however, this study found more benefits in the use of focus groups than drawbacks.

In the focus groups, the participants shared their views about their experience of taking the test (See 3.4.4 for details). The combination of data from the survey and the focus groups allowed to gain more insights into the participants' point of views. The combination of research methods, qualitative and quantitative, is often used to triangulate the data by assessing possible biases emerging from the survey which might quantify the biases (Goulias, 2000).

3.3 Participants

3.3.1 Sampling of participants

This study used the purposive sampling method, which is a type of non-probability sample that often consists of volunteers, or respondents who are hand-picked in the belief that they share particular characteristics (Cohen, Manion & Manion, 2013). In the first stage, 54 participants from four Japanese high schools in the western part of Japan (two from public schools and two from private schools in Kobe, Kyoto and Osaka Prefectures) agreed to participate in the study. Initially, only high-school seniors were targeted; however, it was later decided to recruit high-school juniors and freshman because they had a more flexible schedule than senior students. At the initial stage, the pool of participants consisted of 15 seniors, 29 juniors, 12 freshmen (a total of 56 students), all of whom signed the consent forms. For the eye-tracking experiments, however, this number reduced to 42 students (15 seniors, 19 juniors, 8 freshman), 32 of whom were female students. All these participants spoke Japanese as their first language and studied English as their foreign language. The participants' age ranged from 16 years old to 18 years old, which meant that they had been studying English for at least three years in the middle school environment.

3.3.2 Choosing participants for eye-tracking experiment

In the second stage, the researcher examined cognitive patterns by recruiting test-takers (N=42) who were available to participate in the study from the list of candidates. These forty-two students agreed to continue with the eye-tracking experiment, which entailed a commitment of approximately 120 minutes. The eye-tracking experiment involved collecting data from at least 84 eye-movement recordings to ensure an effective sample size. The detailed procedure is explained below in section 3.4.2 (see Table 6).

3.4 Research Instruments

3.4.1 (Phase 1) The Aptis test for describing English proficiency of participants

The Aptis test was used to gain insights into the participants' independent reading and writing abilities in English, since it is designed for the high-school age range and above and measures English proficiency in all four skills. While most official tests are conducted at designated examination centres on specific examination dates, the Aptis test offers flexibility to academic institutions in terms of when, where and for the number of candidates they want to test. The Aptis test gives both the numerical scores (0-50) and CEFR levels (A1-C) of the test-takers, and these can be used to compare the performance of the students with the TEAP writing test (designed for the A2 to B2 levels). In this study, the participants took the Aptis tests prior to the eye-tracking experiments, and the Aptis scores were used to run a statistical analysis to compare the associations with the eye-tracking variables between test-takers of higher and lower English abilities. For example, the Mann-Whitney U test was conducted using the Aptis reading and writing levels as independent variables and eye-tracking variables as dependent variables. This analysis helped to build a picture of the roles of reading and writing proficiency in the integrated writing processes. Since integrated writing, such as the TEAP Writing Task B test, requires both reading and writing skills, it was important to measure participants' abilities in reading and writing in order to understand the role of reading and writing in integrated writing.

The cost of the Aptis test was covered by a British Council Research Award (Nishikawa, Award recipient for 2015-2016). The Aptis paper test provided by the British Council, Tokyo, was administered by the author in December of 2015 and January of 2016. Figure 4 indicates the proportion of average Japanese high-school students at different levels on the CEFR scale (MEXT, 2014). As can be seen from the figure, 90 percent of high-school students are at or

below A2 levels. The findings from the Aptis scores revealed that the participants in this study had CEFR levels much higher than the average of the Japanese high-school students (only 20% was found to have a proficiency level below A2 in reading and 2.5% below A2 in writing), as described in Chapter 4 (Section 4.3).

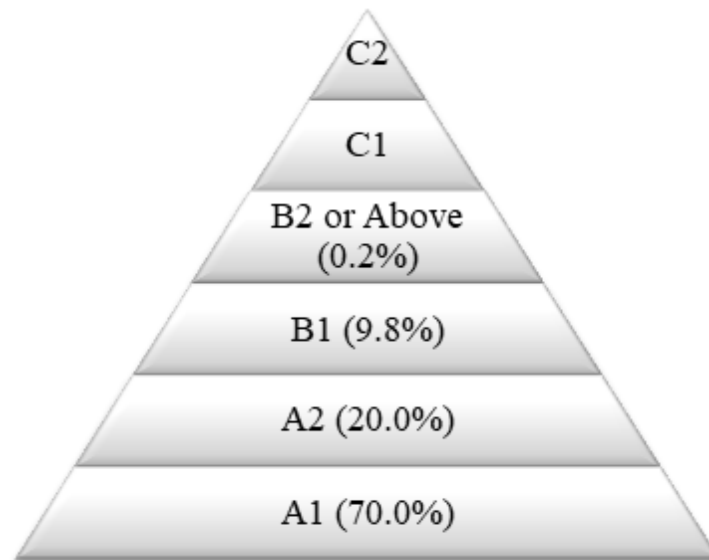


Figure 4. Average Japanese high-school students according to CEFR scale cited in (MEXT, 2014)

3.4.2 (Phase 2) Eye-tracking experiment using a Tobii Eye Tracker

In the second phase, the participants were tested for two different integrated writing tasks using the sample TEAP paper tests that were created for the purpose of this experiment with the help of the editing team at the Eiken Foundation of Japan. The experiment was intended to investigate the effects of the different types of source inputs (i.e., multiple source texts with two types of graphs) on test-takers' cognitive processes for integrated writing. Both of the tasks created for the eye-tracking experiments resembled the TEAP writing test Task B, which is an argumentative essay using two texts and two types of information from graphs (e.g., a pie chart and a bar graph). These sample tests are shown below in Figure 5 and Figure 6 (see the content

details in Appendix A for Task 1; Appendix B for Task 2). The paper version of the test has text appearing beneath the graphs, but this text was moved to the right half of the screen in order to give space for a composition on the screen, so that all components of the tasks could be presented in one screen. Both tests consisted of (1) Task Instruction, (2) Two types of graphs, (3) Newspaper article, (4) Letter to the editor and (5) the areas of essay composition.

How Students Spent After School Hours in 2015

Category	Percentage
Club activities	15%
Homework	25%
TV	35%
Volunteering	25%

Average Hours of Sleep among High-School Students in Greenhill

Year	Average hours of sleep
2000	9.0
2005	8.8
2010	8.5
2015	8.2

Education News

A new report found a worrying trend concerning teenagers in Greenhill. Mike Parker, the principal at North Greenhill High School, thinks that changing the daily routine after school is the first step in dealing with the problem. Parker talked about his ideas at a recent meeting with parents. "We need to educate our children about the importance of sleep," Parker said. "Average teenagers need about nine hours of sleep each night to feel well rested," he said. Parker wants to hold a special session for students to discuss this issue.

However, Parker suggested other steps are also necessary. For example, the government is considering changing school start times in high school. One reason is that high-school students tend to stay up late doing homework and other activities. "To be more realistic," said Parker, "this might be the only way to solve the problem. By delaying the start of morning classes by an hour, we can ensure that students gain extra sleep." Many teachers agree that students perform better when they are well-rested.

LETTER TO THE EDITOR

Dear Editor,

I am very concerned about the recent trend among adolescents regarding sleep. Based upon my many years of experience as a school nurse, I would like to offer some advice. First, exercise can help young people fall asleep faster and sleep more soundly. According to recent studies, it is better to work out earlier in the day than in the evening. So, it might be helpful to schedule morning exercise as a part of regular school activities in order to encourage adolescents to go to sleep before midnight.

I also believe the government should take action to resolve the situation. Some experts recommend adjusting the school times to fit the biological clocks of adolescents. Research has shown that students in classes with a later start time were twice as productive as those in normal classes. It may be worth considering changing school start times in Greenhill high schools.

I'm confident that parents and local educators will find effective ways to address this issue.

Sincerely,
Sarah Case

Figure 5. TEAP Task B Sample Task 1

Food waste in Greenfield (collected as garbage)

Year	Food waste (tons)
2000	100
2005	200
2010	300
2015	400

Sources of Food waste

Source	Percentage
Shops & Restaurants	35%
Households	25%
Schools	15%
Other	25%

City Council News

Members of Greenfield City Council met yesterday to discuss the issue of food waste in the city. Mike Parker, the head of the city's waste collection unit expressed his concern. "The city's garbage collection service has a very heavy workload," he said. He also pointed out that collecting and disposing of food waste has led to increased costs for the city. Parker thinks that targeting restaurants is the first step in dealing with the problem. "We need to encourage restaurants to throw away less food," he suggested, adding that "the only way to do this is to hit them where it hurts, in their pockets." To do this, he suggested introducing a special tax that would be calculated based on the amount of food waste a business produces.

Parker also put forward another idea. "The food that is thrown away by grocery stores cannot be consumed by humans," he said, "but it could be used for a different purpose." "I have a friend who is a farmer, and he told me he'd welcome the savasted food to feed his animals." Parker argued that this would be a good way to reduce costs for the council and benefit many local farmers. The city council will discuss this issue further at its next meeting.

LETTER TO THE EDITOR

Dear Editor,

I am very concerned about this recent trend in Greenfield. In my work as an environmental health officer I sometimes visit local stores. I'm always shocked by how much unsold food they throw away. I know that it is difficult for them to judge how much they will sell each day, but I believe we should try and find a way to reuse some of this food.

I also believe families in Greenfield must change their habits. When I visit my friends' homes, I'm surprised by how much food they discard without even thinking. I suggest it would be a good idea to have classes in schools to teach children about this issue. If children learn from a young age that wasting food is bad, they will grow up to teach their own children the same. This will also please parents, as their children will learn to throw away less food. I read a recent report that showed that families can reduce their monthly spending by twenty percent simply by wasting less food.

I'm confident that the city council will find effective ways to address this issue, and I hope my ideas will be useful for them.

Sincerely,
Sarah Case

Figure 6. TEAP Task B Sample Task 2

Tracking how long participants viewed the different types of graphs and the various parts of the texts may reveal valuable information about their decision-making process in regard to what information they decided to include in their essay. Table 6 below indicates the schedule for the data collection, which took approximately 120 minutes for each participant.

Table 6. Schedule for Eye-tracking Experiments and Survey

Stage	Purpose	Allocated time
Introducing Procedures	Collecting a consent form	5 minutes
	Explaining the procedure	
Getting Started	Calibration	5 minutes
Collecting Data	Task 1	40 minutes
Break		5 minutes
Collecting Data	Task 2	40 minutes
Conducting surveys	Collecting a form of cognitive processing questionnaires	5-10 minutes

3.4.3 [Phase 2] Analytical scores of the essays using the rating scales

The sample TEAP Writing Task B, which featured multiple source texts and information from graphs, uses five criteria: (1) Main Idea, (2) Coherence, (3) Cohesion, (4) Lexical Range & Accuracy, and (5) Grammatical Range & Accuracy. The eye-tracking variables were used to analyse the association with the participants' reading and writing abilities based on the Aptis test that they had taken in the weeks before the eye-tracking experiments. As noted in the literature review, the TEAP test was developed from Weir's cognitive framework based on cognitive, content and scoring validities (See Section 2.4).

The rubric, which was taken from a publicly available online source from the Eiken Foundation of Japan, was modified so as to break it down to nine numerical scales, from 0 to 8. The scores were rated numerically to help quantify the rating for further statistical analysis with eye-tracking data. Two raters, each with professional experience in teaching English as a second language and marking experience of other test products, were given separate rater training, and a chance to review and discuss sample ratings with each other before assessing the written essays. The inter-rater reliability (i.e., correlations between Rater 1 and Rater 2 ratings in the

five criteria) was tested by calculating a correlation coefficient of the two sets of scores rated by the two raters for the same essays. The reason for adopting the analytic sub-scores in this study rather than a holistic score was because it was necessary to find out if there was a relationship between the level of English performance in the five components of the TEAP writing construct and the eye-tracking variables during the writing processes. Analytic scores help understand construct-relevant variation in the essay scoring and are also more accurate. In fact, some studies have argued that a holistic score tends to be more subjective than an analytic score due to the different types of strategies used by raters in their decision-making processes (e.g. Barkaorui, 2011, In the current study, the essays that had less than 50 words were graded as A2 or below (i.e., A1) or zero (0) in the rating. Also, the essays that failed to synthesize information from both texts and graph information were not marked above B1 or B2 level. After discussing all five written constructs areas of the TEAP test, the two raters then also had a chance to mark the sample essays and discuss the results. When the marks for Task 1 and Task 2 were combined, two students received A2 or below, 22 students received A2, 15 students received B1, and one student received a B2 score.

Table 7. TEAP Writing Task B, Construct and Rating Scales Adopted from Eiken Foundation of Japan website

Raw Score	<i>Main Ideas</i>	<i>Coherence</i>	<i>Cohesion</i>	<i>Lexical Range & Accuracy</i>	<i>Grammatical Range & Accuracy</i>
+ 3 —	Synthesizes and evaluates information and arguments from all of the verbal and nonverbal input texts.	Organized as a coherent response to the task; organization of ideas within and across paragraphs is generally clear, though may be formulaic	Uses discourse markers and referential cohesive devices effectively to mark the relationship between sentences and link utterances into clear, coherent discourse.	Uses appropriate synonyms and alternative expressions to convey the main ideas.	Uses a range of sentence structures appropriately; grammatical errors rarely occur and do not impede understanding of the message.
+ 2 —	Provides a basic summary of some of the main points, bringing together information from more than one of the input texts.	Has a logical structure but the organization of ideas may not always be clear; organized into paragraphs, but the paragraph structure may not be completely appropriate.	Sentences and paragraphs are generally connected using discourse markers; use of referential cohesive devices (for example, pronominal reference) is mostly clear.	Gives a basic description of the main ideas in the input texts, but tends to rely on the vocabulary supplied in the input texts. Some inappropriate vocabulary usage is evident.	Grammatical errors occur frequently but tend to be associated with attempts at complex structures and do not impede communication of the message.
+ 1 —	The response refers to some of the elements or points mentioned in one or more of the input texts (verbal and/or non-verbal) but does not synthesize these points or make clear how they are related.	No logical paragraph structure or some separation which is not appropriate; text consists of mainly unconnected sentences with no clear direction or progression across sentences.	Uses conjunctions to link clauses within sentences, but generally does not mark clearly the relationship between sentences. Use of cohesive referential devices (for example, pronominal reference) is generally not clear.	Usage of paraphrasing and synonyms is extremely limited, and alternatives are not appropriate for the task. Errors and unnatural/inappropriate usage common when reusing vocabulary from the input texts.	Grammatical errors occur systematically and may impede communication of the message.
0	<p>Has almost entirely copied from the input text or the number of words is too short (fewer than 50).</p> <p>Is written on a topic different from those assigned,</p> <p>Is connected to the prompt so loosely that the essay could have been prepared in advance,</p> <p>Requires considerable effort to see any connection being made between the composition and the prompt.</p>				

3.4.3 (Phase 3) Cognitive Processing Questionnaire

The cognitive-processing questionnaire in this study also has a section dedicated to investigating the participants' familiarity with computers and with graphs.

Test-taking strategies & graph familiarity:

The questionnaires were developed by referring to some of the inventories used to investigate test-taking strategies in previous studies (e.g., Phakiti, 2003; Bridges, 2010; Chen, 2012; Xu & Wu, 2012; Heo, Stoffa & Kush, 2012; Yang 2012). The contents of the survey were also tailored, however, to match the specific task descriptions of the TEAP test based on Weir's socio-cognitive model (See Appendix C). The survey was conducted with Likert-scale questions immediately after the two eye-tracking recordings (see Appendix D). The questionnaires were mostly composed of 5-point Likert-scale questions which covered cognitive processing, graph familiarity and test-taking strategies. Because this study looks into the effects of graphs and multiple source texts, there were also questions about participants' daily involvement with the use of charts and similar diagrams for other academic courses, in order to validate some of the findings by other scholars (e.g., Yu, Rea-Dickins & Kiely, 2012).

The questionnaires were organized by referring to Weir's socio-cognitive model as a theoretical foundation to help explain some of the quantitative findings from the eye-tracking data and the rationale behind some aspects of the test-takers' decision-making. Thus, the questionnaires were divided into sections of task representation, macro-planning, organization, micro-planning, translating (putting ideas into linguistic forms), and monitoring and revising.

Computer Literacy:

According to Houser & Thornton (2004), Japanese college students typed much faster on mobile devices using their thumbs than typing on PC with their both hands laid on the keyboard. This implied that high-school students, who use pens and notes in classrooms, would require

some kind of training prior to take any computer-based tests. In 2011, the MEXT published its initiative with "The Vision for ICT in Education" which was re-endorsed by the action of Cabinet's members as the "Japanese's Revitalization Strategy: Japan is Back." (Oshima & Muramatsu, 2015). Despite all these efforts, the Executive Summary Report of 2011 published jointly by the Japan Association for Promotion of Education and Technology (JAPET) and Microsoft, Japan offers a grim picture of the current situation of ICT use in Japanese primary, middle and high school classrooms. For example, only 13 percent (45) of teachers in the 258 high schools who participated in the study reported that they use ICT as a part of their curriculum in the classroom. Accordingly, the researcher did not take it for granted that Japanese high-school graduates would have a decent computer literacy. Given these circumstances, the questionnaire also asked a total of seven questions about computer literacy, particularly familiarity with typing on keyboards so as better to understand their writing processes.

3.4.4 (Phase 4) Focus group discussions

Due to the time constraints, it was difficult to replay the eye-tracking recordings on the spot. Without providing some evidence from the experiment, it would be difficult to conduct meaningful discussions, and it was therefore more practical to ask the participants to meet again a few days later (Total of 24 students, 9 from public school and 15 from private schools. The detailed description of their language levels is shown in Table 72). The participants were therefore invited to join the focus-group discussion on another day, and all the students who participated in the study were invited, although it was not mandatory. Twenty-four students participated in the focus group discussions. Groups of four or five students discussed their test-taking experience through semi-structured questions and open discussions. The initial analysis of the eye-tracking data informed the design of the questions for the focus group discussions (See Appendix F). In order to ensure productive discussions, the researcher played a role as a

facilitator to promote an environment in group discussions in which the participants did not have to feel intimidated or singled-out in sharing their views and experiences. By showing some evidence taken from the eye-tracking experiments on the computer monitor (e.g., video, heat map and gaze plots from some selected participants), the open-guided questions were posed by the researcher to validate whether some of the findings from the heat-map and gaze plots were an accurate reflection of their experience. Their comments and feedback were transcribed and then categorized according to their language proficiency levels to report the most common views of the participants.

3.5 Methods of Data Analysis

This section describes the specific methods of data analysis. It includes what kind of data the eye-tracker produced and the types of analysis for eye-movement data, the data from the cognitive processing questionnaires as well as from the focus group discussions.

3.5.1 Eye-movement data

The *Tobii TX 300* works in tandem with the *Tobii Studio* software, which allowed some of the collected data to be visualized in the forms of heat maps (Figure 7) and gaze plots (Figure 8), as described below. In addition, these data were exported to an Excel sheet which was then imported into IBM SPSS software for statistical analysis. The *Tobii Studio* allows eye-movement measurements in the Areas of Interest (AOIs) to be calculated. AOIs are the specific areas in displays that we are interested in so that certain eye-movement patterns can be studied in those areas (e.g., Fixation Duration, Fixation Counts). Table 8 lists some of the eye-movement variables that could be collected from the eye-tracker (*Tobii Studio* User Manual Version 3.3.2, pp.101-111). This section describes all the data analysis that the current study

uses to explore the effects of source texts and graph information during integrated writing processes.

Table 8. Eye-tracking metrics

Variables	Description
Time to First Fixation	Time from the start of the media display until the participant fixates on the AOI for the first time (seconds)
Fixation Counts	Number of times that the participant fixates on an AOI
First Fixation Duration	Duration of the first fixation on an AOI (seconds)
Fixation Duration	Duration of each individual fixation within an AOI (seconds)
Total Fixation Duration	Duration of all fixations within an AOI (seconds)
Visit Duration	Duration of each individual visit within an AOI (seconds)
Total Visit Duration	Duration of all visits within an AOI (seconds)
Visit Count	Number of visits within an AOI
Total Visit Count	Total number of visits within an AOI

Reference: *Tobii User's Manual Ver. 3.4.5 (2016)* extracted from

Time to First fixation is measured from the start of the time the media has begun recording until the time when the participant's eye first fixated on the AOI. The difference between Fixation Duration and Visit Duration are as follows: fixation duration is the sum of the duration of all the individual fixations in the AOI whereas visit duration refers to all the fixations that were recorded in one visit within the AOI (Kim et al., 2012). These variables are useful for understanding participants' decision-making processes. Researchers typically investigate reading behaviours such as regressions, saccades and return sweeps. For example, the past literature, including Bax (2013), suggests that regression may be indicative of readers having difficulties in reading, and saccades could be a sign of expeditious reading (see more details in section 3.2.2).

Gaze-plots and Heat Maps: There are a number of ways to visualize the eye-recording data using *Tobii Studio*. For example, a heat map can be created for a participant given a specified start time and duration within the recording. Heat maps allow the easy visualization of where the participant viewed most. While heat maps show the intensity of the fixations, gaze plots give more precise information as to the direction and the order in which a participant has viewed the screen, with the size of each circle indicating how long the eyes were fixated at each point. Both these techniques help provide insights into the cognitive processes of test-takers during the reading-into-writing tasks.

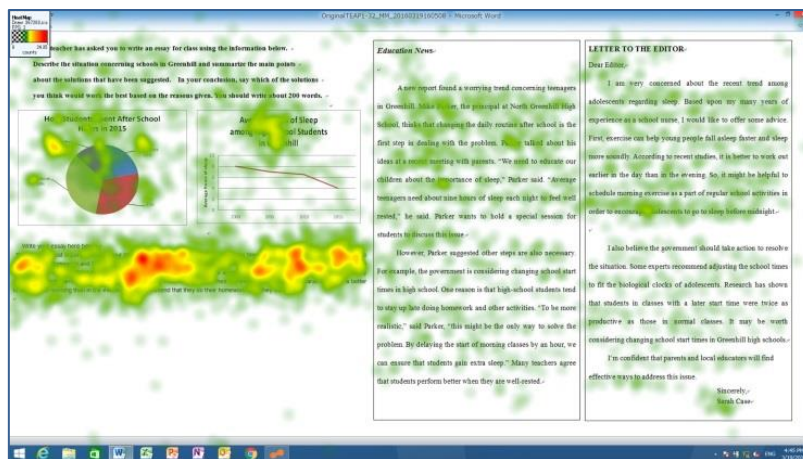


Figure 7. Example of a Heat Map

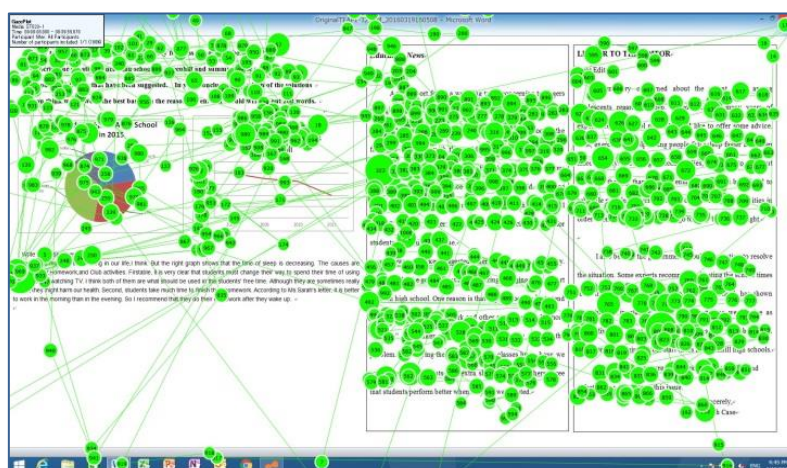


Figure 8. Example of a Gaze Plot

Areas of Interests (AOIs)

Defining and quantifying the eye-movements in the areas of interests makes it possible to compare and find a pattern of behaviours among the participants. Here, Task 1 and Task 2 consisted of six different AOIs, as set out in the figures below (Task Instruction, Pie Chart, Line Graph, Bar Chart, Newspaper Article, Letter-to-the-Editor and Essay Composition).

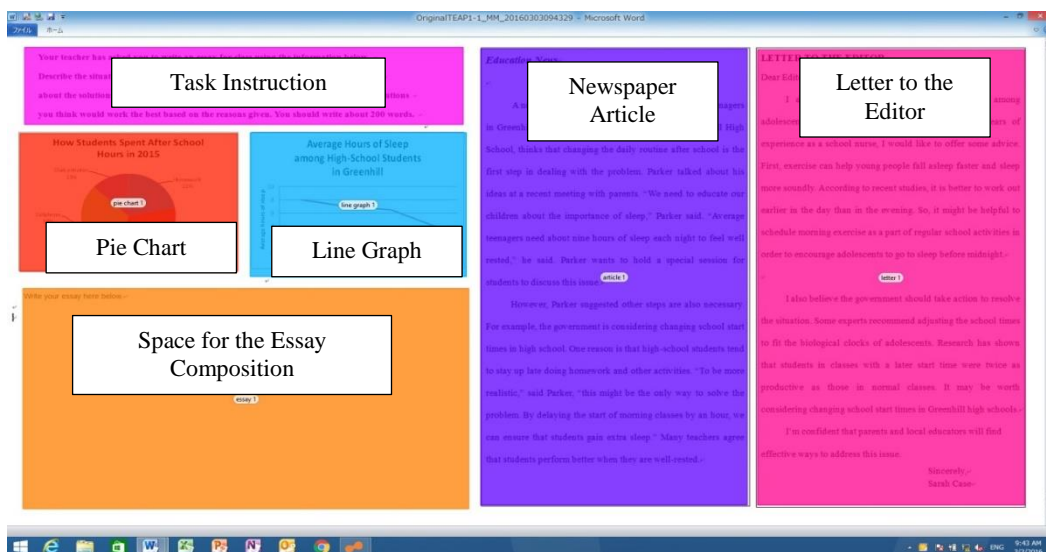


Figure 9. Areas of Interests for Task 1

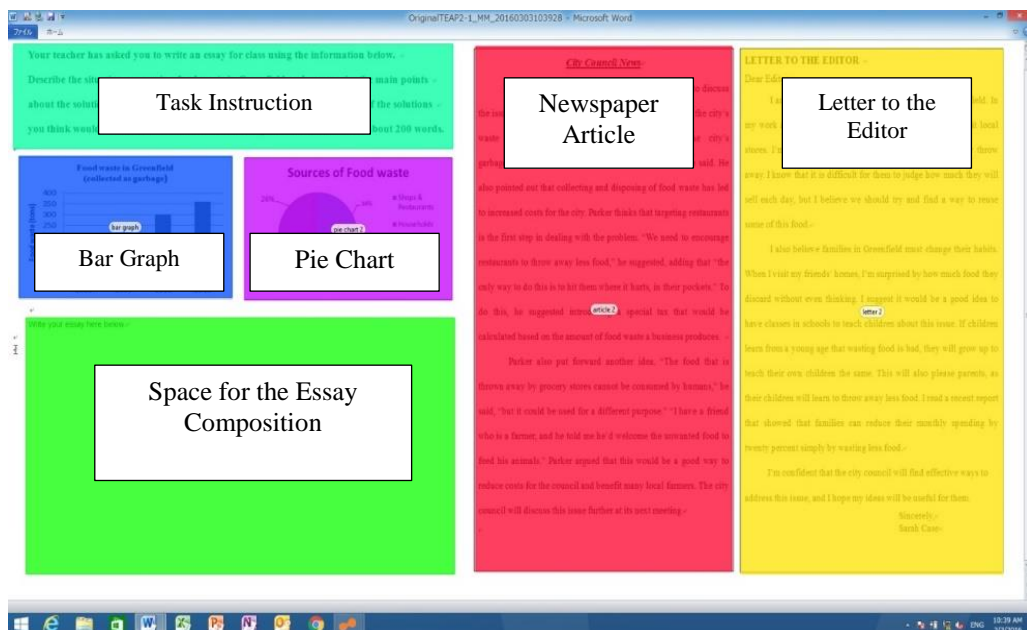


Figure 10. Areas of Interests for Task 2

To answer the five sub-research questions, the AOIs were further divided into sub-sections, as shown in Figure 11 for Task 1 and Figure 12 for Task 2. The sub-sections included the title for each graph, the heading of each source text, and the paragraphs of source texts and the signature.

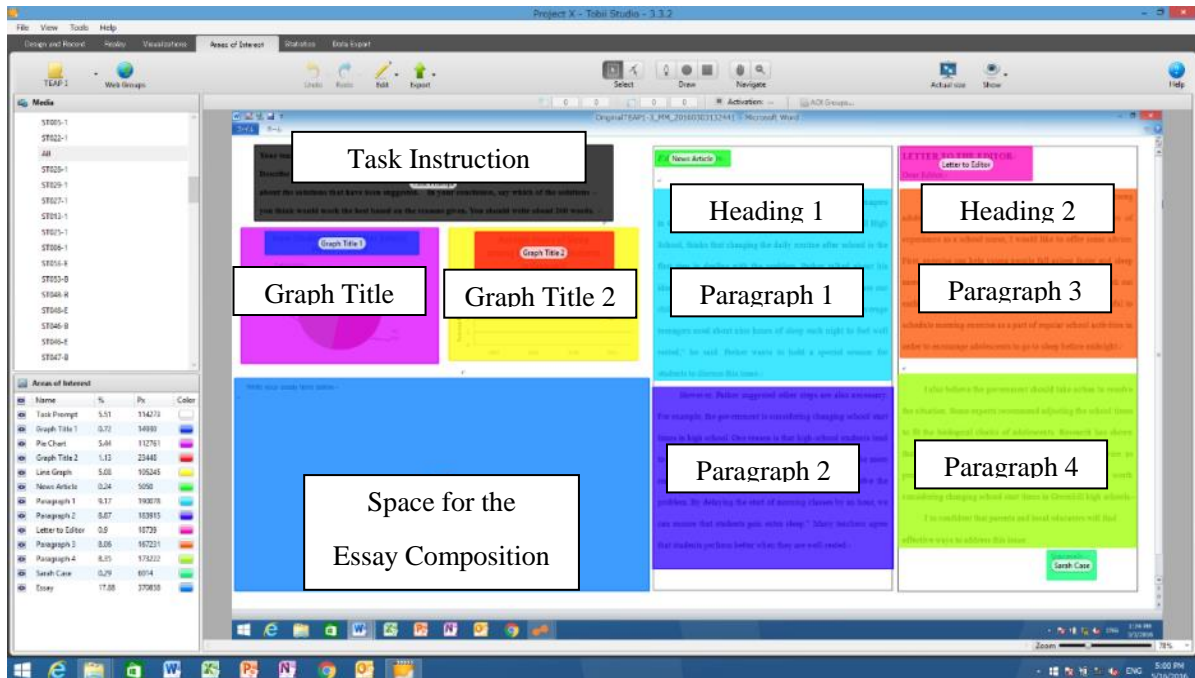


Figure 11. Sub-sections of the Areas of Interest for TEAP Sample Task 1

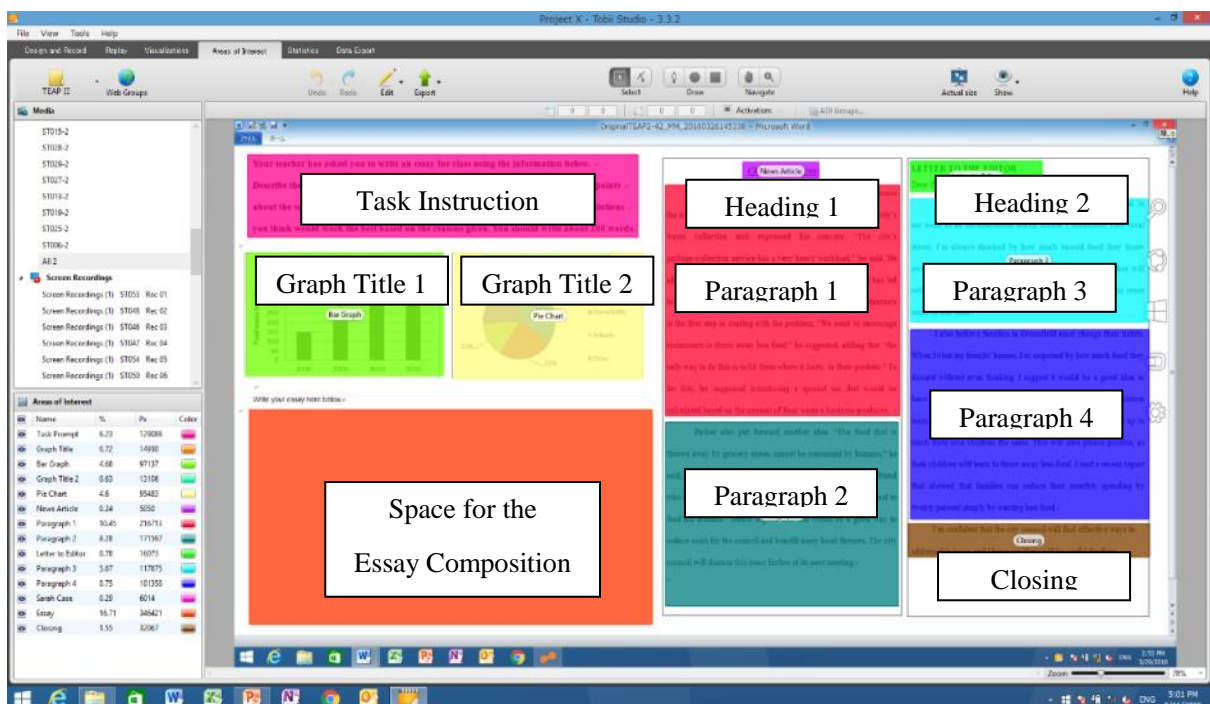


Figure 12. Sub-sections of the Areas of Interest for TEAP Sample Task 2

3.5.2. Statistical analysis

AOI Switches: AOI switches were measured in order to identify the paragraph reading behaviours during the first ten minutes of recordings. The Tobii Eye-tracker (*TX300*) gives 300 data samples for each second. This means 300 columns of Excel data were reported in each second of its recording, which amounted to 1,800 data samples in one minute and 72,000 for a ten-minute recording. The *Tobii Studio* allows fixation data to be exported as binary numbers of either 0 or 1 in time index for specified areas of interests. From the data, it was calculated in which direction the fixation moved, from one area to another. The number and the direction of the AOI switches were compared during the first ten minutes of recording so as to study some of the participants' reading behaviours during the initial cognitive processes for integrated writing in the areas of interests (i.e., Paragraph 1 (P1), Paragraph 2 (P2), Paragraph 3 (P3), Paragraph 4 (P4) and any other areas (defined as PX)).

The assumption was made that the longer eyes were fixated in an AOI, the more the participants were either having difficulty understanding the AOI or getting more information from the corresponding AOI. The challenge, therefore, was identifying whether participants were looking at the graph longer because they were confused by it and trying to work out its relevance (i.e. failing to acquire useful information from it); or because they fully understood it and its relevance to the test and were thus focusing in order to gather information useful to writing an answer. In other words, the same fixation data could have two completely opposite interpretations. This would be dealt with by analysing the data according to the participants' reading and writing proficiency levels.

Saccades: According to Frenck-Mestre, the level of difficulties readers face in the reading process can be predicted by counting the numbers of re-fixations, or the length of saccades between fixations. Fixations and re-fixations can provide clear indications of readers'

processing of the information (e.g., *saccades* (jumps), *fixations* (stops) and *regressions* (re-takes) (Frenck-Mestre, 2005, p. 176). Bax (2013) used the findings of Rayner (1998) as a benchmark for the measurement of the eye-movements and reported that the average eye-fixation typically falls between 200 and 250 micro-seconds and that the mean saccades are about 7 and 9 letter spaces (Bax, 2013, p. 444). If the average saccade was less than 7-9 letters, then it would indicate that readers were having difficulties with the source texts and graphical information.

Wilcoxon signed-ranks test: A non-parametric test, which is the equivalent of the parametric paired sample t-test, was performed to understand the differences in the same participants' eye-movements in AOIs between Task1 and Task 2. The Shapiro-Wilk test confirmed that the eye-tracking variables collected for Task 1 and Task 2 were not normally distributed ($p < .05$). This, combined with the relatively small sample size, led to the selection of the non-parametric test for the statistical analysis. Both Task 1 and Task 2 were designed with similar formats. The only major difference was that the Task 1 included a pie chart and a line graph while Task 2 included a pie chart and a bar graph. By conducting the Wilcoxon signed-ranks test for different types of graphs (Pie chart vs. Line graph in Task 1, Line graph in Task 1 vs. Bar graph in Task 2, and Bar graph vs Pie chart in Task 2), it was possible to compare the differences in the eye-tracking metrics between the two test features as well as the familiarity of tasks.

Mann-Whitney U test: Mann-Whitney U tests for non-parametric data were conducted to find any significant differences in the eye-movement variables over AOIs between the upper-intermediate and lower-intermediate participants for the overall 40 minutes of the test, the first 10 minutes and the last 10 minutes of recordings. The rationale behind the use of different segments of tests for analysis was based on the assumption that the test-takers would show different behaviours during the integrated writing processes. The TEAP integrated writing tasks were developed by referring to Weir's socio-cognitive framework (2005) which consists

of three main writing stages: pre-writing (i.e., macro-planning, organization, micro-planning), during writing, and post writing (i.e., monitoring and revising) activities. By observing how long it took, on average, for the test-takers to study the task instruction and source information, the researcher felt the first 10 minutes and the last 10 minutes would give an insight into their cognitive processes and operations during the 40-minute tests.

Using the Aptis reading and writing scores, the students were classified into upper-intermediate (B2 above levels) and lower-intermediate (B1 below) groups. The scores were used as predictors to run a statistical analysis of the differences between the groups, with the intention of understanding the differences in the underlying cognitive processes of the writing process, as shown in Table 4. Similar methods are also adopted in Shaw & Weir (2007), Weir (2014) and Brunfaut & McCray (2015). The description of the measurements guides the investigation of the differences in the cognitive patterns between the upper intermediate (B2 above) and lower intermediate (B1 below) level participants for integrated writing tasks.

Multiple linear regression test: Using both Aptis reading and writing scores as predictors, multiple regression tests were conducted to understand the separate effects of both reading and writing proficiencies on the dependent variables that were collected from the eye-tracker for the overall 40 minutes of the test, the first 10 minutes and the last 10 minutes. If a coefficient was significant ($p < .05$), then the eye-movement dependent variable is indicated to be closely related to a student's ability in reading or writing. By comparing how many significant coefficients were found on reading and writing proficiencies in different segments of eye-recordings (e.g. the first 10 and the last 10 minutes), the findings helped identify test-taking behaviours and cognitive processes in integrated writing processes that are related to either reading or writing proficiency (or both).

Pearson Correlation: The essays that the participants produced during the two eye-tracking experiments were rated using the TEAP rating rubrics. The analytic scores were based on five criteria (i.e., Main Idea, Coherence, Cohesion, Lexical Range and Accuracy, Grammatical Range and Accuracy). The Pearson Correlation tests were used to investigate the associations between the essay scores and the Aptis reading and writing scores. Also, Pearson correlation tests were conducted to find relationships between eye-tracking metrics and the analytical scores. The results were intended to allow an analysis of which of the eye-tracking metrics were associated with the scores in the five criteria.

3.5.3 Cognitive processing questionnaires

Chi-square tests: The results of the questionnaires were used to compare the differences between the groups of successful (B1 and B2 levels) and less successful (below A2 and A2 levels) writers using the TEAP analytic scores. Also, cross-tabulations made it easier to compare answers across participants of different language proficiency levels. Chi-square tests were conducted to acquire overall trends in the responses to each question, thus helping to understand how decision-making processes and test-taking strategies during the integrated writing task differed between more- and less-successful test-takers.

3.5.4 Focus group discussions

The participants were invited to discuss their test-taking experience and receive feedback. The meeting was scheduled some days after the eye-tracking experiments. Appendix E has a list of questions that were discussed. Sixty minutes were allocated for the focus group discussions at each venue. The participants' remarks in response to the focus group discussions were summarized according to their language proficiency levels in CEFR (A2-below, A2, B1 and B2) based on their TEAP essay markings. The tables allowed their answers to be compared to the questions during the group discussions across their writing proficiency levels which helped

to identify their decision-making processes and the areas of difficulties at each language proficiency level.

Six students participated in the Focus Group Discussions on March 9, 2016, and eighteen students participated on April 12, 2016. In both cases, the discussions were conducted by forming a small group of four or six students in each group. High-school English teachers did not attend this meeting so as to facilitate an authority-free environment where the participants could give any feedback or comments. The researcher took the role of a group facilitator of the discussion and asked the groups to discuss a topic written on the paper and write down some opinions on the paper as they openly discussed. In this way, it was ensured that even the students who were not always vocal in giving opinions in groups had something to contribute to the session.

Focus groups allowed the students to see the extent to which their peers had similar or different experiences or opinions on the topic. Instead of an individual interview, in which students might be unsure whether their experience was only theirs, the focus group discussions were intended to give a meaningful and positive feedback for all the participating students without feeling alone. By creating an honest environment in group discussions, the participants did not have to feel intimidated or singled-out in sharing their views and experiences. In one instance, the students already knew each other because it was such a small group before the discussion took place. In another case, in which there was a mixture of students from different grade levels, they did not know each other until the time of the discussion. In both instances, the students were immediately connected as they were asked to share the experience of going through the same test material. They could build a rapport as they discussed the experience they were all familiar with.

The findings of focus group discussions (Chapter 7) and some of the case studies (Chapter 6) were intended to deepen an understanding of the quantitative findings (statistical data analysis). Table 9 explains how the different sets of statistical data analysis will be integrated to address the research questions for this study.

Table 9. The Overview of Data Analysis

Research Questions	Statistical Data Analysis
RQ1. To what extent do test-takers incorporate information from multiple texts and graphs?	-Wilcoxon signed-ranks tests to help identify the difference between how participants viewed AOIs between Task 1 and Task 2. (Chapter 4) -Chi-squared test on how they answered the questionnaires relating to RQ1 (Chapter 5).
RQ2. To what extent do features of the graphs (e.g. line graphs and bar graphs) affect cognitive processes of integrated writing tasks?	-Wilcoxon signed-ranks tests to help identify the difference between how participants viewed AOIs between Task 1 (line graph) and Task 2 (bar graph). (Chapter 4) -Chi-squared test on how they answered the questionnaires relating to RQ2 (Chapter 5).
RQ3. What role does language proficiency play in integrated writing tasks?	-Mann-Whitney U test to compare two groups of participants who have upper and lower-intermediate English proficiencies. (Chapter 4) -Multiple Regression Analysis, using both Aptis reading and writing scores as to identify whether language levels in reading and writing can separately predict the eye-tracking variables. (Chapter 4) -Pearson Correlation test of the TEAP analytic scores to writing scores? (Chapter 4) -Pearson Correlation test to find which of the TEAP writing scale criteria are closely related to eye-tracking variables (Chapter 4) -Chi-squared test to compare the differences between upper and lower intermediate groups based on how they answered the questionnaires relating to RQ3 (Chapter 5).
RQ4. What kinds of test-taking strategies are used for integrated writing tasks?	-Chi-squared test on how they answered the questionnaires relating to RQ4 (Chapter 5).

3.6 Theoretical Perspectives and Researcher Positionality

Since educational research is a part of social science, which “involves the collection of data about people and their social contexts,” there are a variety of research methods available that reflect different philosophical understandings of the nature of knowledge and truth

(*epistemology*) and being (*ontology*) (Somekh & Lewin, 2011, p. 2). Understanding one's philosophical orientation is essential, as it underpins the nature of the research that a researcher will bring to the study (Creswell, 2013; Crotty, 1998).

Epistemologically, it first appeared that this research was aligned with objectivism, which views "objects in the world" as having "meaning prior to, and independently of, any consciousness of them" (Crotty, 1998, p. 27). From this viewpoint, knowledge develops based on careful observation of the behaviour of individuals and the measurement of objective reality. These two variables need to be tested or refined in order to understand the world (Creswell, 2013), and these views may be consistent to some degree with the second phase of the current study which uses a quantitative approach to deduce patterns and characteristics of successful and less successful test-takers during integrated writing tasks. The theoretical perspective in the third phase of the study, however, was more aligned with constructivism, since each case had to be examined to understand what went on the participants' mind by surveying and focus group discussions. In the view of constructionists, "all knowledge, and therefore, all meaning and reality as such, is contingent upon human practices, being constructed in and out of the interaction between human beings and their world, and developed and transmitted within an essential social context" (Crotty, 1998, p. 42). In other words, constructionists believe there are multiple realities, and that the true meaning of the world cannot be found either objectively or subjectively.

After some consideration, my overall philosophical framework was similar to those of pragmatists, who emphasise what is practically useful for any given piece of research and are thus open to combining both quantitative and qualitative data. A pragmatic perspective is consistent with this study since its research design incorporated both objective and subjective knowledge, and thus the work was not committed to a particular philosophical system or

understanding of reality. Pragmatists believe that whatever works best at a given time for understanding problems should determine the research design. Also, they place emphasis on what and how to research in a manner that best matches the needs and purposes of a particular study, using both quantitative and qualitative data (Creswell, 2011).

In the view of a pragmatist, there is no such thing as pure “objectivism” in studies in social science. “The nature of qualitative research sets the researcher as the data collection instrument. It was reasonable to expect that the researcher’s beliefs, political stance, cultural background are important variables that affect the research process” (Bourke, 2014, p.2). It is therefore very crucial for researchers who deal with qualitative data, even in a mixed methods studies, to be self-consciously aware and transparent about their own positionality to both participants and readers.

In particular, the relationship between the researcher and the participants might affect the research outcome. In this regard, researchers are often considered as being either “insiders” or “outsiders”, and Bourke (2014) has discussed the complexity of the boundaries between these two positions.

In this context, prior to this research, there were some fears that my position as a university English instructor might influence how students would behave in the focus group discussions (i.e. that the participants might give answers that they feel to be appropriate for someone in my position). These risks could not be completely eliminated and therefore the focus was on minimizing such risks by being sensitive to these positionality issues and by taking some precautions by explicitly announcing to the prospective participants that their responses would not affect participants’ academic performance or results. The positionality issues were taken into consideration throughout the research process.

3.7 Weakness and Strengths of the Chosen Approach

This research design reflects a pragmatic approach, employing mixed methods to investigate the cognitive processes of L2 learners while synthesizing multiple texts and information from graphs into their writing. The strength of the approach came from the fact that this study used a mixed methods explanatory design. This research design allowed the researcher to triangulate both quantitative and qualitative data collected from the use of eye-trackers, the cognitive processing questionnaire, and focus group discussions to study the test subjects in depth.

The weakness of this approach is acknowledged in the limitations of the study. There were three main areas of concern. One was the generalisability of the findings. Since this study used purposive sampling and its sample size was only up to 42 for each task feature, the findings could not be generalized to the overall population of high-school students. The second concern was the time-consuming nature of the experiment. Since the experiment had to be done only one participant at a time, careful planning was needed to minimize the time required for the participants to take part in this study. Lastly, a major concern was the cost of the experiment. The rental of a *Tobii TX300* Eye Tracker cost around three thousand Pounds per month. The Eiken Foundation of Japan funded this part of the lease.

3.8 Ethics

Ethical concerns need to be addressed for any research project. Kumar reviewed several ethical issues in social science; (1) determining the stakeholders in the research, (2) determining ways to collect information, (3) maintaining confidentiality throughout the research process, and (4) avoiding bias for data analysis (Kumar, 2014, pp. 282-289). These ethical issues were reviewed before, during and after my research.

3.8.1 Determining the stakeholders in the research

According to Kumar (2014), it is important to ensure that research is not affected by the self-interest of any party. The stakeholders in this study were high-school students who volunteered to take part in the experiment. The researcher has an ethical responsibility not to expose these student participants to excessive mental stress because this could affect both the well-being of participants as well as the result of the study. This was avoided by informing the participants, through a consent form, that the scores of the Aptis test and essays submitted for eye-recordings would not influence their grades at school and that all the collected data would be kept anonymous.

3.8.2 Collecting information

Kumar (2014) states that it is also important, when obtaining informed consent, to provide participants with a clear aim of the study, including (1) which types of information will be collected, (2) the reason why this information is sought, (3) what is expected of the participants, and (4) the ways in which the results will be made available to the public (Kumar, 2014, pp. 282-289).

In this study, the participants were asked to take the Aptis Reading and Writing package test and were assigned an integrated writing task to measure their cognitive processes on a computer monitor screen. Then, they completed a survey and took part in a focus group discussion a few weeks later. The experiment with an eye-tracker was run twice on the same day. To compensate them for their time, I gave some feedback to the participants with some useful tips for preparing for integrated writing tasks, as well as pre-paid gift cards for 500 yen and some stationery goods.

3.8.3 Maintaining confidentiality

As shown in the consent form (Appendix G) and the information sheet (Appendix H), the participants were guaranteed their right to anonymity in respect to all the data collected, including test results, survey responses and feedback in the group discussion. To ensure anonymity, the researcher used pseudonyms for data collection purposes. The participants had the chance to discuss any questions they might have, including how they would be involved in the research process. The participants were informed of their rights to withdraw at any point during the data collection sessions without giving any reasons. In the event of a participant deciding to withdraw during or within one week after the data collection, the data that the participant provided were destroyed in a secure manner.

Furthermore, the teachers who referred the students taking part in this study were told they would not have access to the personal information in these results. The general description of data might be accessible to them, but school authorities were not permitted to gain access to individual scores from this study. There was a separate letter of consent sent to teachers in this research to inform them of the intent of this study, its procedures and implications for their benefit. A summary of overall trends was reported to each teacher. These ethical concerns have been discussed and summarized in the Ethics Form (Appendix J.)

3.8.4 Avoiding bias

The final ethical consideration was the limitations of this study, including the sample sizes and sample errors for data analysis. This study employed purposive sampling since this allowed the researcher to examine the target population. In purposive sampling, it is important to prescribe sample criteria to define and prioritize the quotas for selection (Ritchie, 2003, p. 107; Ritchie, Lewis & Elam, 2003). Even then, the study may not justifiably be generalized to a wider population. Thus, reliability and the competence of the information need to be carefully

reviewed. One expected source of bias comes from the fact that the participants in this study represent only those who are good at English compared to the public, and the data collected from these students might be biased to represent only high-proficiency students. To keep such bias at a minimum, participants were chosen not from a single high-school but from four schools (two private, two public) from different districts. Also, it was made sure to recruit students with a broad range of general English abilities.

3.9 Conclusions

This chapter has described the multi-phase research design of this study and its mixed-methods methodology. The current study has four phases of data collection: 1) a description of the participants' language proficiency in reading and writing from the Aptis test, (2) the completion of two integrated writing tasks in which eye movements were measured using the *Tobii* eye-tracker, and performance rated according to five criteria, (3) the identification of findings about the decision-making processes by means of cognitive processing questionnaires, and finally (4) the revelation of the test-takers' experiences through focus group discussions. The results of each of these phases are now reported in four separate chapters. Chapter 4 reveals the overall trends by analysing the language proficiency and eye-movement data collected for Task 1 and Task 2. Firstly, the Wilcoxon signed-ranks tests helped identify the difference between how participants viewed AOIs between Task 1 and Task 2. Then, Mann-Whitney U tests showed the roles of reading and writing in cognitive processes during integrated writing, and the differences in eye movements between test-takers of lower and higher proficiency which also helped answer the case analysis in Chapter 6. Furthermore, the chapter reveals what predicts performance in integrated writing by evaluating the test-takers' essays based on the five criteria used for the TEAP Writing Task B rating scales and relating these to the eye-tracking variables. Chapter 5 then discusses the results of the cognitive-processing questionnaires for the three main stages in the cognitive process of writing by referring to the theory proposed by Shaw

and Weir (2007) to seek insights into those cognitive processes. Chapter 6 reports several individual cases in order to offer some interpretations of the eye-tracking data. Chapter 7 discusses the results of the focus group discussions to offer a deeper understanding of the students' test-taking experiences. As a reference, Table 10 below summarizes the working measurements for the various analyses and tests in each chapter.

Table 10. Summary of Working Measurements

Finding	Variables	Analysis
Chapter 4 Findings from eye-tracking recordings	-Aptis test scores -Eye-recording data (Task 1 and Task 2) Overall 40 minutes The First 10 minutes The Last 10 minutes -Rating essay using the TEAP Writing Task B Rating Scales	-Test of Normality -Pearson Correlation Test -Wilcoxon Test -Mann-Whitney U Test -Multiple Regression Test -Pearson correlation Test
Chapter 5 Findings from cognitive-processing questionnaires	-5-point Likert answers from the questionnaires	-Chi-square Tests
Chapter 6 Findings from some cases in different cognitive stages of integrated writing processes	-Written outputs (Selective cases are shown)	Preparing-to-write -gaze-plots (first 5 minutes) -AOI Switches -Saccades/Regression/Return Sweeps Translating and Writing -writing outputs Revising and Monitoring -gaze-plots (last 5 minutes)
Chapter 7 Findings from the focus group discussions	-Reports from the focus group discussions	-Cross-tables by CEFR levels

Chapter 4. Quantitative Findings from Eye-movements

4.1 Introduction

This multi-phase study provided an enormous amount of data for data analysis. This chapter begins by describing the participants' language proficiency in reading and writing as shown in their Aptis scores. It is important to understand participants' independent abilities in reading and writing since their test scores will be used as predictors to find any association with eye-movement variables in order then to figure out what affects integrated writing processes and thus finally to understand the role of reading in integrated writing, as well as the effects of source texts and information from graphs. Next, findings of the Wilcoxon tests were reported to show the difference between how participants viewed AOIs between Task 1 and Task 2. Then, associations of eye-movements with reading and writing proficiencies were reported from the analysis of the full 40-minutes of eye-recordings followed by the analysis of the first 10 minutes and last 10 minutes of the recorded data. Finally, the associations between the eye-movements and the essays marked using the rating scales for the TEAP writing tests were explored. The findings provide insights into which of the criteria used for measuring integrated writing skills are strongly associated with eye-movements.

4.2 Background of the Participants (Gender, Age)

A total of 54 high school students from four Japanese high schools in urban areas in the western part of Japan took part in the Aptis tests between December 2015 and January 2016. Of those 54 students, 42 (10 males and 32 females) decided to participate in the eye-recorded TEAP writing tests. The students were 21 freshmen (Grade 10), 9 juniors (Grade 11) and 12 seniors (Grade 12). The English teachers from the four high schools helped recruit participants whom they believed would be available to take part according to their school time schedules. The participants came from both public (14 students) and private (28 students) schools in Kyoto,

Osaka and Kobe prefectures, 32 of whom were female students. Their ages ranged from 15 to 18 years old, suggesting they had at least three years of English language education at school plus one, two or three years in high school depending on their grade at that time.

4.3 Participants' language proficiency (Aptis test results)

4.3.1 Correlations between reading and writing skills

The Aptis reading and writing scores were used to understand what role reading and writing proficiency play in the integrated writing tasks. The bar charts below show the number of participants by their CEFR-based reading and writing levels (Figure 13).

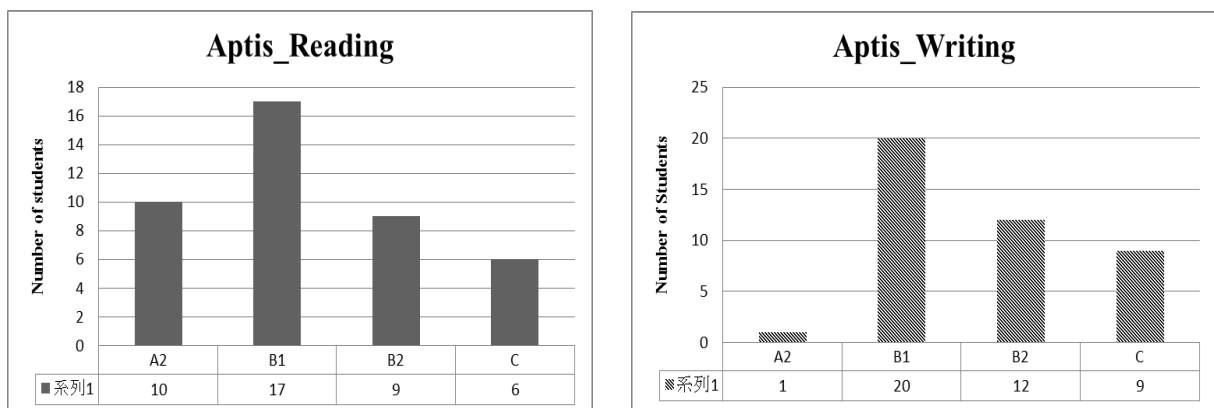


Figure 13. Distribution of CEFR levels for Aptis Reading and Writing Scores

From the Descriptive Statistics in Table 11, it can be seen that the students performed slightly better in the writing test ($M=37.67$, $SD=9.401$) than the reading test ($M=32.35$, $SD=8.66$). It is important to note there was only one participant who scored at A2 level in writing. The Pearson correlation analysis (Table 12) was performed using the scores of the reading, writing and grammar & vocabulary tests. As a result, moderate correlations were found between reading and writing ($r=.561$, $p<.001$), and between reading and grammar & vocabulary ($r=.685$,

p<.001). A large correlation was found between writing and grammar & vocabulary (r=.785, p<.000).

Table 11. Descriptive Statistics of the Aptis Tests

Aptis Scores (N=40)	Mean	Std. Deviation
Reading score	32.35	8.66
Writing score	37.68	9.40
Grammar and Vocabulary score	32.70	5.28

Note: The analysis only includes those students who decided to participate in the eye-tracking studies.

Table 12 Pearson Correlation Analysis for Aptis Reading, Writing and Grammar & Vocabulary

Aptis Scores (N=40)	Reading	Writing	Grammar & Vocabulary
Reading	1	.561**	.685**
Writing	.561**	1	.785**
Grammar & Vocabulary	.685**	.785**	1

** . Correlation is significant at the 0.01 level (2-tailed).

Table 13 summarizes the combination of study participants' reading and writing abilities (by CEFR-levels). The correlations between the reading and writing scores found in the Aptis tests gave opportunities to investigate the independent association of each with the eye-movement variables. For example, regression analysis could help identify whether it is reading or writing skills that are most predictive of a particular eye-movement variable. In chapter 6, some cases are shown to offer interpretations of the eye-tracking data for qualitative analysis. This chapter seeks to use the quantitative data to gain in depth insights into participants' eye-movement patterns.

Table 13. Participants' Reading and Writing Proficiencies (N=40)

Reading proficiency level		Writing proficiency level				Total
		A2	B1	B2	C	
A2	Task 1	1	6	1		8
	Task 2	1	6	1		8
B1	Task 1		10	2	4	16
	Task 2		11	2	5	18
B2	Task 1		2	5	2	9
	Task 2		2	5	2	9
C	Task 1			3	2	5
	Task 2			3	2	5
Total	Task 1	1	18	11	8	38
	Task 2	1	19	11	9	40

Light shaded cells indicate cases where writing proficiency was higher than reading proficiency; darker shaded cells indicate cases where reading proficiency was higher than writing proficiency

4.4 Eye-tracking movements for analysis

Forty-two students took the two tasks for the eye-tracking experiments. On the day of the eye-recording some of the participants were not wearing the eyeglasses that they normally wear and this resulted in a small “gaze percentage”, which could distort the data analysis (even though calibrations were attempted multiple times). The average gaze percentage was calculated based on the proportion of time the participants spent looking at the screen. This was 40.43 and 45.46 percent for Task 1 and Task 2, respectively. There was also one occasion when a participant spent most of the recording time looking down at the keyboard, resulting in a small gaze percentage. After a personal discussion with a member of the Tobii Technical staff based in Tokyo, it was decided to exclude from the subsequent analysis participants who were recorded as spending less than 5% of the time looking at the screen. According to their advice, any data less than 10% should not be counted. There were also a few rare cases where

a participant suddenly moved the screen monitor (without the researcher's permission) by hand to change the angle. Consequently, there were 38 valid recordings for Task 1 and 40 valid recordings for Task 2 out of 42 recordings in each test. The small sample raises some concerns about the validity of eye-recordings, and some future studies should look into introducing some training sessions to familiarize participants with the test format on a computer screen before collecting the eye-tracking data. This may be an important step to ensure the conditions of the two tests in terms of test takers' familiarity of the tasks. Most of the low gaze percentages, however, were a reflection of the multiple tasks this test involved. The participants had to look at different parts of the screens (two texts, two graphs) and the area of essay composition while monitoring their hands on the keyboard to produce their essay. In many cases, the gaze percentage was reduced by participants looking down at the keyboard.

Tobii Studio allows the export of different eye-metrics (e.g., Time to First Fixation, Fixation Duration, Total Fixation Duration, Fixation Counts, Visit Duration, Total Visit Duration., etc.) according to different segments of recordings and by categories such as different reading and writing language proficiency levels. The descriptive statistics of all the metrics used for the current study are summarized in Table 14 below.

Table 14. Descriptive Statistics of Eye-tracking Metrics (N=40)

Eye-tracking metrics	Task 1				Task 2			
	Minimum	Maximum	Mean	Std. Deviation	Minimum	Maximum	Mean	Std. Deviation
Time to First Fixation Essay	0.00	958.88	46.42	169.53	0.00	1091.43	177.77	219.41
Time to First Fixation Pie Chart Title	0.00	1314.76	88.60	228.90	0.00	2290.91	270.08	548.38
Time to First Fixation Line/Bar Graph Title	0.00	1439.32	134.52	308.32	.13	2248.51	150.29	431.18
Time to First Fixation Letter-to-the-Editor	0.00	1854.44	297.36	462.42	0.00	2228.18	474.06	591.74
Time to First Fixation Line Graph	0.00	556.62	65.58	145.41	0.00	485.08	27.59	84.03
Time to First Fixation Newspaper Article	0.00	1859.79	162.44	377.66	0.00	1951.48	228.55	391.22
Time to First Fixation Paragraph 1	0.00	667.95	47.47	117.83	0.00	437.96	45.67	79.41
Time to First Fixation Paragraph 2	0.00	1948.58	219.58	394.27	2.16	1437.14	243.18	299.52
Time to First Fixation Paragraph 3	0.00	2224.66	225.33	399.76	0.00	1476.63	239.73	305.55
Time to First Fixation Paragraph 4	0.00	1531.73	335.55	387.95	.33	1425.26	334.30	324.09
Time to First Fixation Pie Chart	0.00	314.43	18.12	55.54	0.00	567.51	68.02	137.42
Time to First Fixation Sarah Case	0.00	2398.23	432.19	618.30	0.00	2316.41	593.08	735.42
Time to First Fixation Task Instruction	0.00	43.74	5.45	7.88	0.00	1440.32	39.31	227.24
First Fixation Duration Essay	0.00	.77	.19	.15	.06	.46	.15	.09
First Fixation Duration Pie Chart Title	0.00	.56	.16	.11	.02	1.08	.20	.18
First Fixation Duration Line/Bar Graph Title	0.00	.65	.17	.13	.07	.56	.21	.11
First Fixation Duration Letter-to-the-Editor	0.00	.44	.13	.11	0.00	.58	.14	.13
First Fixation Duration Line Graph	0.00	.65	.17	.12	.05	.95	.20	.14
First Fixation Duration Newspaper Article	0.00	.36	.09	.10	0.00	.55	.14	.15

First Fixation Duration Paragraph 1	0.00	.39	.16	.09	.05	.31	.16	.06
First Fixation Duration Paragraph 2	0.00	.64	.18	.12	.06	.66	.22	.15
First Fixation Duration Paragraph 3	0.00	.31	.12	.07	0.00	.42	.15	.08
First Fixation Duration Paragraph 4	0.00	.93	.17	.16	.07	.43	.18	.09
First Fixation Duration Pie Chart	.07	.34	.16	.06	.02	1.08	.20	.18
First Fixation Duration Sara Case	0.00	.58	.15	.15	0.00	.89	.19	.21
First Fixation Duration Task Instruction	0.00	.83	.20	.15	.02	.35	.15	.07
Fixation Duration (Mean) Essay	0.00	.42	.25	.07	.14	.42	.27	.04
Fixation Duration (Mean) Pie Chart Title	0.00	.33	.19	.07	.08	.53	.22	.08
Fixation Duration (Mean) Line/Bar Graph Title	0.00	.34	.19	.07	.10	.38	.22	.05
Fixation Duration (Mean) Letter-to-the-Editor	0.00	.36	.13	.10	0.00	.35	.16	.10
Fixation Duration (Mean) Line Graph	0.00	.31	.19	.06	.15	.38	.21	.03
Fixation Duration (Mean) Newspaper Article	0.00	.32	.09	.09	0.00	.54	.15	.13
Fixation Duration (Mean) Paragraph1	0.00	.28	.19	.05	.13	.37	.24	.04
Fixation Duration (Mean) Paragraph 2	0.00	.29	.18	.06	.11	.32	.21	.04
Fixation Duration (Mean) Paragraph 3	0.00	.31	.19	.07	0.00	.34	.22	.06
Fixation Duration (Mean) Paragraph 4	0.00	.38	.21	.07	.13	.40	.23	.04
Fixation Duration (Mean) Pie Chart	0.00	.31	.20	.05	.11	.34	.21	.04
Fixation Duration (Mean) Sara Case	0.00	.38	.16	.13	0.00	.66	.18	.15
Fixation Duration Task Instruction	0.00	.29	.19	.05	.11	.29	.20	.04
Total Fixation Duration Essay	0.00	880.14	312.04	248.78	2.10	935.44	325.83	240.57
Total Fixation Duration Pie Chart Title	0.00	32.11	6.98	7.77	.08	27.80	6.84	6.48
Total Fixation Duration Line/Bar Graph Title	0.00	46.00	10.42	11.51	.28	21.88	6.33	5.67

Total Fixation Duration Letter-to-the-Editor	0.00	10.11	1.99	2.46	0.00	7.52	1.41	1.82
Total Fixation Duration Line Graph	0.00	95.59	28.27	22.98	3.07	98.42	26.87	20.07
Total Fixation Duration Paragraph 1	0.00	327.10	89.82	76.87	0.00	5.57	1.12	1.43
Total Fixation Duration Paragraph 2	0.00	181.03	59.10	52.65	1.38	464.32	170.79	117.82
Total Fixation Duration Paragraph 3	0.00	199.22	67.10	59.16	.43	238.99	63.84	49.19
Total Fixation Duration Paragraph 4	0.00	219.28	47.95	47.57	0.00	136.92	43.90	35.33
Total Fixation Duration Pie Chart	0.00	130.69	35.76	28.27	.25	259.92	91.52	73.43
Total Fixation Duration Sara Case	0.00	4.94	1.16	1.41	1.18	95.86	34.51	23.31
Total Fixation Duration Task Instruction	0.00	155.40	34.86	36.11	0.00	7.28	1.44	1.82
Fixation Count Essay	0.00	2566.00	1127.32	789.73	.34	67.68	21.08	17.81
Fixation Count Pie Chart Title	0.00	122.00	31.92	34.28	15.00	2914.00	1142.95	793.23
Fixation Count Line/Bar Graph Title	0.00	200.00	47.90	50.69	1.00	148.00	30.62	30.05
Fixation Count Letter-to-the-Editor	0.00	38.00	9.67	10.43	1.00	86.00	26.15	20.41
Fixation Count Line Graph	0.00	390.00	133.40	97.23	0.00	33.00	6.47	7.41
Fixation Count Newspaper Article	0.00	10.00	2.33	2.85	8.00	404.00	122.65	88.44
Fixation Count Paragraph 1	0.00	1611.00	404.02	348.57	0.00	18.00	4.90	5.76
Fixation Count Paragraph 2	0.00	758.00	269.90	215.87	11.00	1605.00	664.90	423.73
Fixation Count Paragraph 3	0.00	882.00	287.32	244.21	4.00	751.00	273.12	183.69
Fixation Count Paragraph 4	2.00	1105.00	377.30	300.22	0.00	531.00	182.57	137.19
Fixation Count Pie Chart	0.00	684.00	163.37	129.50	2.00	1105.00	377.30	300.22
Fixation Count Sara Case	0.00	15.00	4.15	4.46	11.00	403.00	155.12	96.12
Fixation Count Task Instruction	0.00	672.00	155.85	154.65	0.00	17.00	4.82	5.04
Visit Duration (Mean) Essay	0.00	17.00	4.97	3.46	3.00	308.00	94.05	74.15

Visit Duration (Mean) Pie Chart Title	0.00	.82	.36	.18	2.26	1800.51	744.16	442.81863
Visit Duration (Mean) Line/Bar Graph Title	0.00	1.15	.47	.28	.08	1.12	.48	.22
Visit Duration (Mean) Letter-to-the-Editor	0.00	1.23	.32	.30	.10	1.01	.45	.21
Visit Duration (Mean) Line Graph	0.00	6.62	1.19	1.18	0.00	1.00	.31	.27
Visit Duration (Mean) Newspaper Article	0.00	.72	.13	.15	.28	6.73	1.13	1.14
Visit Duration (Mean) Paragraph 1	0.00	6.85	2.21	1.47	0.00	.78	.23	.20
Visit Duration (Mean) Paragraph 2	0.00	12.25	2.12	2.18	.21	9.11	2.79	1.55
Visit Duration (Mean) Paragraph 3	0.00	4.10	1.63	.99	.11	4.37	1.85	.95
Visit Duration (Mean) Paragraph 4	0.00	17.21	2.53	2.76	0.00	7.84	2.05	1.44
Visit Duration (Mean) Pie Chart	0.00	7.32	1.36	1.40	.13	5.97	2.41	1.19
Visit Duration (Mean) Sara Case	0.00	1.80	.3351	.41	.13	3.09	1.05	.48
Visit Duration (Mean)Task Instruction	0.00	54.53	2.35	8.48	.11	5.02	1.30	1.06
Total Visit Duration Essay	0.00	1656.54	747.56	471.00	2.26	1800.51	744.16	442.81
Total Visit Duration Pie Chart Title	0.00	33.88	7.78	8.55	.08	34.84	8.30	8.26
Total Visit Duration Line/Bar Graph Title	0.00	52.33	12.25	13.37	.28	23.33	6.96	6.19
Total Visit Duration Letter-to-the-Editor	0.00	10.46	2.27	2.84	0.00	8.10	1.61	2.04
Total Visit Duration Line Graph	0.00	118.48	48.27	31.90	3.08	343.21	52.78	72.47
Total Visit Duration Newspaper Article	0.00	2.89	.42	.65	0.00	5.74	1.17	1.49
Total Visit Duration Paragraph 1	0.00	480.59	136.68	99.74	1.88	555.02	241.27	143.10
Total Visit Duration Paragraph 2	0.00	277.33	95.23	70.46	.43	319.20	94.90	71.37
Total Visit Duration Paragraph 3	0.00	287.98	96.37	79.67	0.00	166.07	56.16	42.49
Total Visit Duration Paragraph 4	0.00	280.58	75.30	67.05	.25	399.69	133.63	107.73
Total Visit Duration Pie Chart	0.00	217.59	65.73	46.57	1.19	137.65	56.96	36.30

Total Visit Duration Sara Case	0.00	5.21	1.28	1.59	0.00	7.38	1.72	2.17
Total Visit Duration Task Instruction	0.00	205.71	48.55	50.77	.34	100.65	27.31	22.92
Visit Count Essay	0.00	448.00	158.12	103.01	9.00	440.00	180.80	91.75
Visit Count Pie Chart Title	0.00	84.00	18.02	19.58	1.00	55.00	16.57	14.56
Visit Count Line/Bar Graph Title	0.00	105.00	21.15	21.09	1.00	42.00	14.00	9.97
Visit Count Letter-to-the-Editor	0.00	21.00	4.65	4.79	0.00	13.00	3.72	3.35
Visit Count Line Graph	0.00	165.00	49.07	35.89	7.00	153.00	45.32	33.61
Visit Count Newspaper Article	0.00	9.00	1.85	2.25	0.00	12.00	3.25	3.57
Visit Count Paragraph 1	0.00	260.00	68.67	55.17	9.00	234.00	89.62	54.37
Visit Count Paragraph 2	0.00	148.00	48.02	33.47	4.00	141.00	46.32	28.44
Visit Count Paragraph 3	0.00	141.00	44.65	38.44	0.00	90.00	33.20	22.68
Visit Count Paragraph 4	0.00	107.00	30.00	27.30	2.00	178.00	53.37	44.68
Visit Count Pie Chart	0.00	248.00	59.57	46.50	9.00	189.00	54.40	37.45
Visit Count Sara Case	0.00	10.00	2.70	2.84	0.00	14.00	2.85	3.26
Visit Count Task Instruction	0.00	148.00	38.17	31.86	3.00	61.00	22.60	14.68
Valid N (listwise)=40								

milliseconds

To begin with, it is important to acknowledge that significant statistical results could be spurious due to the repeated testing. When the test is repeatedly applied, it would be expected to find at least one significant result by chance in every 20 tests (given a p-value $<.05$). There is therefore a chance that some of the significant results may not be accurate.

In general, the significant results in respect to Task 2 might be considered more reliable than those for Task 1. This is because, as revealed in the survey and the focus group discussion, most of the participants were not familiar with the format of the task in the TEAP test, even though sample tests were distributed to them in the days before the eye-tracking experiment. The results from the second round of the test may therefore be more reliable in this case, since students were more familiar with what was expected of them by the time they took Task 2. This may have caused an unintended bias in the findings which will be discussed later in the research limitation.

4.4.1 Wilcoxon signed-ranks tests in groups of AOIs

A non-parametric Wilcoxon signed rank test was conducted to find differences in the means for the same group of participants (N=38) whose eye-recordings had sufficient gaze percentages for both Task 1 and Task 2 during the entire duration of recordings. Non-parametric tests were used because most of the eye-tracking measures were not normally distributed. Also, given that the study had relatively small samples, the Wilcoxon signed rank test was more appropriate than a paired-sample t-test to find the differences of the two tests.

To gain some overall trends, the Wilcoxon tests were conducted by grouping the Total Visit Duration (Sum) of AOIs. The Wilcoxon tests helped to find the difference between Task 1 and Task 2 for the same AOIs. The groups of AOIs included the areas of Task Instruction, two types of visual inputs, two types of source texts (i.e., the Newspaper Article (Paragraph 1 and

Paragraph 2), and the Letter-to-the Editor (Paragraph 3 and Paragraph 4), and essay compositions. Task 1 had a pie chart and a line graph while Task 2 had a bar graph and pie chart. Both tests contained two source texts, such as one newspaper article and one letter to the editor. Table 15 shows the results of the Wilcoxon test. The list shows variables (i.e. Task Instruction and Newspaper Article) that had statistical significance between Task 1 and Task 2 ($p < .05$) in bold.

Table 15. Wilcoxon signed-ranks Test Results (Task 1 and Task 2 in broader areas (groups) of AOIs for 40 minutes)

Dependent variables	Z	p-value
Total Visit Duration of Essay Composition	-.430a	.667
Total Visit Duration of Task Instruction	-2.997a	.003
Total Visit Duration of Pie Chart	-1.190a	.234
Total Visit Duration of Line Graph or Bar Graph	-1.667a	.096
Total Visit Duration of Newspaper Article	-4.207b	.000
Total Visit Duration of Letter-to-the Editor	-.712b	.476

Note: a based on positive ranks, b based on negative ranks, Significant at $p < .05$

Task Instruction: Understanding the task requirement in the instructions plays a key role for the test-takers in deciding what to do during the test. Given the fact that the participants have taken a similar test twice, it was hypothesized that the participants would spend more time studying the task instruction in Task 1 than in Task 2. The Wilcoxon Signed-Ranks Test between 1 and 2 for the overall analysis of the two 40-minute recordings in Table 16 shows a significant difference between the two tests. The Wilcoxon Signed-Ranks Test indicated that the Total Visit Duration on Task Instruction was statistically higher in Task 1 than in Task 2 ($z = -2.997$, $p < .005$).

The analysis of the ranks shows that 27 out of 40 participants spent more time on Task Instruction, measured by the Total Visit Duration during Task 1 than Task 2, whereas 13

students spent more time on the Task Instruction measured by the Total Visit Duration during Task 2 than Task 1. The test result therefore indicates that more participants spent a longer time on the Task Instruction in Task 1 than in Task 2, as predicted. The main reason for this result could be that the test-takers were not presented the tests (Task 1 and Task 2) randomly. This was later acknowledged as one of the research limitations in the research design. Other possible explanations, however, could be that the test-takers had become familiarized with the task instruction after the first round of the test regardless of which test was presented first. Both tests were designed to be similar in format, and the test instructions in the Task Instruction were identical. Given that there is a time-limit for any language assessment tests, it would be advantageous for the test-takers to familiarize themselves with the required tasks in advance as a part of their test-taking strategy.

Source Texts: The Wilcoxon Signed-Ranks Test showed a difference between time spent on the source texts in Task 1 and Task 2, measured by the total visit duration. The total visit duration on the Newspaper Article was found to be statistically higher in Task 2 than in Task 1 ($z=-4.207, p<.001$). According to the ranks, 33 out of 40 participants spent more time reading the Newspaper Article (i.e., Paragraph 1 and Paragraph 2) in Task 2 than in Task 1 while only seven participants spent more time reading the Newspaper Article in Task 1 than in Task 2.

Significant differences in participants' behavioural activities were therefore found between Task 1 and Task 2. These may come from the repetition of the tasks: the test-takers were probably quick to notice that both Task 1 and Task 2 were similar in structure and were thus able in the second test to identify quickly the parts of the texts in which the main topics and solutions were located. For example, the Newspaper Article (Paragraph 1 and Paragraph 2) introduced the main issues and solutions. On the other hand, the Letter-to-the-Editor (Paragraph 3 and Paragraph 4) just restated the information and gave only one more solution to the problem. Based on these results, it appears that majority of students spent more time in

Task 2 describing the situation from the key information taken from the newspaper article and, possibly, the one or two solutions suggested in the article.

Effects of Graph Types

The descriptive statistics in Table 16 describe how the participants had spent time on gleaning information from graphs in Task 1 (Pie Chart and Line Graph) and Task 2 (Pie Chart and Bar Graph). The Wilcoxon Signed-Ranks Test was used to show any differences between the different types of information from graphs, measured by the total visit duration (Table 17).

Table 16. Descriptive Statistics of Time Spent on Gaining Information from Different Types of Graphs in Task 1 and Task 2

N=40

Task 2	Mean	Std. Deviation	Minimum	Maximum
Total Visit Duration Line Graph (Task 1)	48.27	31.96	.00	31.90
Total Visit Duration Pie Chart (Task 1)	65.74	46.58	.00	46.58
Total Visit Duration Bar Graph (Task 2)	52.79	72.47	3.08	72.47
Total Visit Duration Pie Chart (Task 2)	56.96	36.30	1.19	36.30

Table 17. Wilcoxon signed-rank Tests of Differences in Eye-movement Measurements for Different Graph Types (40 minutes)

Dependent variable	Z	p-value
Total Visit Duration Graph Types (Pie Chart vs. Line Graph) in Task 1	-1.486a	.137
Total Visit Duration Graph Types (Pie Chart vs. Bar Graph) in Task 2	-2.890a	.004
Total Visit Duration Graph Types (Pie Chart 1 in Task 1 vs. Pie Chart 2 in Task 2)	-1.190b	.234
Total Visit Duration Graph Types (Line Graph in Task 1 vs. Bar Graph in Task 2)	-1.768b	.077

a. Based on negative ranks, b. Based on positive ranks
 Note: Negative Ranks means Task 2 < Task 1, Significant at P < .05

The Wilcoxon test was conducted to find whether there are significant results in how participants view different graph types. There was no statistically significant difference found in Task 1 between the Line graph and the Pie Chart. In Task 2, however, significant differences were found in the Total Visit Duration according to the Graph Types. Based on the Wilcoxon signed rank test, the types of graphs did have some impact on the time spent by test-takers on those graphs. In other words, participants spent longer on the Pie than the other graph in both tasks, and this difference was stronger in Task 1 (where the alternative was a Line) than in Task 2 (where the alternative was a Bar).

4.4.2. Wilcoxon signed ranks tests in sub-sections of AOIs

This section reports Wilcoxon tests conducted on sub-sections of AOIs for the full recordings for Task 1 and Task 2. The AOIs included the title of each graph and the heading of each source text. Also, each source text was divided into paragraphs. The Newspaper Article consisted of two paragraphs (Paragraph 1 and Paragraph 2), and the Letter-to-the-Editor also contained two paragraphs (Paragraph 3 and Paragraph 4) and the signature (i.e., Sara Case) of a person who wrote the letter to the newspaper company. Analysis of these smaller AOIs in both tests gave insights into test-takers' processing of information in depth.

Wilcoxon signed-ranks tests for the overall 40 minutes of recording

Rate of Fixation Duration was calculated to investigate the relative amount of time viewing of the AOIs for the purpose of comparing different participants. It shows the proportions of Total Fixation Duration in an AOI over the sum of Total Fixation Duration of all AOIs. The following table is the percentage (rate) of total fixation duration on an AOI. It was calculated by dividing the sum of the total fixation durations of all AOIs by the total fixation duration on that AOI. The Wilcoxon signed-ranks test for the 40 minutes of recordings for each test are summarized below in Table 18.

Table 18. Rates of total fixation duration on sub-sections of AOIs across Task 1 and Task 2 (40 minutes)

Dependent variable	Z	p-value
Rate of Total Fixation Duration for Essay	-1.290a	.197
Rate of Total Fixation Duration of Title (Pie Chart)	-1.290a	.809
Rate of Total Fixation Duration of Title (Line/Bar Graph)	-2.715a	.007
Rate of Total Fixation Duration of Letter-to-the-Editor	-1.697a	.090
Rate of Total Fixation Duration of Bar Graph	-1.438a	.150
Rate of Total Fixation Duration of Newspaper Article	-2.470b	.013
Rate of Total Fixation Duration of Paragraph 1	-4.288b	.000
Rate of Total Fixation Duration of Paragraph 2	-.605a	.545
Rate of Total Fixation Duration of Paragraph 3	-3.447a	.001
Rate of Total Fixation Duration of Paragraph 4	-3.575b	.000
Rate of Total Fixation Duration of Pie Chart	-1.653a	.098
Rate of Total Fixation Duration of Signature (Sarah Case)	-.991a	.322
Rate of Total Fixation Duration of Task Instruction	-5.471a	.000

Note: a based on positive ranks; b based on negative ranks

Note: Negative Ranks means Task 2 < Task 1 Positive Ranks means Task 2 > Task 1

Significant at p<.05

Table 19. Ranking of the Means for Wilcoxon signed-ranks Tests

Rate of Fixation Duration		N	Mean Rank
Graph Title	Negative Ranks	28	21.86
Line/Bar Graph	Positive Ranks	12	17.33
Heading	Negative Ranks	23	19.17
(Letter-to-the-Editor)	Positive Ranks	13	17.31
Heading	Negative Ranks	10	15.30
(Newspaper Article)	Positive Ranks	24	18.42
Paragraph 1	Negative Ranks	7	13.00
	Positive Ranks	33	22.09
Paragraph 3	Negative Ranks	28	22.75
	Positive Ranks	11	13.00
Paragraph 4	Negative Ranks	9	16.00
	Positive Ranks	31	21.81
Task Instruction	Negative Ranks	38	21.50
	Positive Ranks	2	1.50

Note: Negative Ranks means Task 2 < Task 1 Positive Ranks means Task 2 > Task 1

Notable differences between Task 1 and Task 2 were found in Paragraph 1, Paragraph 3 and Paragraph 4. From the Table 19, it can be said that the students viewed Paragraph 3 more in Task 1 than in Task 2, but they viewed Paragraphs 1 and 4 more in Task 2 than they did in Task 1. Paragraph 1 represented the first half of the text of the Newspaper Article, which described the key issue and the first of the two proposed solutions. Paragraph 4 contained information for the alternative solution suggested by Sara Case who wrote the letter to the editor. These findings help identify which part of the texts participants looked at for synthesizing information from the text. Such information helped generate questions for the focus group discussions. The Wilcoxon signed-ranks tests between Task 1 and Task 2 of all other eye-movement metrics are reported in Appendix L, M, and N.

4.4.3 Mann-Whitney U tests for groups of AOIs

Effects of Language Proficiencies: A non-parametric Mann-Whitney U test was run to study the differences in the visit durations between the B1 and below group (lower intermediate) and

the B2 and above group (upper intermediate) as shown in Table 20 below for reading and writing. The rationale behind dividing the groups between CEFR level B1 and B2 for both analyses was because approximately half of the participants fell in either of the categories for writing. Because two participants had their eye-recording data removed on account of their low gaze weights in Task 1, there were 38 test scores available for Task 1 and 40 test scores available for Task 2 for further analysis. As a result, 24 students were classified as B1 and below and 14 students as B2 and above for reading in Task 1. Also, an exactly equal number of students (n=19) was classified as B1 and below and B2 and above for writing. In Task 2, 26 reading scores fell at B1 or below, and 14 reading scores fell at B2 or above, while an equal number of writing scores (n=20) fell in to these two groups. The statistically significant variables are summarized in the tables below according to the total test duration of 40 minutes, the first 10 minutes and the last 10 minutes in the broader areas of interests (groups of AOIs).

The Mann-Whitney U test for differences between the two groups in respect to fixation on the AOI groups revealed no statistically significant differences in the distributions of the eye-tracking variables between low and high proficiency participants, either in terms of reading or writing, either for the full duration of the tests or for the last 10 minutes of the recording. There were, however, some statistically significant differences during the first 10 minutes of recording between lower and upper intermediate groups in Task 2 (Table 22). The Total Visit Duration to the Essay Composition AOI was statistically higher ($U=125.00$, $p<.05$) for the B2 and above group (Mdn=88.15) than for the B1 and below group (Mdn=13.98) in writing. The same trend was found in respect to the Total Visit Duration in Essay Composition ($U=109.00$, $p<.05$), which was statistically higher for the B2 and above group (Mdn=88.15) than for the B1 and below group (Mdn=13.98) in reading. Also, the Total Visit Duration to the Bar Graph was found to be statistically higher ($U=123.00$, $p<.05$) for the B2 and above group (Mdn=26.56) than for the B1 and below group (Mdn=17.59). It is notable that while the same

variables were tested in Task 1 and Task 2, the significance was found only in Task 2 during the first 10 minutes of the test.

Table 20. Significant Results of the Mann Whitney U Test to Compare the Groups of Upper and Lower Groups of Aptis Scores in Groups of AOIs

Total Visit Duration (Task 2)		N	Rank Average	Sum of Ranks	Median	U	Z	P
Bar Graph (Writing)	Lower Intermediate	20	16.65	333.00	17.5850	123.00	- 2.083	.037
	Upper Intermediate	20	24.35	487.00	28.5600			
Essay Composition (Writing)	Lower Intermediate	20	16.75	335.00	13.9800	125.000	-2.029	.042
	Upper Intermediate	20	24.25	485.00	88.1450			
Essay Composition (Reading)	Lower Intermediate	26	17.69	460.00	13.98	109.00	-2.070	.038
	Upper Intermediate	14	25.71	360.00	88.15			

The difference is significant when $p < .05$

4.4.4 Mann-Whitney U tests for sub-sections of AOIs

For more accurate views, the Mann-Whitney U test was conducted for sub-sections of AOIs between the two groups (B2-above and B1-below groups according to both Aptis reading and writing scores). In the table below, the results on the Rate of Total Fixation Duration are reported for the total test duration of 40 minutes, the first 10 minutes and the last 10 minutes of recordings. Among the AOIs, the most significant difference was once again found in the Essay Composition for Task 2 between the groups of students who had higher and lower reading scores. The mean rank was higher for the group with a reading ability of B2 and above group ($Mdn=.456$) than for the B1 and below group ($Mdn=.294$) ($U=95.00$, $p < .05$). This result showed that the participants who had a reading proficiency at B2 and above spent more time viewing the essay composition area in Task 2. This could also possibly mean that these students were better at blind typing. When students were better at blind typing, they did not have to look down at the keyboard for typing. As a result, they would have spent longer on the screen.

Indeed, the questionnaire results showed that the participants' typing skills were significantly different between participants with different proficiency levels. Another significant statistical difference was found in the signature, Sara Case in Task 1, which appears at the end of Paragraph 4 in the Letter-to-the-Editor. The mean rank for the group with a reading ability of B2 and above (Mdn=.0025) was greater than for the B1 and below group (Mdn=.0003) (U=87.000, p<.05). The result appears to show that the test-takers who had higher writing proficiency had a longer total visit duration on the "Sara Case," in the last part of the source text. This also suggests that these students who had the ability to identify who wrote the Letter-to-the-Editor had a better chance of getting higher writing score.

Table 21. Significant Results of the Mann Whitney U Test to Compare the Groups of Upper and Lower Groups of Aptis Scores in Sub-division of AOIs (all 40 minutes)

Rate of Fixation Duration All 40 minutes		N	Rank Average	Sum of Ranks	Median	U	Z	p*
Essay Composition (Reading) Task 2	Lower Intermediate	26	17.15	446.00	.2943	95.000	-2.467	.014
	Upper Intermediate	14	26.71	374.00	.4595			
Sara Case (Writing) Task 1	Lower Intermediate	26	16.85	438.00	.0003	87.000	-2.753	.006
	Upper Intermediate	14	27.29	382.00	.0025			

*The difference is significant when p<.05

Furthermore, there were two statistically significant differences found during the First 10 minutes of Task 1 between the B2 and above and B1 and below test-takers (Table 22). In respect to Task Instruction, the mean rank for the group with a writing ability of B1 and below (Mdn=1.413, Mean Rank=24.48) was greater than for the B2 and above group (Mdn=.089, Mean Rank=16.53) (U=125.000, p<.05). The result indicates that the group with a lower writing proficiency viewed the Task Instruction for longer in the first 10 minutes as measured by the Rate of Total Fixation Duration

The second difference was in respect to the title of the line graph, for which the total fixation duration was statistically longer for the group with a reading ability of B2 and above (Mdn=.030) than for the B1 and below group (Mdn=.011) (U=106.00, $p<.031$). The title of the Line Graph, and the Average Hours of Sleep among High-School Students in Greenhill may have been viewed for longer by those who had a higher reading proficiency because the title contained some keywords and ideas that could be used for their essay.

Table 22. Significant Results of the Mann Whitney U Test to Compare the Groups of Upper and Lower Groups of Aptis Scores in Sub-division of AOIs (First 10 minutes)

Rate of Fixation Duration First 10 minutes		N	Rank Average	Sum of Ranks	Median	U	Z	p*
Task Instruction (Writing) Task 1	Lower Intermediate	20	24.48	489.50	.1413	120.500	-2.151	.032
	Upper Intermediate	20	16.53	330.50	.0892			
Essay Composition (Writing) Task 2	Lower Intermediate	20	15.90	318.00	.0312	108.000	-2.489	.013
	Upper Intermediate	20	25.10	502.00	.1832			
Title of Line Graph (Reading) Task 1	Lower Intermediate	26	17.58	457.00	.0107	106.000	-2.161	.031
	Upper Intermediate	14	25.93	363.00	.0300			

*The difference is significant when $p<.05$

The complete analysis of the Mann-Whitney U tests for the full test duration of 40 minutes, the first 10 minutes and the last 10 minutes are included in Appendices O, P and Q.

Table 23. Eye-tracking measurements from the groups of AOIs: Mann-Whitney tests of differences between participants with lower and upper intermediate skills in reading and writing (Overall 40 minutes, the First 10 Minutes and the Last 10 minutes)

Groups of AOIs Mann-Whitney tests of difference (40 min.) Dependent variable (Sum of Total Visit Duration)	Task 1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Total Visit Duration of Essay Composition	-1.219	.223	-1.115	.265	-1.730	.084	-1.039	.299
Total Visit Duration of Task Instruction	-.482	.630	.208	.208	-.227	.821	-.622	.534
Total Visit Duration of Pie Chart	-.482	.630	-1.633	.103	-.057	.955	-1.011	.312
Total Visit Duration of Line Graph (or Bar Graph)	-1.106	.269	-.377	.706	-.227	.821	-.224	.823
Total Visit Duration of Newspaper Article	-1.049	.294	-.462	.644	-.284	.777	-.471	.638
Total Visit Duration of Letter-to-the-Editor	-.794	.427	-.244	.808	-.199	.843	-1.409	.159

*The difference is significant when $p < .05$

Groups of AOIs Mann-Whitney tests of difference (First 10 min.) Dependent variable	Task 1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Total Visit Duration of Essay Composition	-.993	.321	-1.353	.176	-2.070	.038	-2.029	.042
Total Visit Duration of Task Instruction	-.766	.444	-1.583	.114	-.085	.932	-.460	.646
Total Visit Duration Pie Chart	-1.304	.192	-.122	.903	-1.319	.187	-1.366	.291
Total Visit Duration of Line Graph (or Bar Graph)	-1.319	.192	-1.055	.291	-1.815	.070	-2.083	.037
Total Visit Duration of Newspaper Article	-1.219	.223	-.230	.818	-.255	.799	-.271	.787
Total Visit Duration of Letter-to-the-Editor	-.199	.843	-.609	.543	-1.730	.084	-.839	.402

*The difference is significant when $p < .05$

Groups of AOIs Mann-Whitney tests of difference (Last 10 min.) Dependent variable	Task 1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Total Visit Duration of Essay composition	-.936	.349	-.284	.776	-1.900	.057	-1.515	.130
Total Visit Duration of Task Instruction	-.072	.943	-.192	.848	-.808	.419	-.564	.573
Total Visit Duration of Pie Chart	-.028	.977	-.406	.685	-.766	.444	-.893	.372
Total Visit Duration of Line Graph (or Bar Graph)	-1.915	.055	-.974	.330	-.014	.989	-.244	.807
Total Visit Duration of Newspaper Article	-.908	.364	-.676	.499	-.170	.865	-.433	.665
Total Visit Duration of Letter-to-the-Editor	-.868	.385	-1.371	.170	-.525	.600	-.027	.978

*The difference is significant when $p < .05$

Table 24. Eye-tracking measurements from the sub-division of AOIs on Rate of Total Fixation Durations: Mann-Whitney tests of differences between participants with lower and upper intermediate skills in reading and writing (Overall 40 minutes, the First 10 Minutes and the Last 10 minutes)

Rate of Total Fixation Duration Mann-Whitney tests of difference (40 min.) Dependent variable	Task1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Rate of Total Fixation Duration of Essay Composition	-.879	.379	-.446	.655	-2.467	.014	-1.839	.066
Rate of Total Fixation Duration of Title (Pie Chart)	-.822	.411	-.189	.850	-.369	.712	0.000	1.000
Rate of Total Fixation Duration of Title (Line Graph)	-.142	.887	-.054	.957	-1.560	.119	-1.190	.234
Rate of Total Fixation Duration of Letter-to-the-Editor	-.829	.407	-.368	.713	-1.080	.280	-.922	.356
Rate of Total Fixation Duration of Line Graph (or Bar Graph)	-.170	.865	-.311	.756	-.681	.496	-.298	.766

Rate of Total Fixation Duration of Newspaper Article	-.308	.758	-1.230	.219	-.372	.710	-.957	.339
Rate of Total Fixation Duration Paragraph 1	-1.560	.119	-.852	.394	-1.191	.234	-1.055	.291
Rate of Total Fixation Duration of Paragraph 2	-1.106	.269	-.717	.473	-1.900	.057	-1.380	.168
Rate of Total Fixation Duration of Paragraph 3	-.199	.843	-.406	.685	-.822	.411	-.920	.358
Rate of Total Fixation Duration of Paragraph 4	-.737	.461	-.676	.499	-1.475	.140	-1.461	.144
Rate of Total Fixation Duration of Pie Chart	-1.049	.294	-.771	.441	-.085	.932	-1.001	.317
Rate of Total Fixation Duration of Sarah Case	-2.753	.006	-1.769	.077	-.635	.526	-.716	.474

*The difference is significant when $p < .05$

Rate of Total Fixation Duration Mann-Whitney tests of difference (First 10 min.) Dependent variable	Task1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Rate of Total Fixation Duration of Essay Composition	-.312	.755	-.514	.607	-2.581	.010	-2.489	.013
Rate of Total Fixation Duration of Title (Pie Chart)	-.184	.854	-.771	.441	-1.079	.281	-.785	.432
Rate of Total Fixation Duration of Title (Line Graph)	-2.161	.031	-1.329	.184	-.284	.777	-1.082	.279
Rate of Total Fixation Duration of Letter-to-the-Editor	-.161	.872	-.853	.394	-.974	.330	-1.436	.151
Rate of Total Fixation Duration of Line/Bar Graph	-1.446	.148	-1.474	.140	-1.475	.140	-1.758	.079
Rate of Total Fixation Duration of Newspaper Article	-.515	.606	-.333	.740	-.464	.643	-1.216	.224
Rate of Total Fixation Duration of Paragraph 1	-1.390	.165	-.433	.665	-.284	.777	-.730	.465
Rate of Total Fixation Duration of Paragraph 2	-.625	.532	-.054	.957	-.908	.364	-.162	.871
Rate of Total Fixation Duration of Paragraph 3	-.085	.932	-.759	.448	-1.675	.094	-1.164	.244
Rate of Total Fixation Duration of Paragraph 4	-.800	.424	-.668	.504	-1.306	.192	-.921	.357

Rate of Total Fixation Duration of Pie Chart	-0.397	.691	-.149	.882	-.936	.349	-1.596	.110
Rate of Total Fixation Duration of Sarah Case	-.945	.345	-1.112	.266	-.599	.549	-.016	.987
Rate of Total Fixation Duration of Task Instruction	-1.106	.269	-2.151	.032	-.681	.496	-1.271	.204

*The difference is significant when $p < .05$

Rate of Total Fixation Duration Mann-Whitney tests of difference (Last 10 min.) Dependent variable	Task 1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Rate of Total Fixation Duration Essay Composition	-0.298	.766	-0.379	.705	-0.395	.061	-1.001	.317
Rate of Total Fixation Duration of Title (Pie Chart)	-.089	.929	-.199	.843	-.313	.567	-.574	.566
Rate of Total Fixation Duration of Title (Line Graph)	-1.187	.235	-.741	.459	-.128	.849	-.503	.615
Rate of Total Fixation Duration of Letter-to-the-Editor	-.485	.628	-.889	.374	-.128	.767	-.518	.605
Rate of Total Fixation Duration of Line/Bar Graph	-1.534	.125	-.772	.440	-.644	.809	-.840	.401
Rate of Total Fixation Duration of Newspaper Article	-1.113	.266	-2.859	.004	-.557	.578	-.286	.775
Rate of Total Fixation Duration of Paragraph 1	-.852	.394	-.908	.364	-.557	.843	-.243	.808
Rate of Total Fixation Duration of Paragraph 2	-1.193	.233	-.555	.579	-.085	.571	-1.380	.168
Rate of Total Fixation Duration of Paragraph 3	-1.100	.271	-1.213	.225	-.808	.741	-.466	.641
Rate of Total Fixation Duration of Paragraph 4	-.158	.874	-.315	.753	-.284	.532	-.190	.850
Rate of Total Fixation Duration of Pie Chart	-.170	.865	0.000	1.000	-.227	.202	-.081	.935
Rate of Total Fixation Duration of Sarah Case	-.332	.740	-.918	.359	-.383	.609	-.975	.330
Rate of Total Fixation Duration of Task Instruction	-.101	.920	-.411	.681	-.085	.686	-.193	.847

*The difference is significant when $p < .05$

4.4.5 Multiple regression analysis for groups of AOIs

Roles of Reading and Writing: The eye-recording data in groups of AOIs were analysed by using both the Aptis reading and writing scores as continuous predictors and eye-tracking measures as dependent variables. Multiple regression tests were conducted for the full test duration of 40 minutes (Table 25), the first 10 minutes (Table 26) and the last 10 minutes (Table 27). These tests were run to identify whether language levels in reading and writing can separately predict the eye-tracking measurement variables.

Table 25. Regression results for eye-movement variables for the full 40 minutes (groups of AOIs)

Groups of AOIs	Task 1					Task 2				
	Reading predictor		Writing predictor		R-squared	Reading predictor		Writing predictor		R-squared
	Std beta	p-value	Std beta	p-value		Std beta	p-value	Std beta	p-value	
Dependent variable Sum of Total Visit Duration										
Total Visit Duration Essay	.012	.950	.263	.177	.073	.156	.428	.072	.714	.042
Tot. Visit Duration Task Instruction	.298	.119	-.404	.037	.117	.049	.806	-.072	.718	.004
Tot. Visit Duration_Pie Chart	.353	.068	-.332	.086	.103	-.077	.683	.364	.060	.107
Tot. Visit Duration Line/Bar Graph	.010	.960	-.043	.829	.001	.150	.449	.005	.981	.023
Tot. Visit Duration Newspaper Article	.226	.252	-.057	.771	.040	-.238	.228	.195	.323	.043
Tot. Visit Duration Letter-to-the-Editor	.098	.621	.071	.720	.022	-.035	.861	.025	.901	.001

Note: Each row contains the results from the two separate regressions, with results first for Task 1 and then for Task 2. P-values less than .05 highlighted in bold. N is 39. *The difference is significant when $p < .05$

Table 26. Regression results for eye-movement variables in the First 10 minutes (groups of AOIs)

Groups of AOIs	Task 1					Task 2				
	Reading predictor		Writing predictor		R-squared	Reading predictor		Writing predictor		R-squared
	Std beta	p-value	Std beta	p-value		Std beta	p-value	Std beta	p-value	
Dependent variable										
Total Visit Duration Essay	.152	.416	.261	.166	.136	.252	.194	.042	.827	.077
Tot. Visit Duration Task Instruction	.262	.168	-.413	.033	.118	.091	.647	-.413	.736	.006
Tot. Visit Duration_ Pie Chart	.272	.164	-.293	.134	.070	.169	.392	-.293	.952	.031
Tot. Visit Duration Line/Bar Graph	.073	.712	.086	.665	.020	.341	.065	.086	.470	.183
Tot. Visit Duration Newspaper Article	.311	.113	-.134	.489	.068	-.178	.369	-.134	.378	.027
Tot. Visit Duration Letter-to-the-Editor	-.122	.537	.182	.359	.023	-.268	.175	.182	.286	.053

*The difference is significant when $p < .05$

Table 27. Regression results for eye-movement variables in the Last 10 minutes (groups of AOIs)

Groups of AOIs	Task 1					Task 2				
	Reading predictor		Writing predictor		R-squared	Reading predictor		Writing predictor		R-squared
	Std beta	p-value	Std beta	p-value		Std beta	p-value	Std beta	p-value	
Dependent variable										
Total Visit Duration Essay	.070	.721	.177	.365	.050	.118	.546	.144	.461	.054
Tot. Visit Duration Task Instruction	-.069	.729	.091	.647	.006	.044	.825	.041	.836	.006
Tot. Visit Duration_ Pie Chart	-.069	.729	.091	.647	.006	.117	.554	.029	.883	.018
Tot. Visit Duration Line/Bar Graph	-.256	.168	.486	.011	.162	-.221	.258	-.045	.816	.062
Tot. Visit Duration Newspaper Article	-.047	.813	.000	.999	.002	.013	.949	.082	.679	.008
Tot. Visit Duration_ Letter-to-the-Editor	.053	.788	.120	.546	.024	.124	.531	.041	.837	.023

*The difference is significant when $p < .05$

For instance, when regressed against the full 40 minutes of recording, the Aptis writing score was statistically significantly negatively related to the Total Visit Duration to the Task Instruction, ($\beta = -.40$, $p < .05$) in Task 1, but not in Task 2. Likewise, in the first 10 minutes of recording, the Aptis writing score also statistically significantly predicted the Total Visit Duration to Task Instruction, ($\beta = -.41$, $p < .05$) in Task 1. The last 10 minutes of recording found one variable that was statistically different in terms of Total Visit Duration. The writing score predicted a significant difference in the Total Visit Duration to the Line Graph in Task 1 ($\beta = -.26$, $p < .05$). No other statistically significant variable was found in the last 10 minutes of recordings in the broader areas (groups) of AOIs. It is important however to note that findings that were only significant in Task 1 may not be applicable in an applied setting.

4.5 Summary of Aptis Test and Eye-tracking Results

This chapter began with a description of the participants' language proficiencies in reading and writing. To gain a sense of the overall picture, the broader areas of AOIs (groups) were analysed followed by sub-section of AOIs. Wilcoxon signed-ranks tests were conducted to find the differences between Task 1 and Task 2, as well as between different types of source texts and graphs within the two sets of tasks and the three different time segments (40 minutes, the first 10 minutes and the last 10 minutes). In addition, Mann-Whitney U tests were used to test for significant differences between the upper and lower intermediate groups when using the Aptis reading and writing scores as predictors for the eye-tracking metrics in respect to groups of AOIs. Lastly, multiple regression tests were conducted using the Aptis reading and writing scores to see the associations between the eye-tracking metrics and language proficiencies. The results were reported in the hope of understanding the relationships between language proficiency and the effects of the source texts and information from the graphs.

- 1) Task Instruction:** The findings from the Wilcoxon tests (Table 15) suggested that the students viewed the Task Instructions for longer in Task 1 than in Task 2. This finding could be the result of an increased familiarity with the test format. This finding could be due to the fatigue effect (i.e., the tasks were not given in a randomized order). Some discrepancies were also found in Task 1 between those with writing proficiencies of B1 and below and B2 and above. The finding from the Mann-Whitney U Test suggests that students with a higher writing proficiency (B2 and above) spent less time than the B1 and below students fixating on the Task Instruction. It may be that the students with lower writing proficiency did not fully understand the task requirements for the test, thus it took longer for them to comprehend what they were supposed to write in the essay. An understanding of the task prompt affected whether participants could fulfil the writing tasks, which in turn influenced their writing scores.
- 2) Types of graphs:** The Wilcoxon tests found that, in both tests, the participants viewed the Pie Charts longer than other types of graphs, such as Line Graph and Bar Graph. The Wilcoxon tests also indicated that the test-takers had spent more time on the Pie Chart than on the Bar Graph in Task 2. Although there was a difference it was not statistically significant, less proficient students tended to view the Pie Charts for their main source of information more than they did the Line Graph or the Bar Graph (Table 17). The questionnaire will later reveal which type of graphs the test-takers thought they had mostly depended on as a source of information.
- 3) Heading of the source texts and graph titles:** There were two source texts used for both Task 1 and 2 (i.e. the Newspaper article and the Letter-to-the-Editor) that were nearly identical in terms of word lengths, format and the range of lexical

difficulties. One text was given in the form of a newspaper article for addressing the issues of concern, with some proposals for a solution. The second text was in the form of a Letter to the Editor, which reviewed the issue and the solution and suggested another approach to the problem. The topics for the two tests were different. Task 1 was about the lack of sleep among teenagers and Task 2 was about the reduction of food waste in the city.

The findings from the Wilcoxon tests (Table 18 and Table 19) indicated that the participants spent more time reading the newspaper article in Paragraph 1 and 2 than they did in the letter to the editor in Paragraph 3 and 4. Also, the finding of the Mann-Whitney U test found a statistically significant difference in terms of how upper and lower intermediate students viewed the second part of the source texts (as in Paragraph 3 and Paragraph 4), as well as the signature (i.e. Sara Case) placed at the end of the Letter-to-the-Editor. The result may indicate that reading proficiency affected how quickly the test-takers were able to read both source texts, and which source text they relied on in terms of the information they presented or included in their essay.

- (4) **Essay compositions:** Mann-Whitney U tests were conducted to find significant differences between the B1 and below and the B2 and above groups during the 40 minutes of recording. In Task 2, there was a significant difference between the two groups when the Total Fixation Durations on Essay composition were tested in relation to the participants' reading and writing scores. In both cases, the B2 and above groups spent significantly longer in the Essay Composition area.
- (5) **Roles of language proficiencies:** The multiple regression tests were used to test whether the Aptis reading and writing scores significantly predicted the eye-

tracking variables. The results of the regression indicated that the predictors explained the Total Visit Duration of the Task Instruction (beta=-.40, $p<.05$) for the overall 40 minutes of recording and (beta=-.41, $p<.05$) for the first 10 minutes of recording. The Mann-Whitney U test found a statistical difference in the time spent on the Task Instruction between the upper and lower writing groups in the first 10 minutes. Regression analysis made it possible to confirm the predictive role of writing skills for the length of time spent on Task Instruction by holding the level of reading skills constant, not only during the first 10 minutes of the test but also for all 40 minutes of the tests in Task 1.

Other than the obvious fact that test-takers spent longer on the Task Instruction in Task 1 than Task 2, there were few robust findings using the eye-tracking measurements to support the claim for strong associations between independent reading and writing skills as measured by eye-tracking experiments. In other words, the results indicated that reading and writing proficiency, at least as measured independently by the Aptis tests, was a poor predictor of the eye movements during the integrated writing tasks. The finding is consistent with the earlier studies (e.g., Flashive & Bailey (1993) and Shanhan & Tierney (1990)) that reading and writing abilities were only moderately correlated, but not with others (e.g., McGinley, (1992); Elley (1991) and Wagner & Stanovich (1996)).

It could be that the sample (N=38 for Task 1, N=40 for Task 2) was not large enough to show any significant trends. Close investigation of individual data, however, suggested that eye movements could provide many insights into what we can learn about the test-takers at different language proficiency levels. This will be explored further in the qualitative analysis.

4.6 TEAP Test and Eye-tracking Results

The essays produced during the eye-tracking experiments were rated based on the rating scales for the TEAP writing rubrics. This section reports on the inter-rater reliability by correlating the analytical scores of rater 1 and rater 2, followed by Pearson Correlation tests between the Aptis outcomes and the essay marks, as well as Pearson Correlation tests on the scores of the five criteria for the TEAP writing, in order to find the associations with eye-tracking metrics.

4.6.1 The Correlations between Rater 1 and Rater 2 ratings for the five criteria

Two raters marked the same essays independently (40 essays for Task 1 and 39 essays for Task 2 respectively). The reliability test was examined between the two raters. Descriptive statistics found that Rater 1 was consistently more generous than Rater 2 in all five categories.

Table 28. Descriptive statistics of two raters of the TEAP essays

Item Means	Mean	Minimum	Maximum	Range	Variance	N of Items
Task 1	3.78	3.27	4.30	1.03	.13	10
Task 2	3.86	3.41	4.36	.95	.07	10

Note: N=40 (Task 1), one excluded from Task 2 (N=39)

Table 29. Correlations between Rater 1 and Rater 2 ratings of the five criteria

TEAP Writing Scale Criteria	Cronbach's Alpha		Cronbach's Alpha Based on Standardized Items		N of Items
	Task 1	Task 2	Task 1	Task 2	
Main Idea	.804	.834	.811	.846	2
Coherence	.680	.819	.690	.821	2
Cohesion	.703	.669	.701	.670	2
Lexical Range & Accuracy	.622	.688	.631	.691	2
Grammatical Range & Accuracy	.305	.701	.311	.701	2

Valid N=40 (Task 1), N=39 (Task 2* one essay was not recorded by error)

Table 30. Inter-rater reliabilities between TEAP Task 1, Task 2 and the Average

Intraclass Correlation Coefficient							
	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.831 ^a	.736	.900	15.752	39	78	.000
Average Measures	.937 ^c	.893	.964	15.752	39	78	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Inter-Item Correlation Matrix			
	TEAP Task 1	TEAP Task 2	Average TEAP
TEAP Task 1	1.000	.696	.938
TEAP Task 2	.696	1.000	.902
Average TEAP	.938	.902	1.000

Based on the correlations between Rater 1's and Rater 2's ratings for the five criteria, the Cronbach's Alpha in Task 2 was higher for all five criteria except Cohesion. What is considered to be an acceptable level of agreement for the correlations to be significant differs depending on the sources. In this study, we referred to one source from Cicchetti's paper on the guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology (Cicchetti, 1994) which defined 00-.19 (very weak); .20-.39 (weak); .40-.59 (moderate); .60-.79 (strong); and .80-1.0 (very strong). Accordingly, anything below .60 (in our case Grammatical Range and Accuracy) needed to be re-examined. The biggest difference, (low correlation) in the criteria, was found in the Grammatical Range & Accuracy in Task 1 (alpha=.305). The discrepancy may have been caused by the extent to which the rater evaluated sentences that the test-taker had lifted from the source texts. When a test-taker had copied a long string of word or a sentence, there were few grammatical mistakes shown in the outputs. As a rule of thumb, the test-takers should not be given any credit for chunks of texts (i.e., at sentence level or above) that were lifted directly from the texts. In some cases, however,

additional credits can be assigned if the evidence shows that the test-takers used words and phrases that were linked by some discourse markers that they came up with on their own. In marking the Task 1 analytic scores, the two raters scored the Grammatical Range & Accuracy criterion quite differently from each other. To improve the rating validity, the two raters had an online discussion on the scale descriptor for Grammar Range & Accuracy. An agreement was made not to give credit for sentences that the test-taker had borrowed from the source texts when evaluating their grammatical range and accuracy. As a result, the Cronbach's Alpha for grammatical range and accuracy improved significantly from .311 in Task 1 to .701 in Task 2. As shown in Table 30, the inter-rater reliability test using the absolute agreement method showed the reliability score of the average TEAP scores as .93, which suggests high inter-rater reliabilities between the two raters.

4.6.2 Pearson Correlation test (Aptis scores and TEAP analytic scores)

A Pearson Correlation test was conducted using the TEAP analytic scores, based on the average of the two raters, and the Aptis reading and writing scores. The results show that the scores were only moderately correlated.

Table 31. Descriptive statistics of Aptis writing, reading, TEAP scores

	N	Mean	Maximum	Minimum	Std. Deviation
Average TEAP Scores	40	1.50	81.50	38.2125	18.12807
Aptis Writing Scores	40	10.00	50.00	37.6750	9.40100
Aptis Reading Scores	40	18.00	50.00	32.3500	8.66336

Table 32. Pearson Correlation Test on Aptis Reading, Writing and TEAP scores

Variables	1	2	3
1. Average TEAP Scores	-		
2. Aptis Reading Scores	.566	-	
3. Aptis Writing Scores	.660	.561	-

In addition, the Pearson Correlation tests with the five criteria making up the TEAP scores confirmed that the Aptis writing scores had a slightly higher correlation with the TEAP scores than the Aptis reading scores across the criteria (see the tables below). This means that the TEAP writing test score was slightly better predicted by the Aptis writing score than it was by the Aptis reading score of the test-takers. Although the study does not intend to validate the writing constructs of the Aptis and TEAP writing tests, the result has the important implication that different test products measure different types of writing skills. Integrated writing tasks, which combine reading and writing skills are more complexed than independent writing tasks due to the combination of a different set of skills.

The findings support early findings discussed in the literature review. Some scholars argued that reading and writing skills are only moderately correlated with integrated writing skills (e.g., Delaney, 2008), while others are concerned about the effects of textual borrowing (e.g., Cumming et al., 2005; Hyland, 2005; Shi, 2004; Yu, 2009). The current study also found it to be difficult to predict the integrated writing ability by just referring to either reading or writing scores.

Table 33. Aptis writing, reading, and TEAP scores in five criteria: correlations and descriptive statistics (N=40, Task 1 and Task 2)

Task1

Variables	1	2	3	4	5	6	7
1. Aptis Writing score	-						
2. Aptis Reading score	.56	-					
3. TEAP Main Idea	.63	.50	-				
4. TEAP Coherence	.62	.45	.93	-			
5. TEAP Cohesion	.63	.52	.82	.90	-		
6. TEAP Lexical Range & Accuracy	.65	.64	.87	.87	.85	-	
7. TEAP Grammatical Range & Accuracy	.62	.58	.80	.79	.84	.90	-

Task 2

Variables	1	2	3	4	5	6	7
1. Aptis Writing score	-						
2. Aptis Reading score	.56	-					
3. TEAP Main Idea	.51	.48	-				
4. TEAP Coherence	.55	.42	.90	-			
5. TEAP Cohesion	.56	.49	.91	.91	-		
6. TEAP Lexical Range & Accuracy	.47	.43	.88	.86	.91	-	
7. TEAP Grammatical Range & Accuracy	.52	.42	.87	.85	.91	.94	-

To gain an insight into test-takers' behaviours during the integrated writing tasks, this section of the chapter explored which of the eye-movement metrics can predict the five criteria that the TEAP Writing B has used for rating (i.e., Main Idea, Coherence, Cohesion, Lexical Range & Accuracy, Grammatical Range & Accuracy) by testing correlations between eye-tracking metrics and the five criteria scores (See Table 34 below).

In Task 1, the Time to First Fixation in Paragraph 1 showed statistically significant associations with the five criteria. Those who spent a long time from the start of the test before their eyes first fixated on Paragraph 1 returned statistically lower scores on the Main Idea, Coherence, Cohesion, Lexical Range & Accuracy, and Grammatical Range & Accuracy. This would mean that the participants were looking elsewhere, such as information from graphs or the source texts before reading the test instruction. It could be inferred that these test-takers tended to perform poorly on the TEAP integrated writing test when their first eye fixation on the Task Instruction was delayed. This calls for further investigation by qualitative analysis.

In Task 2, the First Fixation Duration on Paragraph 4 showed a statistically significant and positive associations across the five criteria. When the first fixation on the paragraph was long, the participants tended to have higher marks on the five criteria scores. Similarly, the Total Visit Duration on Paragraph 3 also showed statistically significant and positive associations across the five criteria.

In general, in both Task 1 and Task 2, the Total Fixation Durations of Paragraph 1, Paragraph 2, Paragraph 3 and Paragraph 4 seemed to be good indicators of how students performed in the tests. In Task 1, statistically significant positive associations were found with the Main Idea. In Task 2, the Total Fixation Duration on Paragraph 3 had statistically significant positive associations with all five analytical scores. Similar patterns prevailed in the Fixation Count and Visit Count of the paragraphs in Task 1 and 2.

Both positive and negative values were found within the same eye-movement metrics. The results called for a careful consideration of each case. For example, it cannot be automatically assumed that the test-takers would perform well when they viewed certain AOIs for longer. It could also be that the test-takers were struggling with the information in the AOI or they could be viewing the particular AOI to lift sentences from the source texts. As discussed earlier, test-takers who spent longer on certain AOIs did perform better on the TEAP test. Therefore, it seems unlikely that the test-takers had longer fixation because they were struggling with the information. To clarify this point, it was important to triangulate the data by looking more closely at the individual cases.

Table 34. Correlations between eye-tracking metrics and five criteria scores (40 minutes)

Pearson Correlations (N=40)

M=Main Idea, C=Coherence, CC=Cohesion,
L=Lexical Range & Accuracy, G=Grammatical Range & Accuracy
1=Task 1, 2=Task 2

Variables

	M1	C1	CC1	L1	G1	M2	C2	CC2	L2	G2
1. Time to First Fixation Essay Composition	.29	.32*	.28	.33*	.10	-.23	-1.01	-.00	-.23	-.02
2. Time to First Fixation Title (Pie Chart)	-.11	-.03	-.03	-.02	.01	-.17	-.18	-.22	-.21	-.16
3. Time to First Fixation Title (Line/Bar Graph)	-.06	-.12	-.11	-.10	-.20	-.20	-.13	-.28	-.32*	-.21
4. Time to First Fixation Letter-to-Editor	.10	.04	-.02	.02	-.07	-.08	-.00	-.02	-.12	-.09
5. Time to First Fixation Line/Bar Graph	.00	-.07	-.12	-.10	-.12	.06	-.07	.10	.09	.28
6. Time to First Fixation Newspaper Article	-.26	-.26	-.25	-.25	-.29	.07	.22	.32*	.14	.12
7. Time to First Fixation Paragraph1	-.35*	-.35*	-.33*	-.38*	-.39*	-.09	-.09	.12	.03	.18
8. Time to First Fixation Paragraph2	-.16	-.08	-.13	-.19	-.23	-.23	-.04	-.17	-.23	-.23
9. Time to First Fixation Paragraph3	-.28	-.25	-.26	-.32*	-.31	.12	.08	.27	.28	.25
10. Time to First Fixation Paragraph4	.02	.06	.11	.08	.07	.08	.00	.07	.12	.14
11. Time to First Fixation Pie Chart	-.19	-.19	-.23	-.23	-.22	.08	.14	.01	-.19	-.08
12. Time to First Fixation Sarah Case	.26	.26	.32*	.19	.23	.19	.28	.36*	.33*	.33*
13. Time to First Fixation Task Instruction	-.25	-.18	-.09	-.22	-.22	-.18	-.19	-.19	-.22	-.19
14. First Fixation Duration Essay Composition	-.19	-.20	-.10	-.11	-.08	.12	.11	-.06	.03	-.11
15. First Fixation Duration Title (Pie Chart)	.12	.12	.08	.26	.05	-.13	.08	-.03	-.00	-.09
16. First Fixation Duration Title (Line/Bar Graph)	-.07	-.07	-.10	-.04	-.11	.18	.10	-.03	.17	.07
17. First Fixation Duration Letter-to-Editor	.20	.19	.26	.19	.03	.10	.33	.25	.04	.10
18. First Fixation Duration Line/Bar Graph	-.00	.01	.04	.04	-.03	-.05	-.04	-.15	-.04	-.08
19. First Fixation Duration Newspaper Article	-.25	-.25	-.14	-.16	-.06	-.06	-.13	.15	-.07	-.02
20. First Fixation Duration Paragraph1	-.08	-.09	-.10	-.06	-.05	.25	.26	.16	.05	-.07
21. First Fixation Duration Paragraph2	.20	.16	.20	.13	.12	-.04	.18	.07	.05	-.00
22. First Fixation Duration Paragraph3	.04	.09	.20	.13	.14	-.02	.15	.06	-.03	.08
23. First Fixation Duration Paragraph4	.17	.16	.18	.09	.08	.36*	.33*	.44*	.39*	.38*
24. First Fixation Duration Pie Chart	.02	.10	.07	.10	-.14	-.02	.16	.11	.10	.01
25. First Fixation Duration Sarah Case	.10	.09	.18	.15	.24	-.02	.05	-.13	.07	-.07
26. First Fixation Duration Task Instruction	-.18	-.16	-.22	-.11	-.18	.15	.21	.11	.31*	.11
27. Fixation Duration Essay Composition Mean	-.05	-.05	.03	.05	-.03	.23	.32	.25	.28	.24
28. Fixation Duration Title (Pie Chart) Mean	.14	.03	.10	.25	.19	-.05	.11	-.00	.01	-.08
29. Fixation Duration Title (Line/Bar Graph) Mean	.04	.01	.04	.09	.06	.13	.11	-.06	.13	-.07

30.	Fixation Duration Letter-to-Editor Mean	.22	.21	.28	.18	.11	.01	.16	.22	-.07	.03
31.	Fixation Duration Line/Bar Graph Mean	.12	.06	.11	.21	.09	.10	.15	.08	.09	.00
32.	Fixation Duration News Article Mean	-.22	-.22	-.08	-.18	-.01	-.10	-.08	.15	-.17	-.09
33.	Fixation Duration Paragraph1 Mean	.14	.06	.18	.25	.13	.15	.20	.00	.09	-.05
34.	Fixation Duration Paragraph2 Mean	.19	.09	.15	.21	.12	.30	.41*	.04	.18	-.00
35.	Fixation Duration Paragraph3 Mean	.21	.11	.21	.31*	.24	.25	.28	.17	.26	.13
36.	Fixation Duration Paragraph4 Mean	.04	.07	.17	.12	.02	.01	.08	.15	.05	-.07
37.	Fixation Duration Pie Chart Mean	.11	.06	.13	.18	.03	.32*	.41*	.28	.30	.23
38.	Fixation Duration Sarah Case Mean	.19	.19	.26	.26	.31	-.00	.12	-.05	.14	-.00
39.	Fixation Duration Task Instruction Mean	.05	-.01	.07	.15	.09	.24	.21	.27	.29	.18
40.	Total Fixation Duration Essay Composition	.22	.21	.24	.29	.24	.37*	.31*	.17	.41*	.28
41.	Total Fixation Duration Title (Pie Chart)	.11	.07	.19	.14	.10	.17	.18	.24	.11	-.05
42.	Total Fixation Duration Title (Line/Bar Graph)	.01	-.08	.02	.2	.02	.26	.14	.11	.13	-.03
43.	Total Fixation Duration Letter-to-Editor	.23	.15	.22	.22	.29	-.05	.08	.17	-.06	-.04
44.	Total Fixation Duration Line Graph	.12	.01	.13	.11	.12	.15	.10	.23	.13	.05
45.	Total Fixation Duration News Article	-.06	-.10	-.08	-.10	.03	.04	.10	.03	-.18	-.09
46.	Total Fixation Duration Paragraph1	.31*	.22	.26	.24	.24	.31*	.32*	.23	.17	.1
47.	Total Fixation Duration Paragraph2	.33*	.29	.30	.25	.27	.27	.32*	.00	.10	.09
48.	Total Fixation Duration Paragraph3	.41*	.27	.28	.29	.28	.42*	.38*	.47*	.32*	.32*
49.	Total Fixation Duration Paragraph4	.31*	.22	.22	.22	.22	.35*	.28	.22	.16	.17
50.	Total Fixation Duration Pie Chart	.08	-.04	.05	.01	-.04	.30	.19	.22	.31*	.14
51.	Total Fixation Duration Sarah Case	.17	.15	.20	.23	.20	.02	.09	-.11	.05	-.10
52.	Total Fixation Duration Task Instruction	-.02	-.11	-.06	-.03	-.02	-.07	-.11	.15	-.05	-.06
53.	Fixation Count Essay Composition	.21	.17	.21	.28	.25	.33	.27	.11	.39*	.24
54.	Fixation Count Title (Pie Chart)	.08	.04	.16	.09	.05	.13	.19	.19	.06	-.06
55.	Fixation Count Title (Line/Bar Graph)	-.01	-.10	.00	-.00	-.00	.23	.13	.15	.10	-.00
56.	Fixation Count Letter-to-Editor	.17	.08	.14	.17	.22	-.05	.07	.17	-.07	-.05
57.	Fixation Count Line/Bar Graph	.07	-.02	.10	.05	.09	.09	.05	.19	.08	.02
58.	Fixation Count News Article	.08	-.13	-.12	-.12	-.00	.13	.18	.05	-.12	-.06
59.	Fixation Count Paragraph1	.30	.22	.26	.22	.24	.36*	.35*	.30	.21	.21
60.	Fixation Count Paragraph2	.32*	.27	.29	.23	.29	.25	.2	-.00	.11	.14
61.	Fixation Count Paragraph3	.39*	.25	.26	.26	.27	.43*	.39*	.49*	.32*	.36*
62.	Fixation Count Paragraph4	.31*	.22	.32*	.27	.40*	.35*	.26	.22	.14	.19
63.	Fixation Count Pie Chart	-.04	-.11	-.00	-.05	-.09	.24	.10	.19	.26	.11
64.	Fixation Count Sarah Case	.22	.17	.22	.28	.24	.07	.13	-.02	.16	-.00
65.	Fixation Count Task Instruction	-.03	-.13	-.09	-.05	-.04	-.10	-.12	.12	-.08	-.08
66.	Visit Duration Essay Composition Mean	-.11	-.14	-.07	.05	-.06	-.24	-.16	-.21	-.16	-.20
67.	Visit Duration Title (Pie Chart) Mean	.32*	.18	.23	.35*	.28	-.13	.10	.03	-.07	-.17
68.	Visit Duration Title (Line/Bar Graph) Mean	.05	-.02	.04	.12	.17	.07	.04	-.04	.02	-.09

69.	Visit Duration Letter-to-Editor Mean	.21	.18	.25	.28	.18	.05	.11	.03	-.03	-.05
70.	Visit Duration Line/Bar Graph Mean	-.19	-.10	-.12	-.19	-.26	-.12	-.11	-.00	-.05	.06
71.	Visit Duration Newspaper Article Mean	-.01	-.04	.00	-.00	.16	-.08	-.07	.02	-.26	.05
72.	Visit Duration Paragraph1 Mean	.01	.09	.06	.11	-.01	-.13	-.08	.06	-.16	-.12
73.	Visit Duration Paragraph2 Mean	-.14	-.08	-.00	-.18	-.18	.13	.23	-.02	.00	.00
74.	Visit Duration Paragraph3 Mean	-.05	-.02	.10	-.04	-.06	.18	.26	.15	.13	.02
75.	Visit Duration Paragraph4 Mean	-.15	-.14	.01	.00	.10	.14	.11	.08	.12	.14
76.	Visit Duration Pie Chart Mean	-.05	.03	.03	-.00	-.12	-.06	.00	.27	-.03	-.01
77.	Visit Duration Sarah Case Mean	.28	.26	.31	.32*	.40*	.02	.12	.01	.18	.04
78.	Visit Duration Task Instruction Mean	-.21	-.17	-.18	-.19	-.20	.01	-.09	-.09	-.02	.00
79.	Total Visit Duration Essay Composition	.15	.09	.14	.29	.23	.21	.14	-.05	.28	.07
80.	Total Visit Duration (Title Pie Chart)	.11	.06	.18	.13	.09	.11	.15	.22	.07	-.08
81.	Total Visit Duration (Title Line/Bar Graph)	-.01	-.12	-.02	-.00	.00	.22	.12	.09	.09	-.00
82.	Total Visit Duration Letter-to-Editor	.20	.12	.21	.21	.28	-.04	.06	.17	-.02	-.00
83.	Total Visit Duration Line /Bar Graph	.03	-.01	.06	-.04	-.03	-.06	-.06	.07	.00	.05
84.	Total Visit Duration Newspaper Article	-.03	-.08	-.08	-.08	.05	.04	1.0	.03	-.19	-.09
85.	Total Visit Duration Paragraph1	.27	.21	.24	.21	.24	.22	.25	.26	.11	.17
86.	Total Visit Duration Paragraph2	.15	.14	.20	.09	.14	.15	.21	-.1	.04	.12
87.	Total Visit Duration Paragraph3	.32*	.17	.19	.18	.22	.40*	.37*	.48*	.32*	.36*
88.	Total Visit Duration Paragraph4	.28	.17	.27	.22	.37*	.27	.21	.19	.12	.22
89.	Total Visit Duration Pie Chart	-.12	-.12	-.03	-.08	-.14	.13	.07	.30	.16	.07
90.	Total Visit Duration Sarah Case	.22	.20	.23	.30	.25	.03	.08	-.05	.11	.00
91.	Total Visit Duration Task Instruction	-.19	-.24	-.21	-.18	-.18	-.14	-.17	.07	-.10	-.07
92.	Visit Count Essay Composition	.35*	.23	.26	.31*	.24	.54*	.41*	.18	.47*	.35*
93.	Visit Count (Title Pie Chart)	-.02	-.06	.04	-.01	-.05	.29	.22	.32	.21	.06
94.	Visit Count (Title Line/Bar Graph)	-.08	-.18	-.09	-.09	-.12	.17	.08	.16	.09	.03
95.	Visit Count Letter-to-Editor	.20	.12	.16	.14	.22	-.08	.08	.26	-.06	.03
96.	Visit Count Line/Bar Graph	.02	-.06	.04	-.00	.01	.09	.02	.13	.08	.0
97.	Visit Count Newspaper Article	-.20	-.23	-.19	-.22	-.11	.14	.19	.10	-.06	-.01
98.	Visit Count Paragraph1	.25	.16	.18	.14	.17	.40*	.33*	.27	.31*	.27
99.	Visit Count Paragraph2	.30	.025	.25	.23	.25	.17	.13	.00	.14	.18
100.	Visit Count Paragraph3	.47*	.31*	.31	.30	.30	.43*	.33*	.43*	.35*	.41*
101.	Visit Count Paragraph4	.44*	.33*	.37*	.29	.39*	.29	.21	.21	.19	.22
102.	Visit Count Pie Chart	-.10	-.17	-.07	-.08	-.11	.27	.12	.18	.24	.11
103.	Visit Count Sarah Case	.15	.11	.18	.28	.19	.03	.07	-.06	.09	-.04
104.	Visit Count Task Instruction	-.16	-.23	-.19	-.15	-.16	.02	.02	.27	.01	.01

Note: Numbers in bold means significantly significant

*The difference is significant when $p < .05$

4.6.3 Summary of the TEAP essay ratings and eye-tracking results

This section has explored associations between eye-tracking data and the TEAP essays based on the analytical scores. The purpose was to investigate which of the five assessment criteria were closely related with eye-tracking metrics.

Key findings include that eye-tracking metrics were perhaps more associated with the Main Idea among the five criteria. For example, Total Fixation Durations and Fixation Counts in Paragraph 1, Paragraph 2, Paragraph 3 and Paragraph 4 were found to be significant ($p < .05$) for the Main Idea. In particular, Total Visit Duration in Paragraph 3 and Visit Count in Essay Composition also had close associations with the scores of the Main Idea in both Task 1 and 2.

There were not many variables that had statistically significant values for Coherence and Cohesion, except Time to First Fixation in Paragraph 1 and Visit Count in Paragraph 4. Lexical Range & Accuracy and Grammatical Range and Accuracy also did not find many variables that were associated with eye-tracking metrics, although some variables, such as Fixation Count, Visit Count and Total Visit Duration on Paragraph 3, were associated with them.

Although the eye-tracking tests were conducted twice to avoid one shot results, the findings reported in this study are still exploratory as the samples were limited to 40 students. The eye-movements showed virtually no association with proficiency when measured by the Aptis tests, but quite a lot of association with proficiency in terms of performance on the TEAP test itself.

However, the results cannot be replicated given different settings, test-taking strategies of the individuals, and therefore the author acknowledges that further research is needed before definitive conclusions can be drawn from any of the findings reported above.

In addition to the quantitative findings, the study also attempts in later chapters to offer some interpretations of the eye-tracking data by looking at some individual cases qualitatively. By triangulating data from the cognitive processing questionnaires and the focus group discussions,

it might be possible to shed some light on how students behave during the task, and what information these eye movement metrics contain.

Chapter 5. Findings of Cognitive Processing Questionnaires

The cognitive processing questionnaires were conducted immediately after the eye-tracking experiment. The questionnaire was prepared based on Weir's socio-cognitive processing framework for L2 writers (See Table 35 for overall structure of the questionnaire). Chi-squared tests (N=37) were conducted to compare the differences between the groups of lower (A2 and Below-A2, N=21) and higher (B1 and B2, N=16) level students based on the overall TEAP rating scores. There were only a small number of students below A2 (N=3) and just one B2 level student.

Table 35. A cognitive processing framework for L2 writers

(1) Preparing to Write	
Task Representation	Did you understand what is required for the task?
Macro-planning	How did you plan when you started to write an essay?
Micro-planning	How did you decide which information to include?
(2) Translation and Writing	
Describing	What kind of strategies did you use during the test?
Synthesizing	What did you often do while you were writing
Choosing positions	
Giving reasons	
(3) Monitoring and Revising	
Task requirement	What did you do after you finish writing an essay?

Reference: Adopted from Socio-Cognitive Framework for Language Testing (Shaw & Weir, 2007)

The questionnaire had three components. The first segment was focused on computer literacy, the second segment was about familiarity with the graph types, and the third section focused on identifying the different test-taking strategies used during different stages of cognitive processes when writing. Although some questions (e.g., computer literacy) appear to have no direct connection with the integrated writing processes, the eye-tracking experiments were conducted on the computer rather than using the paper tests, and it was therefore important to consider whether familiarity with the keyboard had affected the test-takers' writing processes.

5.1 Computer Literacy

This section provides descriptive statistics on patterns of computer literacy and their association with proficiency on the TEAP test. In general, chi-squared tests found few significant differences in computer literacy between the more- and less-proficient students.

Figure 14 describes at what age the participants began using the keyboard in order to find out their computer literacy for writing activities.

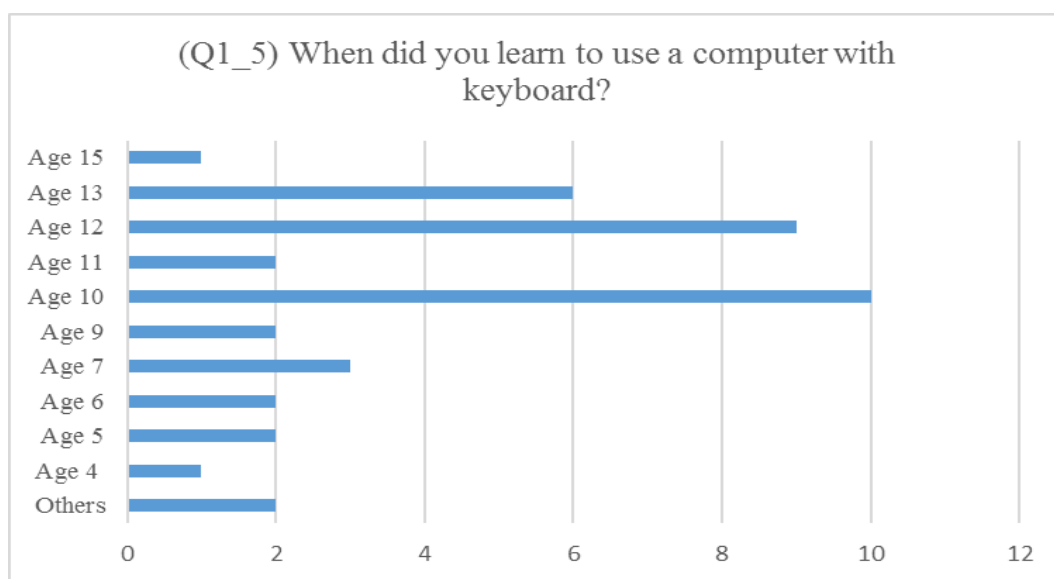


Figure 14. Questionnaire answers on Age of First Computer Access (Q1_5)

Table 36. Question 1-1

Computer access at home (N=37)	Q1-1		Total
	Yes	No	
TEAP A2 and Below-A2	76%	24%	22 (100%)
TEAP B1 and B2	100%	0%	15 (100%)

$X^2 (1, N =37) = 3.94, p<.05$

As shown, the majority of students (approximately 76% of A2 and below used a computer at home, and 100% of B1 and above) had access to computers at home.

Table 37. Question 1-2

Frequency (at home) (N=37)	Q1-2						Total
	Never	1-2 times /semester	1-2 times /month	1-2 times /week	3-4 times /week	No Answer	
TEAP A2 and Below- A2	14%	19%	38%	14%	10%	5%	22 (100%)
TEAP B1 and B2	6%	12%	38%	6%	38%	0%	15 (100%)

$\chi^2 (3, N = 37) = 8.168, p < .05$

However, significance was found in the group on how often they use a computer at home. Approximately 14% of A2 and below and 6% of B1 and above students never used computer at home. This implied that the students from the lower proficiency group probably did not have much practice at home.

Table 38. Question 1-3

Computer access at school (N=37)	Q1-3		Total
	Yes	No	
TEAP A2 and Below-A2	81%	19%	22 (100%)
TEAP B1 and B2	81%	19%	15 (100%)

$\chi^2 (3, N = 37) = 8.82, p < .05$

At school, most of the students seemed to have computer access. Compared to the access at home (76%), more participants at A2 and below levels had slightly better access at school (81%).

Table 39. Question 1-4

Frequency (at school) (N=37)	Q1-4					Total
	Never	1-2 times /semester	1-2 times /month	1-2 time /week	3-4 times /week	
TEAP A2 and Below-A2	48%	14%	9%	29%	0%	22 (100%)
TEAP B1 and B2	37%	0%	44%	6%	13%	15 (100%)

$\chi^2 (1, N =37) =3.94, p<.05.$

A statistically significant difference was found between the two groups of students in the analysis in terms of how many times they used computers at school. Forty-eight percent of the students at A2 and below levels answered that they never use computers, even though they have access to a computer at school, while about the same percentage of the B1 and B2 level students said that they use computers at school once or twice per month.

Table 40. Question 1-6

Use of Keyboard (N=37)	Q1-6			Total
	Hiragana	Alphabet	No Reply	
TEAP A2 and Below-A2	14%	86%	0%	22 (100%)
TEAP B1 and B2	0%	94%	6%	15 (100%)

$\chi^2 (1, N =37) =5.13, p>.05.$

Japanese computers usually have two ways of typing letters on the keyboard-Hiragana (Japanese) or the Roman alphabets. Most participants seemed to know how to type the keyboard with alphabetical letters. As many as 14 % of A2 and A2 below students use Hiragana as a means of typing. Still, the majority of them knew how to use the keyboard using alphabetical letters. Although the students seemed more comfortable writing essays with a pen and paper, the participants in this group were still fairly confident with typing on the computer screen.

Table 41. Question 1-7

How easy or difficult to type English using Keyboard (N=37)	Q1-7					Total
	Very Difficult	Difficult	Neutral	Easy	Very Easy	
TEAP A2 and Below-A2	9%	19%	48%	19%	5%	22 (100%)
TEAP B1 and B2	0%	31%	25%	44%	0%	15 (100%)

$\chi^2 (4, N =37) =9.59, p<.05$

When the participants were asked how easy or difficult it was to use the keyboard, nine percent of the A2 and below students responded that it was very difficult. The results suggest that those who learned to type with alphabetical letters on a keyboard had an advantage in terms of writing essays on a computer.

Based on the results of the survey of computer literacy, it can be concluded that the writing proficiencies of participants in this group were somewhat affected by familiarity of the keyboards.

5.2 Familiarity with Graphs

The students were asked to rate each graph (i.e., the two pie charts, line graph and bar graph) in order of difficulty on a scale of 5, with 1 being the most difficult to understand and 5 being the easiest. In designing the two TEAP sample tests for eye-recordings, different types of graphs were presented to study the effects of graph types. In Task 1, a line graph entitled, “Average Hours of Sleep among High-School Students in Greenhill” was placed in the upper left-hand corner, as well as a pie chart entitled, “How Students Spent After School Hours in 2015” in the upper right-hand corner. In Task 2, a pie chart entitled, “Sources of Food Waste” was in the upper left-hand corner, and a bar graph entitled, “Food Waste in Greenfield

(collected as garbage)” in the upper right-hand corner. The results showed that the majority of students found it easy or very easy to understand all types of graphs.

Table 42. Question 2-1

Understanding of the Line Graph in Task 1 (N=37)	Q2-1					Total
	Very difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	5%	5%	19%	57%	14%	22 (100%)
TEAP B1 and B2	0%	6%	6%	44%	44%	15 (100%)

$\chi^2(10, N =37) =12.93, p=>.05$

Over eighty percent of the B1 and B2 level students found that Line Graph in Task 1 was either “Easy” or “Very easy” to understand. The majority of A2 and A2 level students also chose either “Easy” or “Very Easy” although there were about 10% who said, “Very difficult” or “Difficult.”

Table 43. Question 2-2

Understanding of the Pie Chart in Task 1 (N=37)	Q2-2					Total
	Very difficult	Difficult	Neutral	Easy	Very Easy	
TEAP A2 and Below-A2	5%	0%	19%	43%	33%	22 (100%)
TEAP B1 and B2	0%	6%	6%	38%	50%	15 (100%)

$\chi^2(2, N =37) =3.56, p=>.05$

Likewise, most of the students answered that the Pie Chart in Task 1 was also either “easy” or “very easy” to understand.

Table 44. Question 2-3

Understanding of the Bar Graph in Task 2 (N=37)	Q2-3					Total
	Very difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	5%	0%	19%	43%	33%	22 (100%)
TEAP B1 and B2	0%	0%	6%	50%	44%	15 (100%)

$X^2(4, N = 37) = 2.47, p = >.05$

Similarly, the same participants found that the Bar Graph was also “easy” or “very easy” to understand. The proportion of the students who stated “easy” and “very easy” is also comparable with the ones for the Pie Chart.

Table 45. Question 2-4

Understanding of the Pie Chart in Task 2 (N=37)	Q2-4					Total
	Very difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	5%	0%	14%	52%	29%	22 (100%)
TEAP B1 and B2	0%	0%	6%	44%	50%	15 (100%)

$X^2(4, N = 37) = 4.64, p = >.05$

The results indicated similar trends in respect to all the graph types presented. The majority of the participants seemed to find it easy or very easy to understand the information contained in the graphs. These answers appeared to be contradicted, however, when the same participants were later asked whether it was easy or difficult to interpret the trends in the graphs and to summarize those trends in English (Q5-4 & Q5-5). Further details will be shown in the next section (Section 5.3) which summarizes the answers in respect to the different cognitive stages of writing.

5.3 Test-taking strategies during different cognitive stages in the writing process

The third part of the questionnaire focused on understanding participants' decision-making during the integrated writing processes. There were three main stages in the cognitive processes of writing according to Weir's socio-cognitive approach; (1) preparing-to-write, (2) translation and writing, and (3) monitoring and revising stages. The writing processes mirrored the decision-making processes of each participant. The purpose of conducting the survey was to identify or characterize the use of test-taking strategies according to different language proficiency levels. Initially, cross-tabulations were created according to four CEFR levels of writing proficiency based on the TEAP essays (Below-A2, A2, B1 and B2). The number of test-takers who scored below A2 or B2 were minimal. To make sure the numbers of students in the two groups was even, the rest of the students were divided into two groups which consisted of a lower group (Below-A2 and A2 level students, N=21) and a higher group (B1 and B2 level students, N=16). Chi-square tests were conducted between the two groups. As reported below, only few statistically significant differences were found in these variables between TEAP proficiency groups.

5.3.1. Preparing-to-Write

Task Representations

To judge the extent to which the participants understood the requirements of the integrated writing task for the test, they were asked whether they followed the instructions, how difficult or easy it was to fulfil the task, and which one of the tasks they found most challenging.

Table 46. Question 3-1

Did you understand the instructions on how to write your essay?	Q3-1		Total
	Yes	No	
TEAP A2 and Below-A2	52%	48%	22 (100%)
TEAP B1 and B2	81%	19%	15 (100%)

$X^2 (4, N =37) =1.94, p>.05$

Nearly half of the students who were at A2 or Below-A2 levels seemed to have struggled with understanding the task requirement for the tests.

Table 47. Question 3-2

How easy or difficult did you find it to fulfil the task requirement?	Q3-2					Total
	Very difficult	Difficult	Neutral	Easy	Very Easy	
TEAP A2 and Below-A2	5%	33%	48%	9%	5%	22 (100%)
TEAP B1 and B2	0%	31%	50%	19%	0%	15 (100%)

$X^2 (3, N =37) =8.82, p>.05$

Approximately one-third of the students in both groups said it was difficult to fulfil the task requirement (36% for A2 and Below-A2 level students, 31% for B1 and B2 level students).

Table 48. Question 3-3

Which part of the task requirements did you find most challenging?	Q3-3				Total
	Describing the situation	Summarizing the main points	Stating which solution might work best	Giving the reasons for the choice you made	
TEAP A2 and Below-A2	0%	29%	33%	38%	22 (100%)
TEAP B1 and B2	6%	19%	19%	56%	15 (100%)

$X^2 (4, N =37) =2.62, p>.05$

When they were asked which tasks they found most difficult, both groups responded that “giving the reasons for the choice” was the hardest. It is worth mentioning that 6% of the upper group chose “describing the situation” as the most challenging.

Macro-Planning [Response format]

The following questions were asked to see whether the students had engaged in pre-writing tasks by identifying the purpose of the composition, deciding which solutions would work the best (i.e., opinions), and organizing their ideas by outlining and planning how many paragraphs the essay would be.

Table 49. Question 4-1.

Did you identify the purpose of the essay?	Q4-1			Total
	Yes	No	No Answer	
TEAP A2 and Below-A2	62%	33%	5%	22 (100%)
TEAP B1 and B2	75%	25%	0%	15 (100%)

$X^2(1, N = 37) = 1.77, p > .05$

The majority of the participants seemed confident that they knew the purpose of the essay, given the topics and the instruction.

Table 50. Question 4-2

Did you think about which solutions would work the best?	Q4-2			Total
	Yes	No	No Answer	
TEAP A2 and Below-A2	47%	47%	4%	22 (100%)
TEAP B1 and B2	75%	25%	0%	15 (100%)

$X^2(4, N = 37) = 1.68, p > .05$

Furthermore, it is worth noting that half of the A2 and Below-A2 level participants claimed that they did not think about which solutions would work best in the essay, and what the most important task requirement in the essay was.

Table 51. Question 7-1

Did you make an outline BEFORE writing your essay?	Q7-1			Total
	Yes	No	I don't remember	
TEAP A2 and Below-A2	5%	90%	5%	22 (100%)
TEAP B1 and B2	13%	81%	6%	15 (100%)

$X^2 (1, N =37) =1.69, p>.05$

The result for Q7-1 on whether they made an outline before writing your essay did not find much difference between the two groups. Most of the students did not have any pre-planning of ideas.

Table 52. Question 7-2

Did you decide how many paragraphs there should be in your essay?	Q7-2				Total
	Yes	No	I don't remember	No Answer	
TEAP A2 and Below-A2	33%	53%	14%	0%	22 (100%)
TEAP B1 and B2	31%	50%	13%	6%	15 (100%)

$X^2 (4, N =37) =2.80, p>.05$

Approximately half of the participants in both groups did not think about how many paragraphs there should be in their essay. Each paragraph usually represents one main idea. If the test-takers wanted to develop a cohesive essay, they needed to organize sub-topics in a multi-paragraph essay.

Micro-Planning [Information from graphs]

This section of the questionnaire was concerned about the information from graphs. The students were asked whether they experienced any difficulties in reading the titles of the graphs, values on the graphs, the units on X-and Y-axis, interpreting the information in the graphs and summarizing the main trends in English.

Table 53. Question 5-1

How easy or difficult was it for you to read the titles of the graphs?	Q5-1					Total
	Very difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	9%	5%	43%	33%	10%	22 (100%)
TEAP B1 and B2	0%	12%	19%	50%	19%	15 (100%)

$X^2 (3, N =37) =2.98, p>.05$

Most of the participants found the titles of the graphs either “easy” or “neutral” to understand.

Table 54. Question 5-2

How easy or difficult was it for you to read the values on the graphs?	Q5-2					Total
	Very Difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	0%	10%	14%	52%	24%	22 (100%)
TEAP B1 and B2	0%	6%	12%	69%	13%	15 (100%)

$X^2 (3, N =37) =1.72, p>.05$

An even bigger proportion of the students (78 percent) rated that it was easy to read the values on the graphs.

Table 55. Question 5-3

How easy or difficult was it for you to read the units on the x- and y-axis on the graphs?	Q5-3					Total
	Very difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	14%	0%	29%	33%	24%	22 (100%)
TEAP B1 and B2	6%	0%	13%	56%	25%	15 (100%)

$X^2 (3, N =37) =3.12, p>.05$

When they were asked about the units on the x- and y-axis on the graphs, although some students claimed that they were very difficult to understand, most of the participants in both groups found it either “easy” or “very easy” to read the units on the graphs.

Table 56. Question 5-4

How easy or difficult was it for you to interpret the information in the graphs?	Q5-4					Total
	Very difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	10%	19%	33%	33%	5%	21 (100%)
TEAP B1 and B2	0%	6%	19%	69%	6%	16 (100%)

$\chi^2 (4, N =37) =6.32, p>.05$

The Chi-square test did not find much difference in the difficulty in interpreting the information from the graphs between those with higher (B2 and B1) and lower (A2 and Below-A2) levels of writing skills, as measured by the TEAP rating scores.

Table 57. Question 5-5

How easy or difficult was it for you to summarize the main trends of the graphs in English?	Q5-5					Total
	Very difficult	Difficult	Neutral	Easy	Very easy	
TEAP A2 and Below-A2	19%	48%	19%	9%	5%	22 (100%)
TEAP B1 and B2	12%	25%	44%	19%	0%	15 (100%)

$\chi^2 (3, N =37) =3.07, p>.05$

Finally, the survey (Q5-5) unveiled that more than half of the lower group students found that summarizing the essential information from the graphs in English was very difficult (Q5-4). As far as graph familiarity was concerned, the students across the language proficiency levels

found they were able to understand the information presented in all types of graphs. There was a statistically significant difference between the two groups of students with different writing proficiencies (Q5-4) in terms of interpreting the information from graphs. Summarizing the main trends of graphs in English was difficult for students in both groups, and it appeared they lacked the vocabulary to describe the trends in the graphs.

Micro-Planning [Text Information]

This section of the questionnaire focused on the ways in which students synthesized the text information in their essays. The first question simply asked whether they re-read the parts of the essay to decide which information to include. The students with lower proficiency tended to answer “No”. The following question asked how they decided which text information to include. The students at the A2 and below level relied on the keywords in the texts and some even wrote the essay “by memory”, which would probably have meant that they wrote an independent essay on the topic.

Table 58. Question 6-1

Did you re-read the parts where you thought it was important to include in your essay?	Q6-1		Total
	Yes		
TEAP A2 and Below-A2	81%	19%	22 (100%)
TEAP B1 and B2	87%	13%	15 (100%)

$X^2 (3 N =37) =2.91, p>.05$

Most of the students (81% at lower level, 87% at higher level) answered that they re-read the parts where they believed the key information was to be found in the source texts.

Table 59. Question 6-2

How did you decide which information to include from the texts?	Q6-2					Total
	Never thought about it	By re-reading the whole text once or twice	By going back to a specific paragraph of some importance	By scanning and looking for the keywords	By memory	
TEAP A2 and Below-A2	5%	14%	38%	33%	10%	21 (100%)
TEAP B1 and B2	0%	44%	31%	19%	6%	16 (100%)

$X^2 (4, N =37) =.586, p>.05$

The decisions as to which information to include from the texts are shown in Table 60. The participants at B1 and B2 levels said that they did this by re-reading the whole text once or twice, or by going back to specific paragraphs. Surprisingly, approximately one-third of the participants at A2 and Below-A2 levels answered that they decided which information to include by scanning for the keywords and 10% reported that they did it based on their memory. This response may have meant that they wrote an independent essay based on their own experience of the topic.

Table 60. Question 6-3

How did you connect the ideas from the texts?	Q6-3					Total
	Never thought about it	By copying the sentences from the source texts	By borrowing some words and phrases from the source texts	By referring to ideas from the source texts	By memory	
TEAP A2 and Below-A2	5%	24%	38%	24%	9%	22 (100%)
TEAP B1 and B2	6%	37%	38%	13%	6%	15 (100%)

$X^2 (4, N =37) =8.12, p>.05$

Question 6-3 revealed the participants' recollection of how they inferred the ideas from the source texts for summarizing the essay. It is worth noting that about one-third of the students (24% at A2 and A2-below, 37% at B1 and B2 levels) admitted that they either copied the sentences directly from the source texts or borrowed some phrases. Those who answered that they summarized the ideas based on memory probably produced an independent essay based on their past experience about the topic as often witnessed among the essays written by students in lower proficiency groups (see more evidence in Chapter 6).

5.3.2. Translating and Writing

Fulfilling the task requirements

This section of the questionnaire focused on the self-evaluation of the task responses during the essay writing. The students were asked how well they described the situations, how well they summarized the main points (of the solutions), how well they described which option would work as a solution, and how well they gave the reasons to support such a choice.

Table 61. Question 8-1

How well do you think you have described the situation?	Q8-1					Total
	Very poor	Poor	Average	Good	Very good	
TEAP A2 and Below-A2	14%	29%	38%	19%	0%	22 (100%)
TEAP B1 and B2	0%	31%	19%	44%	6%	15 (100%)

$X^2(4, N=37) = .928, p > .05$

A total of forty-three percent of the A2 and Below-A2 students thought they did either very poorly (14%) or poorly (29%) in describing the situations. On the contrary to their self claim, the finding from an early question (Q3-3) showed that none of this group chose describing the situation as the most challenging task.

Table 62. Question 8-2

How well do you think you have summarized the main points?	Q8-2					Total
	Very poor	Poor	Average	Good	Very good	
TEAP A2 and Below-A2	5%	38%	48%	9%	0%	22 (100%)
TEAP B1 and B2	6%	19%	37%	38%	0%	15 (100%)

$X^2 (2, N =37) =1.04, p>.05$

In the previous questions (Q3-3), only one-third of the lower group students thought that summarizing the main point was the most challenging task to fulfil. Here in Question 8-2, however, up to 43 percent of the A2 and below A2 students rated that they did “very poorly” or “poorly” in summarizing the main points of the source texts.

Table 63. Question8-3

How well do you think you have stated which solution would work best?	Q8-3					Total
	Very poor	Poor	Average	Good	Very good	
TEAP A2 and Below-A2	14%	52%	24%	10%	0%	22 (100%)
TEAP B1 and B2	12%	19%	31%	48%	0%	15 (100%)

$X^2 (3, N =37) =1.58, p>.05$

An even bigger proportion of the A2 and Below-A2 students claimed that stating the best solutions was done poorly or very poorly in their essay writing, while the majority of B1 and B2 level students claimed they did “good” or “average” on the job.

Table 64. Question 8-4

How well do you think you have given the reasons for the choice made?	Q8-4					Total
	Very poor	Poor	Average	Good	Very good	
TEAP A2 and Below-A2	14%	48%	29%	9%	0%	22 (100%)
TEAP B1 and B2	6%	44%	37%	13%	0%	15 (100%)

$X^2 (4, N =37) =7.70, p>.05$

The question (Q8-4) validated that giving the reasons for the choice they made was the most challenging job in the essay task. The answer was consistent with their previous response (Q3-3).

Use of Language

This part of the questionnaire was concerned with the use of language for essay composition. For example, the students were asked about their awareness of the verb tense and the use of pronouns. They were also asked which verb tense or form of pronouns they mostly used. The assumption was made that they would only use the first-person pronoun in the conclusion where they needed to state their opinion as to which solution works the best. However, they should mostly choose to use the third-person pronoun in the rest of the essay.

Table 65. Question 8-5

Did you think what verb tense form to be used before writing?	Q8_5		Total
	Yes	No	
TEAP A2 and Below-A2	62%	38%	22 (100%)
TEAP B1 and B2	56%	44%	15 (100%)

$X^2 (3, N =37) =2.53, p>.05$

Table 66. Question 8-6

When did you decide which verb to use?	Q8-6					Total
	Don't Remember	After I finished writing the Essay	When I was writing the essay	Before I started writing the essay	No Answer	
TEAP A2 and Below-A2	14%	5%	76%	0%	5%	22 (100%)
TEAP B1 and B2	13%	6%	69%	6%	6%	15 (100%)

$X^2 (3, N =37) =3.1, p>.05$

When synthesizing the information from the past events as described in the newsletter article, one would assume that the past tense would be used. Many of the students, however, reported that they mostly used the “Present tense” in the essay (Q8-7).

Table 67. Question 8-7

Which tense did you mostly use?	Q8-7					Total
	Don't Remember	Future tense	Present tense	Past tense	No Answer	
TEAP A2 and Below-A2	14%	5%	67%	9%	5%	22 (100%)
TEAP B1 and B2	13%	6%	56%	25%	0%	15 (100%)

$X^2 (3, N =37) =.781, p>.05$

Table 68. Question 8-8

Did you think which pronoun form should be used for writing?	Q8_8				Total
	Yes	No	Don't remember	No Answer	
TEAP A2 and Below-A2	62%	9%	10%	19%	22 (100%)
TEAP B1 and B2	88%	6%	0%	6%	15 (100%)

$X^2 (1, N =37) =.836, p>.05$

Table 69. Question 8-9

Which pronoun form did you use mostly use?	Q8-9				Total
	Proper nouns	First person pronouns (I / We)	Third person pronouns (It/She/He/They)	No Answer	
TEAP A2 and Below-A2	5%	48%	14%	33%	22 (100%)
TEAP B1 and B2	13%	56%	33%	0%	15 (100%)

$X^2 (4, N =37) =2.90, p>.05$

As far as the use of pronouns was concerned, most of the students responded “Yes”, that they did consider what types of pronoun would be appropriate while writing their essay (Q8-8).

When they were asked which ones they used the most (Q8-9), approximately half of the

participants said they have used “the first-person pronoun (I/You).” One-third of the students at A2 and Below-A2 levels did not even answer the question. For the use of pronouns, the writers would have only needed to use first-person pronouns where they needed to state their own opinions and reasons behind the choice. For most of the essay, the writers should have used the third person pronouns to describe the situation and summarize the main points by inferring to the person who brought the ideas. Although no significance was found, the questionnaires suggested that the participants would have probably struggled to summarize tasks to a certain degree.

5.3.3. Monitoring & Revising

The last segment of the questionnaires was concerned about the students’ monitoring and revising of their written texts as a part of the writing process. The students were asked whether they checked grammar (Q9-1), spellings (Q9-2), the ideas connected with graphs (Q9-3), logical coherence (Q9-4), fulfilment of the tasks (Q9-5), avoiding repetition of the same words (Q9-6), appropriate lengths (Q9-7), and avoiding the use of memorized phrases (Q9-8). While there were no significance differences found in most of the questions including Table 70 on reviewing for spelling mistakes, it was interesting to find a significance found in Q 9-5 on fulfilling the task requirement.

Table 70. Question 9-2

My spelling was correct	Q9-2					Total
	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree	
TEAP A2 and Below-A2	5%	5%	0%	90%	0%	22 (100%)
B1 and B2	6%	25%	12%	44%	13%	15 (100%)

$X^2 (5, N =37) =8.27, p>.05$

Table 71. Question 9-5

I checked if I had fulfilled the task requirement by going back to the instruction	Q9-5					Total
	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree	
TEAP A2 and Below-A2	5%	5%	0	90%	0	22 (100%)
B1 and B2	7%	27%	13%	40%	13%	15 (100%)

$\chi^2 (4, N =37) =12.46, p<.05$

Overall, the cognitive processing questionnaire which included computer familiarity and test-taking strategies, was conducted immediately after the eye-tracking tests. The self-reported answers were based on the test-takers' understanding of their performance. The questions were laid out according to Weir's cognitive processes of writing. It was interesting how students contradicted themselves when the questions were asked differently (e.g. familiarity with the graphs). For example, fewer people thought that summarizing the main points was difficult to do, but when asked for the second time how well they have summarized the main points, more participants corresponded it was they did "very poorly" or "poorly."

The Chi-square tests between the upper (TEAP B1 and B2 levels) and lower (TEAP Below-A2 and A2 levels) groups of students in questionnaire helped identify some interesting trends that could not be explained by the eye-tracking data alone. For example, the participants did not seem to have any problems understanding the information presented in the graphs. The lower proficiency students, however, claimed that they had difficulties both in interpreting the main trends and describing the patterns in their second language. The eye-tracking data alone would not explain which task requirements the students found most challenging according to their language proficiency. The next chapter describes some individual cases to further deepen the understanding of different stages in the cognitive processes of writing.

Chapter 6. Findings from Selected Cases

Tobii Studio turned data from the eye-tracker into a comprehensive information report by identifying different Areas of Interests in order to show the general trends in the integrated writing process behaviours of participants with different language proficiencies. This data, however, did not fully explain or characterize the integrated writing processes of each participant. This chapter, therefore, provides some individual case studies of participants with different levels of reading and writing proficiencies. Although positivists, and even some constructivists, do not believe that the rationale behind the selection of individual cases is important for the purposes of generalization, some researchers would still argue that “generalizability is simply incompatible to develop in-depth understanding” (Power & Gendron, 2015. p. 158). To be clear, the author acknowledges that it is impossible to generalize the findings from the limited sample in this study (N=38). Each participant showed individual differences in how they viewed and processed the reading-to-writing tasks during the eye-recording experiments. In this qualitative chapter, therefore, some cases were purposively chosen among participants with a different combination of reading and writing proficiency levels. Individual cases in this chapter are displayed only for an exploratory purpose, with the main purpose being to use an alternative perspective to illustrate the kind of information that can be gleaned from eye-tracking data in terms of understanding different stages of the cognitive processes during integrated writing. The following cases are chosen for in-depth analysis. The table summarizes their language proficiency levels.

Table 72. Cases shown for in-depth analysis (Language Proficiencies)

ID	Aptis Reading	Aptis Writing	TEAP Writing
ST030	C	B2	B1
ST037	B1	C	B2
ST033	B2	B2	B1
ST053	B1	B1	Below A2

The findings are reported according to three main stages of cognitive processes of writing: (i) Preparing to write, (ii) Translation to write, and (iii) Monitoring and Revising.

6.1 Preparing to Write

This section of the chapter will review the initial process referred to as “preparing to write” according to the Socio-Cognitive Framework for Language Testing (Shaw & Weir, 2007).

6.1.1 Task Representation

The task instruction for the test is as follows:

Your teacher has asked you to write an essay for class using the information below. Describe the situation concerning schools in Greenhill and summarize the main points about the solutions that have been suggested. In your conclusion, say which of the solutions you think would work the best based on the reasons given. You should write about 200 words.

To be successful, test-takers need to understand the required tasks for the essay clearly. Failure to fully understand the purpose of an assigned task often causes them to produce an essay with an inappropriate focus. For example, some test-takers misunderstood the instructions, and wrote an opinion essay about the topic based on their experience and personal knowledge instead of summarizing the most important points suggested in the source texts and graphs. Gaze plots revealed that some students viewed the Task Instruction area more carefully than others. Failure to read instructions is a general indication of low academic achievement. In

other words, poor academic performers in general are also likely to have lower language proficiency and to exhibit this kind of approach to a task as reported in the regression analysis in the section 4.4.5 as well as in the survey (Table 71). Thus, the assumption was made that the participants' performance in the tests would be low if the Gaze plots were not detected in the first 5 minutes of recordings. The cases of four participants seemed to validate that participants with low TEAP scores struggled to understand the requirements for the writing test.

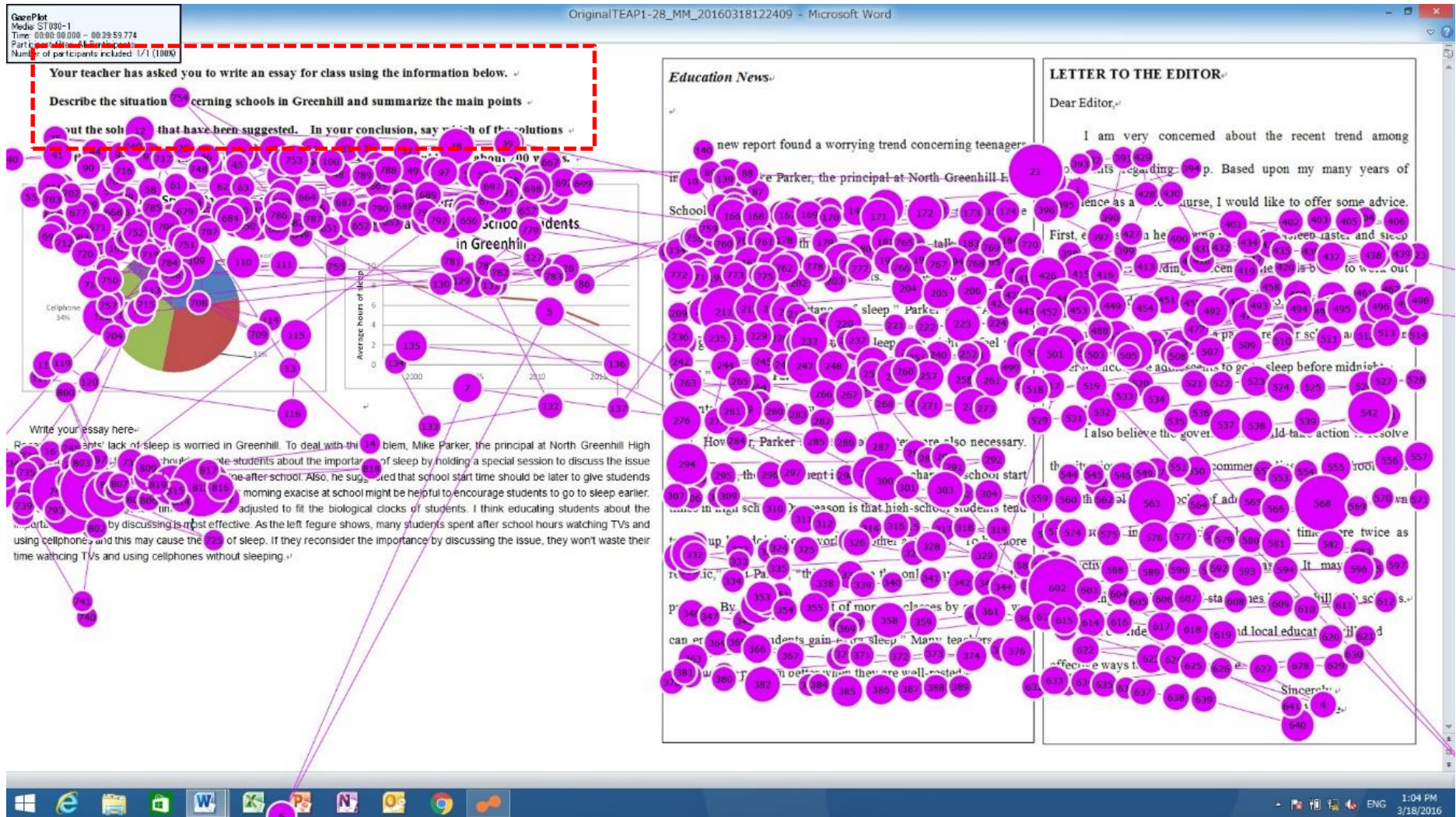
[Case 1]

Figure 15 and Figure 16 are Gaze Plots captured after five minutes' recording of two participants (ST030 and ST037 respectively) during Task 1. ST037 has stronger writing proficiency (B2) than ST030 (B1). The two gaze-plots taken during the first five minutes of recording give a better understanding of which language proficiency, reading or writing, has played a key role in their integrated writing. The red dotted line square indicates the area of the Task Instruction for the essay. ST030 has gone straight into reading information on the graphs and source texts in the first five minutes of recording. On the other hand, ST037 began processing by reading the Task Instruction. This may be an indication that ST037 has a better chance of understanding the required tasks for the essay. Although one could argue that the task instruction is straightforward and only needs the briefest of reads to understand it, most of the participants in the current study were not familiar with the TEAP test formats to begin with. Spending the first five minutes on the Task Instruction may be an indication of how successful they were on the test.

[Case 2]

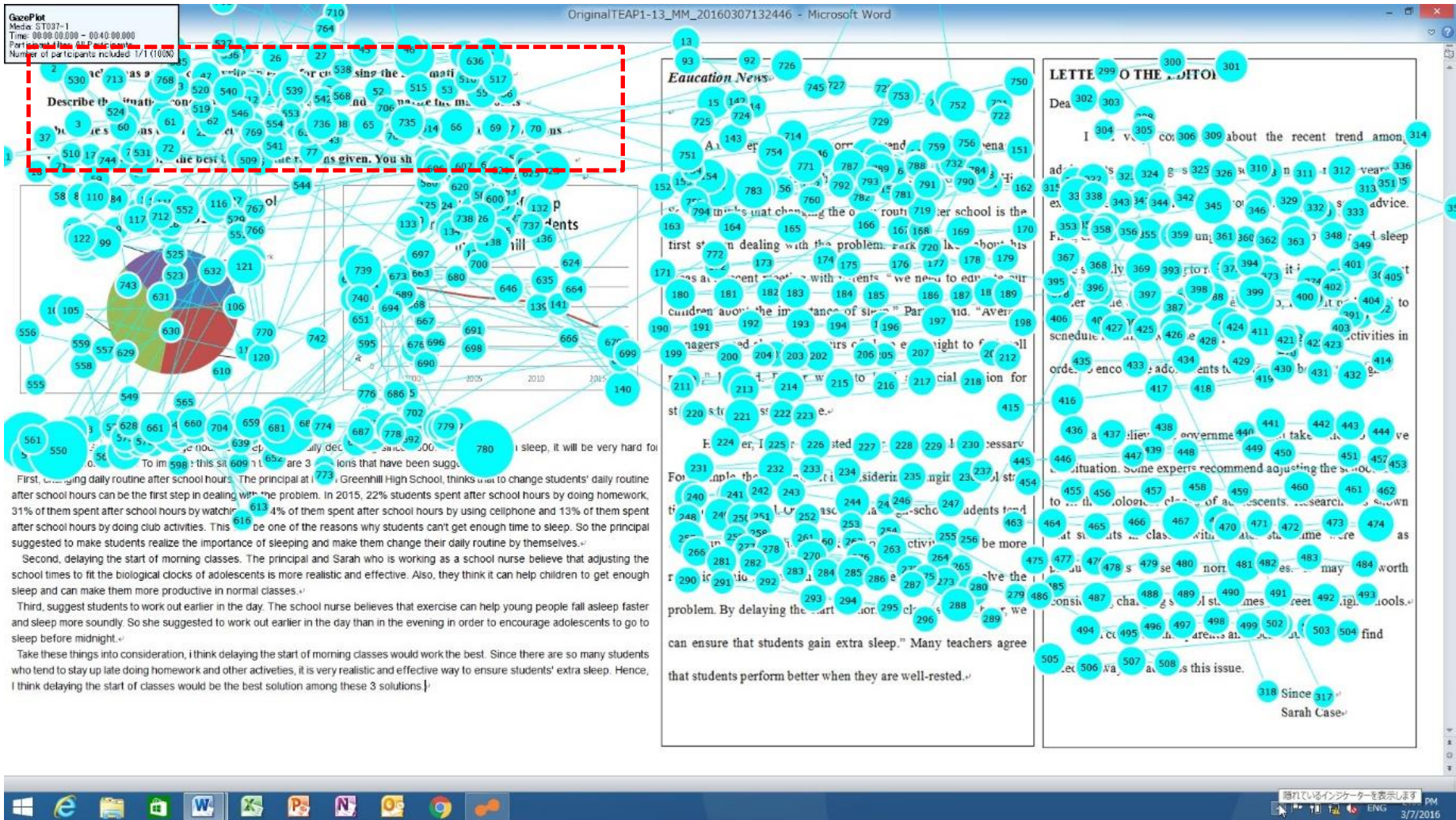
Figure 17 and Figure 18 are Gaze Plots captured after five minutes' recording of two participants (ST053 and ST033 respectively) during Task 1. As shown in the gaze plots, both participants seem to have viewed the task instruction to some degree during their first minutes

of recording in the test, but did not appear to have read the instruction sentences in order. In both cases, it was evident that they started to compose their essay within five minutes of eye-recording before they even tried to read the source text. This may be an indication that both participants were struggling to understand the task requirement for the essay.



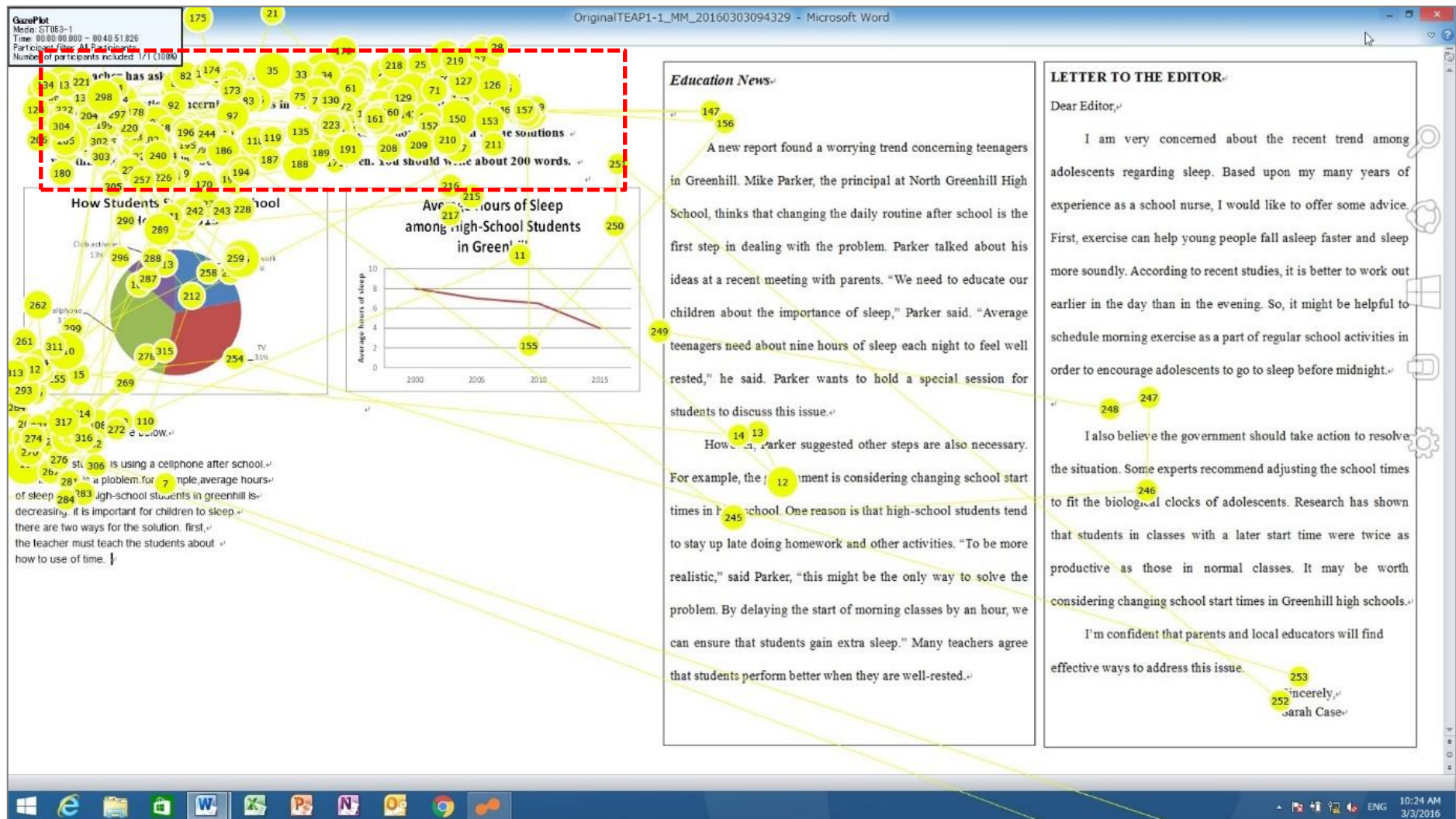
[Case 1]

Figure 15. Gaze-plot for ST030 [Task 1] ST030 (Aptis Reading C / Aptis Writing B2 / TEAP B1, N=1)



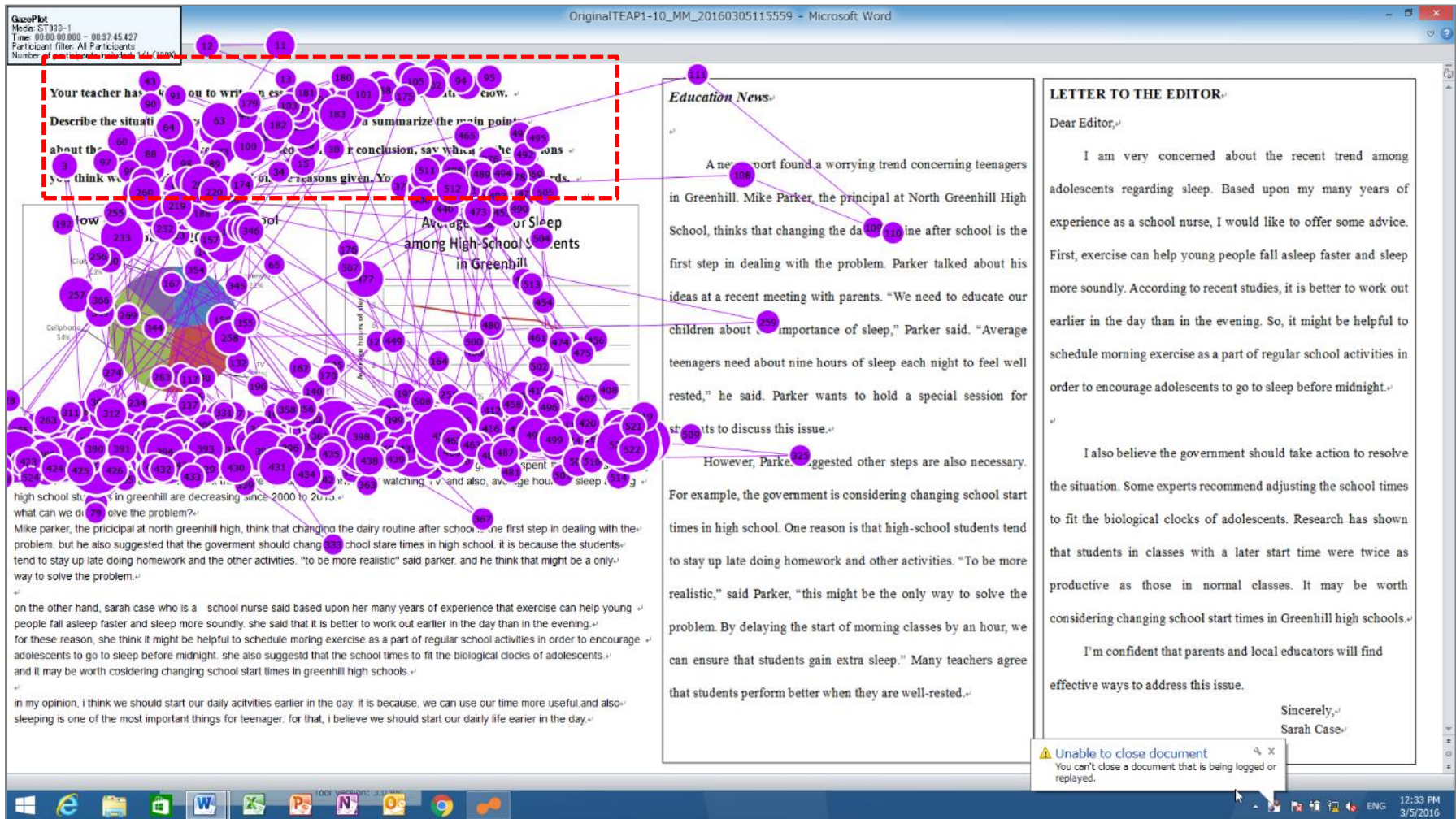
[Case 1]

Figure 16. Gaze-plot for ST037 [Task 1] ST037 (Aptis Reading B1 / Aptis Writing C / TEAP B2, N=1)



[Case 2]

Figure 17. Gaze-plot for ST053, First 5 min. [Task 1] ST053 (Aptis Reading B1/Aptis Writing B / TEAP A2-Below, N=1)



[Case 2]

Figure 18. Gaze-plot for ST033, First 5 min. [Task 1] ST033 (Aptis Reading B2 /Aptis Writing B2 /TEAP B1, N=1)

6.1.2 Macro-Planning

In the macro-planning stage, writers are concerned with the planning of the essay, which includes identifying the main idea (topic sentences), the writing styles (discourse, genre), structure of the essay (number of paragraphs), and registers (tense, use of first, second and third person pronouns). Some of these questions are related to test-taking strategies, particularly in determining how the writers chose a language format for the essay. This section of the chapter looks into this in more detail by showing some of the reading patterns followed by the same four participants as above.

Use of Source Texts (AOI Switches):

The selected participants' AOI switches were analysed in order to study their eye-movements during the first ten minutes of recording in each paragraph of the source texts. The analysis was done in an effort to study their reading behaviours. Here are two examples of AOI switches for ST030 and ST037. The cases might exemplify how their language reading patterns and behaviours (not reading proficiency levels) played a role in micro-planning. From the data, it was evident that both the students visited P1 (Paragraph 1) the most. Furthermore, there was an indication that ST030 did not read P4 (Paragraph 4) after P3 (Paragraph 3). In other words, ST030 skipped reading Paragraph 4 and did not read every paragraph in order during the first ten minutes of recording.

Table 73. Number of AOI Switches for 10 minutes of recording

ST037 (Aptis Reading B1 / Aptis Writing C / TEAP Writing B2)

ST037-1	P1	P2	P3	P4	PX*	# of AOI Switches
P1	0	2	1	0	32	35
P2	0	0	1	5	5	11
P3	2	0	0	1	5	8
P4	0	4	1	0	2	7
PX	33	5	5	1	0	44

Note: N=1. PX refers to the space other than Paragraph 1 (P1), Paragraph 2 (P2), Paragraph 3 (P3) and Paragraph 4 (P4)

ST030 (Aptis Reading C / Aptis Writing B2 / TEAP Writing B1)

ST030-1	P1	P2	P3	P4	PX	# of AOI Switches
P1	0	5	6	1	28	40
P2	3	0	1	0	7	11
P3	5	0	0	0	8	13
P4	0	1	0	0	6	7
PX	32	5	5	6	0	48

Note: N=1. PX refers to the space other than Paragraph 1 (P1), Paragraph 2 (P2), Paragraph 3 (P3) and Paragraph 4 (P4)

ST033 (Aptis Reading B2 / Aptis Writing B2 / TEAP Writing B1)

ST033-1	P1	P2	P3	P4	PX	# of AOI Switches
P1	0	2	0	0	42	44
P2	1	0	0	0	3	4
P3	0	0	0	0	1	1
P4	0	0	0	0	0	0
PX	43	2	1	0	0	46

Note: N=1. PX refers to the space other than Paragraph 1 (P1), Paragraph 2 (P2), Paragraph 3 (P3) and Paragraph 4 (P4)

ST053 (Aptis Reading B1 / Aptis Writing B1 / TEAP Writing A2-Below)

ST053-1	P1	P2	P3	P4	PX	# of AOI Switches
P1	0	0	0	0	0	0
P2	0	0	0	1	2	3
P3	0	0	0	0	0	0
P4	0	0	0	0	2	2
PX	0	3	0	1	0	4

Note: N=1. PX refers to the space other than Paragraph 1 (P1), Paragraph 2 (P2), Paragraph 3 (P3) and Paragraph 4 (P4)

The number of AOI switches for some other participants are shown in Table 74. There was a tendency for the participants to skip or jump around the paragraphs during the first ten minutes of recordings, especially in Paragraph 3 and Paragraph 4.

Table 74. Skipped Paragraphs during the First 10 Minutes

ID	Eye-tracking (Task 1)	Eye-tracking (Task 2)
ST037 (TEAP B2)	No Skipping	P3
ST030 (TEAP B1)	P4	P4
ST033 (TEAP B1)	P3, P4	P3, P4
ST053 (TEAP A2 below)	P2, P3, P4	P2, P3, P4

The participant with the highest proficiency level (ST037/B2 level) seemed to have read each paragraph of the source texts in order, while other students did not view some sections of the texts during the first ten minutes of the tests. ST033 and ST053 were found to begin writing their essays before they had finished reading the entire text.

Uses of Source Texts (Saccades)

Saccades and Regressions may help understand how readers process text. It is possible to export from *Tobii Studio* the binary numbers of either "0" or "1" in each cell to show where a participant has fixated his or her eyes in certain areas of interest such as Paragraph 1. By looking at the direction (positive or negative) of two fixation points in time, the reader's text processing could be defined either as a saccade (forward jumps) or regression (backward jumps). Return sweeps refers to eye-movements back to the next line. In this study, when the regressions were more than one-third of the line (i.e., 60 spaces), the author classified them as return sweeps. The sums of saccades and regressions were calculated to find the average numbers of saccades and regressions. Table 75 shows the number of saccades, regressions and return sweeps in Paragraph 1 of Task 1 for the same participants whose number of AOI switches during the first ten minutes of recordings were given earlier. The data is intended to show patterns in how the participants read the source texts.

Table 75. Saccades, Regression, Return, and Sweeps found in Paragraph 1 (Task1)

ID	Number of Saccades	Number of Regression	Number of Return Sweep	Average Saccades (letters)	Average Regression (letters)
ST037 (TEAP B2)	202	119	36	13.03	22.74
ST030 (TEAP B1)	239	121	17	6.87	12.29
ST033 (TEAP B1)	168	102	25	9.18	14.28
ST053 (TEAP A2-below)	42	24	4	6.93	8.29

From the findings, it can be said that ST037 read Paragraph 1 the fastest, as shown in the average saccades. To investigate how successfully they looked for the key information from the source texts, the individual responses to the questionnaires are carefully reviewed below.

Uses of Source Texts

The use of source texts was queried as: (1) Did you re-read the parts that you thought were important to include in your essay? (Q6_1); (2) How did you decide which information to include from the texts? (Q6_2); and (3) How did you connect the ideas from the texts? (Q6_3).

Table 77 shows how these individual participants responded.

Table 76. Questionnaire answers on the use of source texts

Student	Q6_1	Q6_2	Q6_3
ST037 (TEAP B2)	Yes	Re-reading a whole text	By borrowing some words
ST030 (TEAP B1)	Yes	Going back to a specific paragraph	By copying the sentences from source text
ST033 (TEAP B1)	Yes	Going back to a specific paragraph	By copying the sentences from source text
ST053 (TEAP A2-below)	No	By memory	By borrowing some words

It appears ST053 used some keywords from the texts and wrote an independent essay because she wrote the essay based on memory. ST030 and ST033 relied on the source texts and even

copied sentences from the source texts in specific paragraphs of their essay. ST037 re-read the entire text and wrote an essay based on the key information.

Uses of Information from graphs

To understand how much of the information the participants took from graphs, the following survey questions were examined.

Table 77. Questionnaire answers on the use of information from graphs

Student	Q5_1 Titles	Q5_2 Understanding values on units	Q5_3 Reading X and Y Axis	Q5_4 Interpreting information on graphs	Q5_5 Summarizing the trends of graphs
ST037 (TEAP B2)	Very Easy	Easy	Very Easy	Very East	Easy
ST030 (TEAP B1)	Neutral	Easy	Neutral	Easy	Neutral
ST033 (TEAP B1)	Neutral	Neutral	Easy	Easy	Neutral
ST053 (TEAP A2- below)	Neutral	Neutral	Neutral	Difficult	Difficult

It is evident that, of these cases, ST053, with the lowest writing proficiency (Below-A2 on the TEAP score) struggled to interpret information from the graphs and to summarize the trends in the graphs.

Familiarity with Graph Types

Earlier, the findings from the questionnaires in respect to the participants' familiarity with graphs did not find any significant difference (Q5-1, Q5-2, Q5-3, Q5-4). Similarly, the participants seemed to agree that they were familiar with all types of graphs presented in the tests. Interpretation of the information from graphs was not of much concern to them.

Table 78. Questionnaire Answers on the Familiarity of Graph Types

Student (R/W)	Q2_1 Line graph	Q2_2 Pie Chart (1)	Q2_3 Bar Graph	Q2_4 Pie Chart2	Q2_5 Ability to interpret graph info.
ST037 (TEAP B2)	Very Easy	Very Easy	Very Easy	Very Easy	Average
ST030 (TEAP B1)	Easy	Very Easy	Very Easy	Very Easy	Average
ST033 (TEAP B1)	Very Easy	Very Easy	Very Easy	Very Easy	Average
ST053 (TEAP A2- below)	Very Easy	Very Easy	Very Easy	Very Easy	Average

The questionnaire shows that the students had more difficulty with summarizing the graph trends in English, presumably because they were unfamiliar with some of the English words needed to describe the graph trends (Q5-5).

6.1.3. Summary of the Preparing to Write stage

The quantitative analysis using eye-recording data from the first ten minutes was helpful for an understanding the role of language proficiency in integrated writing tasks. Using questionnaire results, AOI switches, saccades of some individual cases, how participants decided to include the source text and information from graphs were reported. For example, the AOI switches of individual students indicated a tendency for participants with higher reading and writing proficiencies to read the paragraphs of source texts in order.

6.2 Translating and Writing

6.2.1 Examples of students' writing

The second stage of cognitive writing takes place when a writer starts composing an essay. For integrated writing, such as in the TEAP test, writers were asked to (1) describe the situations using information from graphs, (2) summarize the main points described in the text, (3) state in a conclusion which solution might work best, and (4) give the reasons for the choice made.

The eye-tracking tests offered insights into how participants gazed at the multiple texts and information from graphs and helped understand their behaviours during the reading-into-writing tasks. In addition to the statistical analysis, the written outputs generated by the participants can also give some valuable references for their cognitive processes.

There were six main tasks required for writing the essay: (1) Reading the texts (2) Understanding the graphs, (3) Describing the situation concerning food waste in Greenfield by interpreting the information from the graphs, (4) Summarizing the main points about solutions that have been suggested in the texts, (5) Choosing the best solution and giving the reasons for this choice, and (6) Writing about 200 words in all. By looking at the written outputs at different language levels, it became clear that the students struggled with different aspects of these tasks required to write the integrated essay. In this section, two comparisons are made to showcase the roles of reading and writing skills in integrated writing by looking at the students' written essays. Here is the sample essay answer for Task B, Task 1.

[Model Answer for Task 1]

Teenagers in Greenhill are getting less sleep over the years. According to the statistics, the average hours of sleep among high-school students dropped by half from 8 hours to 4 hours between 2000 and 2015. The survey in 2015 indicates that 34 percent of after-school hours were spent on cell phones, followed by 31 percent on TV, 22 percent on homework and 15 percent on club activities.

Mike Parker, the Principal of North Greenhill High School, made two suggestions. First, he suggested educating his students and their parents about the importance of sleep. The second suggestion was to delay the school start times in high schools. This way, he said we can ensure the students get enough sleep every day.

The idea of changing the school times to fit the “biological clocks of adolescents” was also echoed by Sarah Case, who is a school nurse. In addition, she said introducing morning exercises might help the students fall asleep faster at night and sleep more soundly.

In my opinion, having morning exercises would be the best solution. I believe delaying the start of morning classes would only cause the students to stay up much later than they already do. This is not a healthy way to solve the problem. (208 words)

[Case 1: ST030 and ST037]

Extract 1 was written by Participant ST030, who was rated as B1 on the TEAP rating scale and Extract 2 was written by Participant ST037, who had a B2 writing proficiency. The written outputs are reproduced here exactly as they were written. To compare both texts, three types of notes were added to the texts. The parts of texts that are highlighted in grey show where the writers made mistakes in spellings. Red fonts indicate the words or phrases that were borrowed directly from the source texts. The underlined text indicates where the writers attempted to give credit to a person or information in the source text.

Note:

- (1) Misspelled words
- (2) **Verbatim source use**
- (3) A part of speech that gives credit to a person or information in the source text

Extract 1. ST030 (Aptis Reading C/ Aptis Writing B2 / TEAP Writing B1)

Recently, students' lack of sleep is worried in Greenhill. To deal with this problem, Mike Parker, the principal at North Greenhill High School suggested that teachers should **educate students about the importance of sleep by holding a special session to discuss the issue** to encourage students to **change the daily routine after school**. Also, he suggested that school start time should be later

to give students extra sleep. Sarah Case, a school nurse, said that morning exercise at school might be helpful to encourage students to go to sleep earlier. She also believed that school times should be adjusted to fit the biological clocks of students. I think educating students about the importance of sleep by discussing is most effective. As the left figure shows, many students spent after school hours watching TVs and using cellphones and this may cause the lack of sleep. If they reconsider the importance by discussing the issue, they won't waste their time watching TVs and using cellphones without sleeping. (165 words)

Extract 2. ST037 (Aptis Reading B1 / Aptis Writing C / TEAP Writing B2)

In the school in Greenhill, the average hours of sleep is gradually decreasing since 2000. Without enough sleep, it will be very hard for students to perform better. To improve this situation there are 3 solutions that have been suggested.

First, changing daily routine after school hours. The principal at North Greenhill High School, thinks that to change students' daily routine after school hours can be the first step in dealing with the problem. In 2015, 22% students spent after school hours by doing homework, 31% of them spent after school hours by watching TV, 34% of them spent after school hours by using cellphone and 13% of them spent after school hours by doing club activities. This can be one of the reasons why students can't get enough time to sleep. So the principal suggested to make students realize the importance of sleeping and make them change their daily routine by themselves.

Second, delaying the start of morning classes. The principal and Sarah who is working as a school nurse believe that adjusting the school times to fit the biological clocks of adolescents is more realistic and effective. Also, they think it can help children to get enough sleep and can make them more productive in normal classes.

Third, suggest students to work out earlier in the day. The school nurse believes that exercise can help young people fall asleep faster and sleep more soundly. So she suggested to work out earlier in the day than in the evening in order to encourage adolescents to go to sleep before midnight.

Take these things into consideration, i think delaying the start of morning classes would work the best. Since there are so many students who tend to stay up late doing homework and other activities, it is very realistic and effective way to ensure students' extra sleep. Hence, I think delaying the start of classes would be the best solution among these 3 solutions. (324 words)

Misspelling of words was more apparent in Extract 1 (ST030) than in Extract 2 (ST037). Both students seemed to have borrowed several strings of words from the source texts. The patchy

work hindered their ability to paraphrase sentences. Participant ST030, however, showed a greater lexical knowledge, such as in the phrase, “lack of sleep”, which was not provided in the source text. On the other hand, Participant ST037 managed to write her composition using the words that appeared in the source texts. In addition, both referred to information from the graphs in their essays; however, ST030 did not correctly transform the information into written form. ST037 wrote a paragraph describing the trends from both graphs, using these as a rationale or support for the topic of the essay. Finally, the essay written by ST037 in Extract 2 illustrated her ability to compose an essay with a good organizational structure. The five-paragraph essay was layered in a logical sequence. Thus, the essay was more coherent to follow. This participant obviously understood what constitutes a paragraph and the basic structure of what an academic essay would look like. Overall, it can be said that ST037 had longer sentence lengths and synthetic complexity. For this reason, the essay was marked at B2 level by the two raters.

[Case 2: ST033 and ST053]

Extract 3. ST053 (Aptis Reading B1/ Aptis Writing B1 / TEAP Writing A2-Below)

most of the student is using a cellphone after school.
for this, there is a ploblem.for example,average hours
of sleep among high-school students in greenhill is
decreasing. it is important for children to sleep.
the way for the solution is (40words)

Extract 4. ST033 (Aptis Reading B2 / Aptis Writing B2 / TEAP Writing B1)

schools in Greenhill are facing a difficult problems. acording to the data, 22% of students in greenhill spent their after school by homework. but, most of the students spent thier time by using a cellphone, or watching TV. and also, average hours of sleep among high school students in greenhill are decreasing since 2000 to 2015. what can we do to solve the problem?
Mike parker, the pricipal at north greenhill high, think that **changing the dairy routine after school is the first step in dealing with the problem.** but he also suggested that the government should changing school stare times in high school. it is because the **students**

tend to stay up late doing homework and the other activities. "to be more realistic" said parker. and he think that might be a only way to solve the problem.

on the other hand, sarah case who is a school nurse said based upon her many years of experience that exercise can help young people fall asleep faster and sleep more soundly. she said that it is better to work out earlier in the day than in the evening.

for these reason, she think it might be helpful to schedule moring exercise as a part of regular school activities in order to encourage adolescents to go to sleep before midnight. she also suggestd that the school times to fit the biological clocks of adolescents.

and it may be worth cosidering changing school start times in greenhill high schools.

in my opinion, i think we should start our daily acitvities earlier in the day. it is because, we can use our time more useful.and also sleeping is one of the most important things for teenager. for that, i believe we should start our dairly life earier in the day. (297 words)

As shown above, both Extract 3 and Extract 4 had some mechanical errors, not only in terms of the spelling of the words but also in punctuation. Most notably, Extract 3 was very short, indicating ST053 did not have competency in writing. ST053 claimed she found typing keywords very difficult (Q1_6). The essay length, therefore, may have been affected by her experience with typing on a computer.

ST033's essay was longer, but with longer strings of words and in some cases, an entire sentence taken directly from the original source texts. Although paraphrasing skills did not seem to be fully developed by the participant (ST033), he was able to fulfil most of the task requirements such as summarizing the main solutions described in the text and stating which solution he thought would work best, and the rationale behind the choice he made. In short, what made the essay written by the participant ST033 far better than ST053 was the ability to cite references while copying part of the information from the source texts.

6.2.2 Test-taking strategies

To understand what kind of strategies were applied by the participants throughout the reading-into-writing processes, a section of the questionnaires focused on test-taking strategies.

Participants at the B2 level tended to borrow directly from source texts as if they were their ideas. Table 80 shows the survey answers, revealing those four participants' responses to the survey questions on the response format (Q7_1, Q7_2), and translation for writing (Q8_1, Q8_2, Q8_3, and Q8_4).

Table 79. Questionnaire on Response Format and Translation for Writing

Questionnaire / Participants (TEAP Ratings)	ST037 (B2)	ST030 (B1)	ST033 (B1)	ST053 (A2 below)
[Micro-planning: Response format] Did you make an outline BEFORE writing your essay? (Q7_1)	Yes	No	Yes	Yes
Did you decide how many paragraphs there should be in your essay? (Q7_2)	Yes	No	Don't remember	No
[Fulfilling the task requirements] How well do you think you have described the situation? (Q8_1)	Good	Good	Good	Average
How well do you think you have summarized the main points? (Q8_2)	Good	Average	Average	Average
How well do you think you have stated which solution would work best? (Q8_3)	Good	Average	Average	Poor
How well do you think you have given the reasons for the choice made? (Q8_4)	Good	Average	Average	Poor

As shown in Table 52, most students did not plan to make any outlines before they began writing the essay (Q7_1). In the questionnaire analysis (Chapter 5), it was also reported that half of the test-takers had no plan as to how many paragraphs they were going to write in the essay (Q7_2). This perhaps reflected their prior knowledge of academic writing. Some participants had written only one paragraph (e.g., Extract 1 and Extract 3). Some participants had some paragraphs, but one paragraph would only consist of a single sentence, or a few sentences without a topic sentence or clear paragraph structure (e.g., Extract 4). Based on this, it was evident that the students' prior knowledge of academic writing made a difference in respect to the format of their written output. In essay composition, Japanese students do not necessarily learn about paragraph writing (i.e. what constitutes a paragraph). As introduced earlier in the review for the TEAP writing (see section 2.8), English composition lessons in

Japanese middle and high schools are often based on translation skills, using some target words and phrases from a Japanese sentence to an English sentence and vice versa. The lack of experience of an academic essay, such as writing a multi-paragraph essay, paraphrasing source texts, and synthesizing information by citing the sources, etc., could have played a role in the outcomes of the TEAP writing test.

6.3 Monitoring and Revising

According to Shaw and Weir (2007), monitoring and revising involves mechanical accuracy, such as checking the correct spelling and grammar at a basic level, as well as the structural and organizational development of paragraphs at a more advanced level. Basic monitoring could be ongoing during essay writing, since some participants stopped at each sentence or paused at the end of a paragraph to correct some spelling and grammatical mistakes. Most basic monitoring and revising, however, was observed to take place in the gaze-plots during the last stage of writing in both eye-tracking tests. This section, therefore, looks at the last five minutes of the gaze-plots as evidence to illustrate the differences between the participants with a different language proficiency.

6.3.1 Gaze-plots for the last five minutes

This section highlights some of the gaze-plots taken during the last five minutes of recordings for Task 1 for selected participants.

Figure 20 shows the gaze spot analysis for ST037 during the last five minutes of the test. The gaze plots showed some evidence that ST037 had reviewed the task instructions during Task 1. Such behaviour could imply that she made an effort to conduct a quick review of the task instructions by visiting the task instruction during the last five minutes. On the contrary, Figure 19, showing the gaze plots of ST030 in the last five minutes, indicates that this participant was

still in the midst of reading into writing tasks during the last five minutes of recording for Task 1.

Figures 21 and 22 show the gaze plots for ST053 and ST033, respectively, during the last five minutes of the test. The gaze plots indicated that their attention was scattered around during the last five minutes.

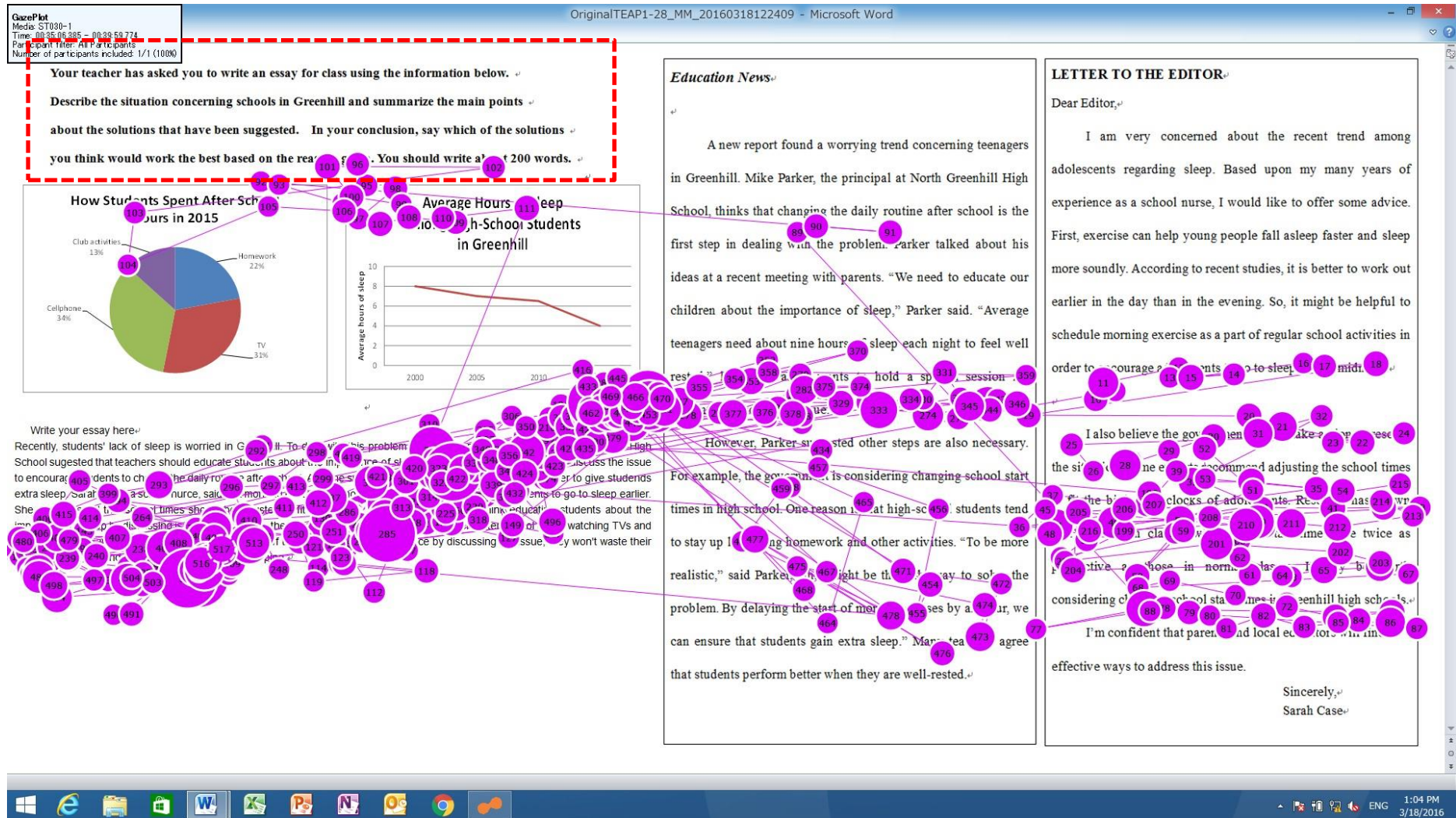


Figure 19. Gaze-plot for ST030, Last 5 min. [Task 1]N=1

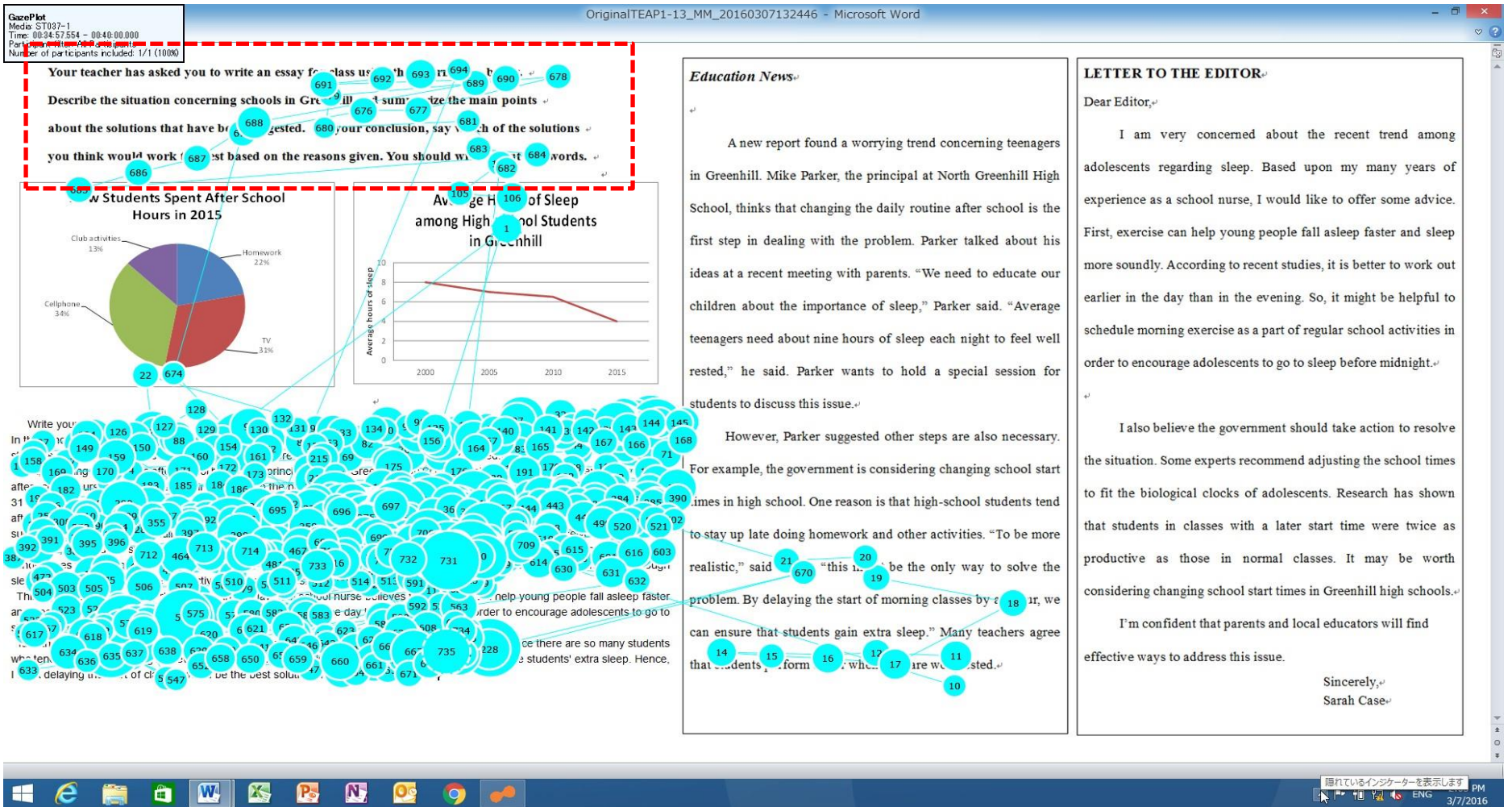


Figure 20. Gaze-plot for ST037, Last 5 minutes [Task 1]N=1

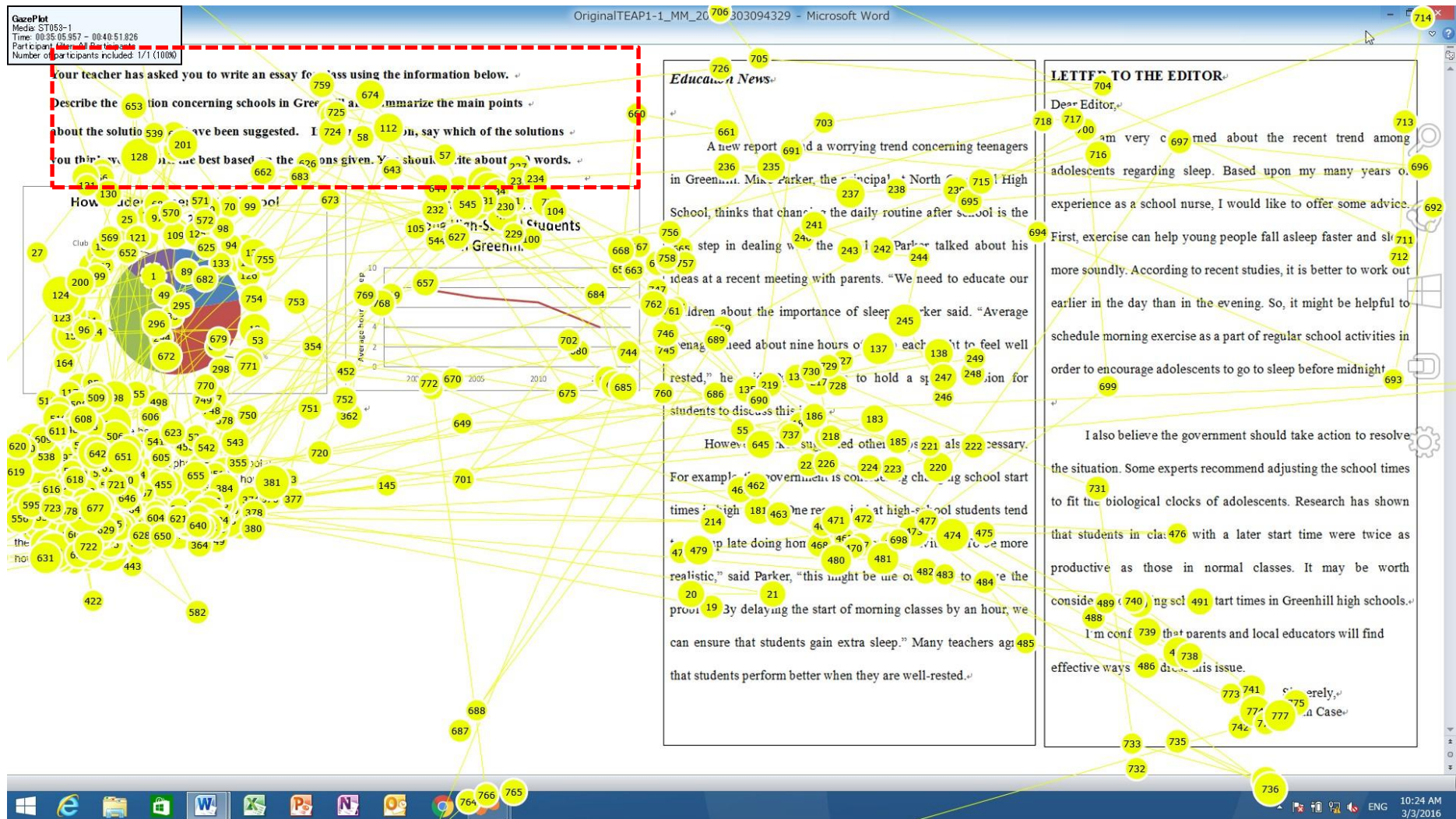


Figure 21. Gaze-plot for ST053, Last 5 min. [Task 1]N=1

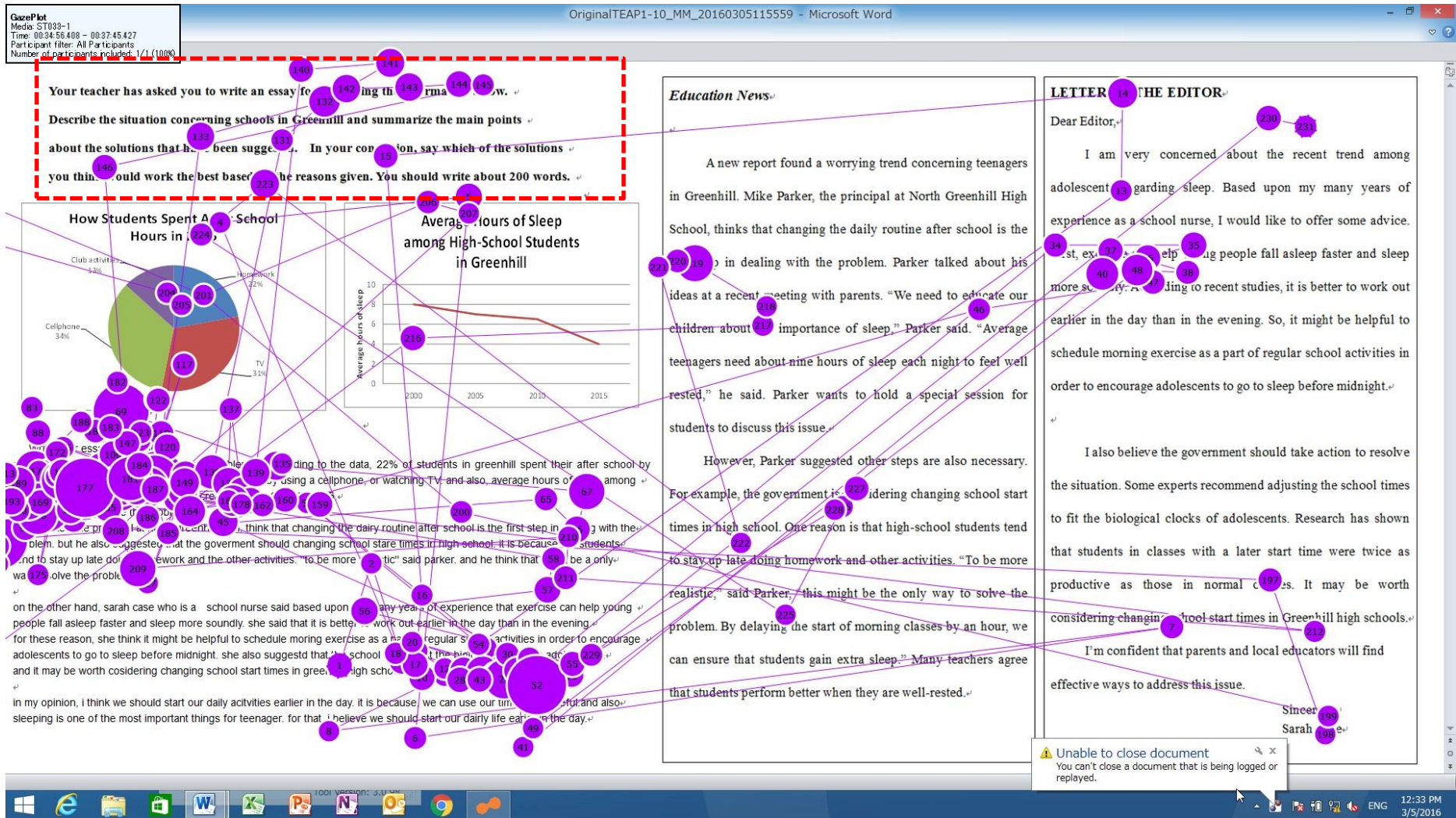


Figure 22. Gaze-plot for ST033, Last 5 minutes [Task 1] N=1

The gaze plots taken from the last five minutes of recording exemplified the participants' behaviours at different language levels. Some of the participants reviewed their written output entirely while other participants missed an opportunity to examine their essay in the last five minutes. They kept writing until the very last minute, and they only monitored their grammatical and spelling mistakes as they wrote at a basic level. Overall, the last five minutes of recording could be a good reflection of the monitoring and revising processes as far as these selected cases are concerned.

6.4 Comparing the performance of the same participant between Task 1 and Task 2

6.4.1 Evidence from the Gaze-plots of the two tests

As explained earlier in the methodology section, the participants took a similar test twice for the purpose of the eye-recordings. Both tests were similar in length and format. This section compares the gaze-plots taken from the first and last five minutes of eye-recordings of the participants selected in this chapter (i.e., ST030, ST033, ST037, and ST053). The purpose of this analysis was to explore whether a test-taker showed a similar pattern in their writing processes every time they took a similar test.

The two gaze-plot figures below are taken from the first and last five minutes of the tests from participant ST033. The lengths of the essays that the participants produced during Task 1 and Task 2 were very similar, but the participant clearly showed a better performance during Task 2. The gaze-plots for the first five minutes of recording on Task 1 indicated that the participant had spent the first five minutes trying to compose an essay immediately after reviewing the

Task Instruction without reading the source texts any further. On Task 2, however, there was evidence that the same participant was skimming the source texts extensively.

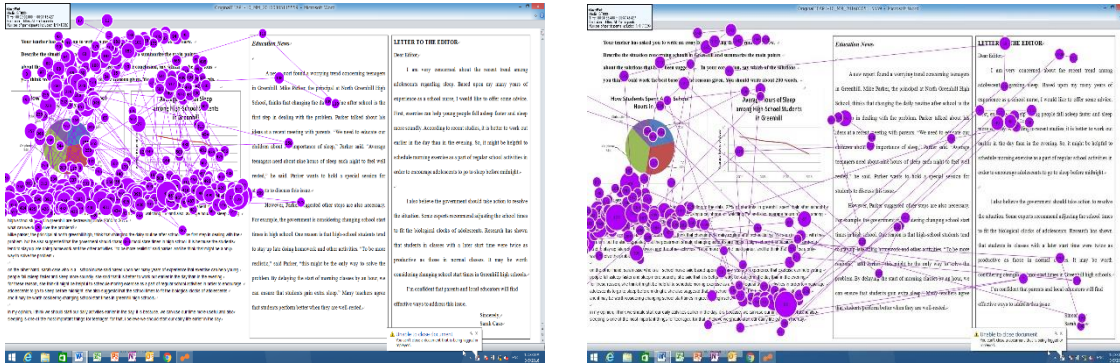


Figure 23. ST033's Gaze-Plots for the First 5 and the Last 5 minutes during Task 1

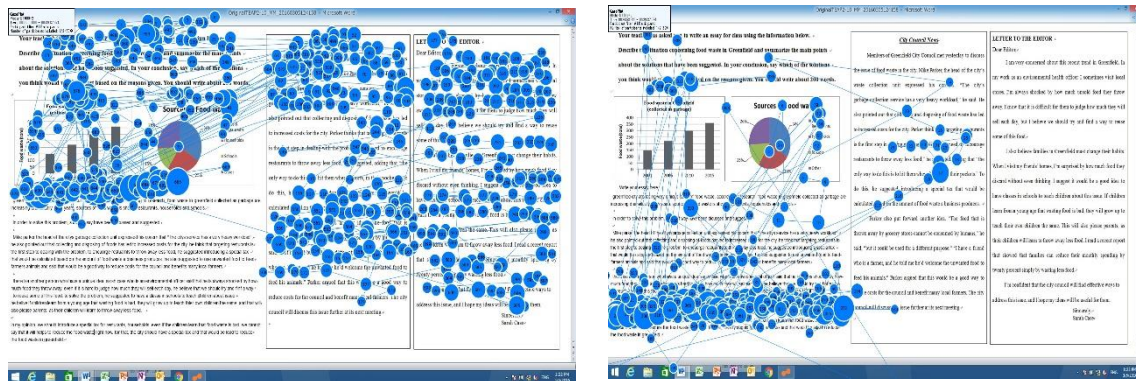


Figure 24. ST033's Gaze-Plots for the First 5 and the Last 5 minutes during Task 2

A similar pattern was found for the participant ST053. From a comparison of the gaze-plots for Task 1 and Task 2, the participant seems to have made more effort to refer to the source texts during Task 2 than in Task 1. This difference did not lead to an improved score on Task 2 in this case, however.

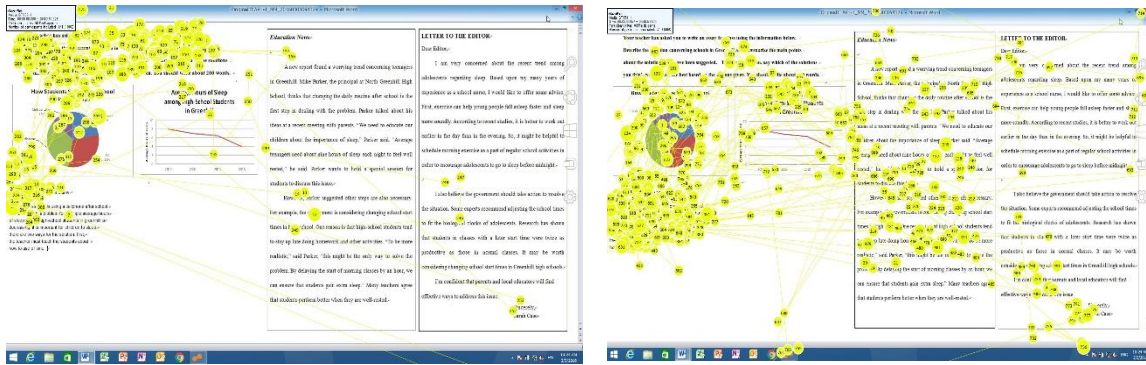


Figure 25. ST053's Gaze-Plots for the First 5 and the Last 5 minutes during Task 1

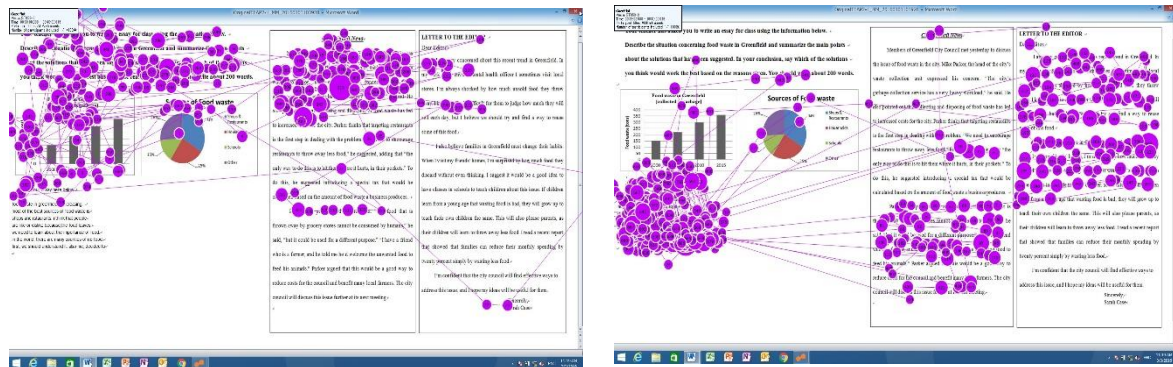


Figure 26. ST053's Gaze-Plots for the First 5 and the Last 5 minutes during Task 2

Figures 27 and 28 below are extracted from the gaze-plots for the first and last five minutes of the participant ST030. Comparing the gaze-plots during the two tests, it is evident that the patterns were very similar for the first and last five minutes of the recordings. Although ST030 had a higher proficiency, the gaze-plots showed that she did not read sentences in order during

the first five minutes. During the last five minutes, the participant seemed to have reviewed her own essays by revisiting some parts of the paragraphs.

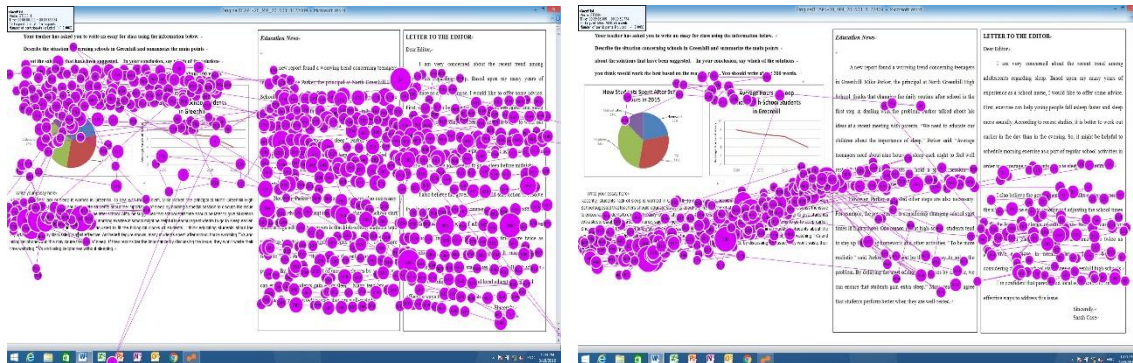


Figure 27. ST030's Gaze-Plots for the First 5 and the Last 5 minutes during Task 1

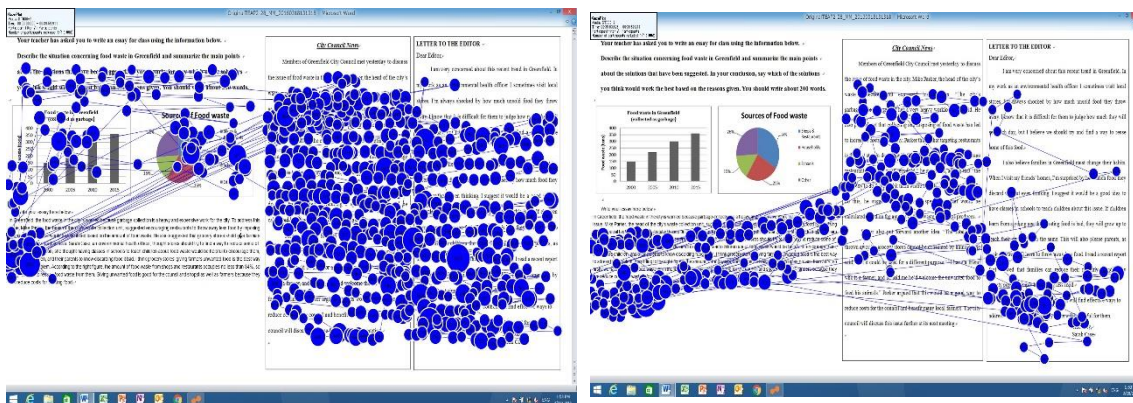


Figure 28. ST033's Gaze-Plots for the First 5 and the Last 5 minutes during Task 2

The following gaze plots are taken from the participant ST037, who was rated as B2 level on the TEAP rating scale. She also showed a very similar pattern in the first and last five minutes of eye-recordings. What impressed the most was that she thoroughly reviewed her written outputs during the last five minutes of recording, which is an indication of monitoring processes taken place. Also, there was a sign that the Task Instruction was thoroughly examined in the first five minutes in both tests.

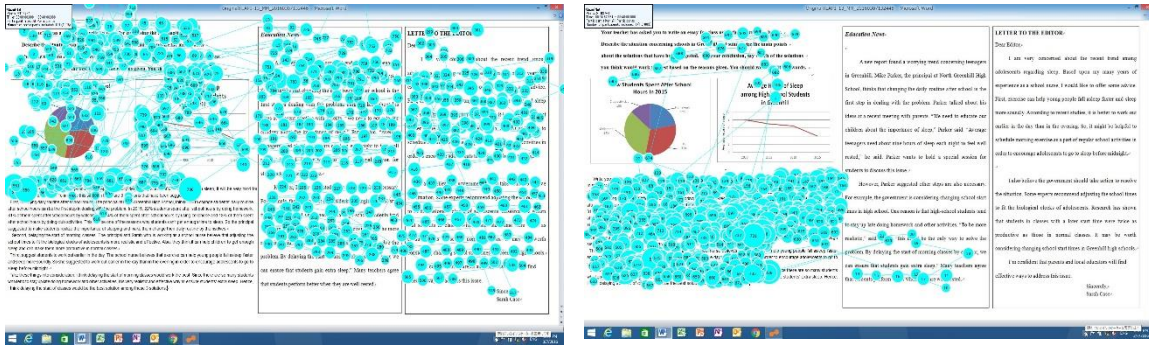


Figure 29. ST037's Gaze-Plots for the First 5 and the Last 5 minutes during Task 1

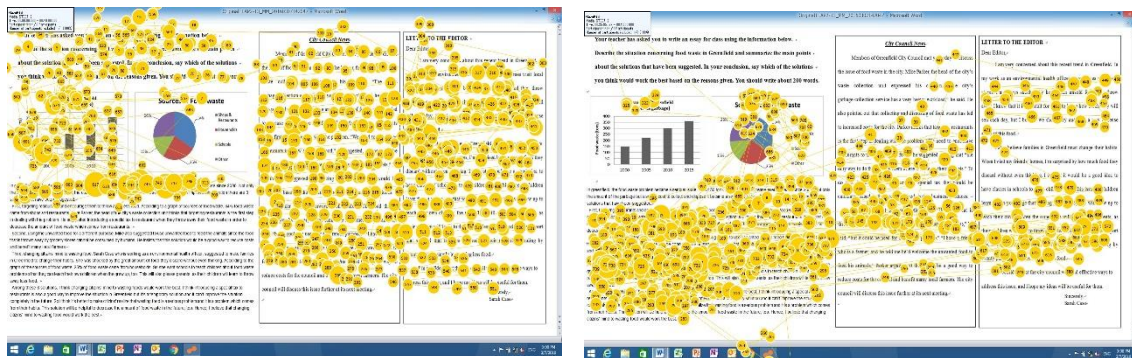


Figure 30. ST037's Gaze-Plots for the First 5 and the Last 5 minutes during Task 2

6.4.2 Evidence from the written outputs of the two tests

The essays produced by the participants across the two tests were very similar in length and did not seem to have many differences regarding the style of their writings and the lexical ranges and accuracy. Since the written outputs looked very similar it was easy to match essays produced by the same writer. Here are the two essays written by participant ST053.

Table 80. Written outputs by ST053

Task 1	Task 2
<p>most of the student is using a cellphone after school.</p> <p>for this, there is a ploblem.for example,average hours</p> <p>of sleep among high-school students in greenhill is</p> <p>decreasing. it is important for children to sleep.</p> <p>the way for the solution is</p>	<p>food waste in greenfield is increasing.</p> <p>most of the best sources of food waste is</p> <p>shops and rstaurants. i think that people</p> <p>are like or dislike.because,the food leaves.</p> <p>we need to learn about the importance of food.</p> <p>in the world, there are many counties of no food.</p> <p>first, we should understand it. also me, decided to</p>

The participant has written an independent essay on the given topic for both tests. Based on the written outputs, one can suspect she has some difficulties with typing the letters on the keyboard. Given the fact that participants had at least three years of English education from junior high-school, they should know that every sentence starts with a capital letter. It may be, however, that they did not know which key (Shift) on the keyboard to press to make a small letter into a capital letter. All the sentences were written in the present form or present progressive. The essay was almost entirely copied from the input text and the number of words was too short (fewer than 50 words). Also, the essay required considerable effort to see the connections as it had no logical paragraph structure, and the sentences had no clear progression. Thus, two raters marked the essay as Below-A2 level.

The following are essays written by participant ST033. The essays are very similar in respect to the language styles, and the mistakes the participant made in Task 1 reoccurred in Task 2.

For example, the student seems to struggle with a subject-verb agreement, which was frequently incorrect in both essays, although those errors did not impede communication of the messages. Also, the student does not appear to know what constitutes a paragraph. The piece has many paragraphs that are not logically developed, although she knows some discourse markers, such as “on the other hand,” and “in my opinions.” Most importantly, both essays showed extensive verbatim source use. Accordingly, the essay was marked as at the B1 level by the two raters.

Table 81. Written outputs by SST033

Task 1	Task 2
<p>schools in Greenhill are facing a difficult problems. acording to the data, 22% of students in greenhill spent their after school by homework. but, most of the students spent thier time by using a cellphone, or watching TV. and also, average hours of sleep among high school students in greenhill are decreasing since 2000 to 2015.</p> <p>what can we do to solve the problem?</p> <p>Mike parker, the pricipal at north greenhill high, think that changing the dairy routine after school is the first step in dealing with the problem. but he also suggested that the goverment should changing school stare times in high school. it is because the students tend to stay up late doing homework and the other activities. "to be more realistic" said parker. and he think that might be a only way to solve the problem.</p> <p>on the other hand, sarah case who is a school nurse said based upon her many years of experience that exercise can help young</p>	<p>greenfiled city are facing very difiicult issue of food waste. according to research, food waste in greenfield collected as garbage are increasing dramaitcally in 15 years. sources of food waste is shops, restaurants, households and schools.</p> <p>in order to solve this problem, a lots of way have been dicussed and suggested.</p> <p>Mike parker the head of the citys garbage collection unit expressed his cocern that " the citys service has a very heavy workload".</p> <p>he also pointed out that collecting and disposing of foods has led to increased costs for the city. he think that targeting resturants is the first step to dealing with the problem. to encuorge restuarnts to throw away less food, he suggested introducing a special tax that would be calculated based on the amount of food waste a business produces. he also suggested to use unwanted food to feed</p>

<p>people fall asleep faster and sleep more soundly. she said that it is better to work out earlier in the day than in the evening. for these reason, she think it might be helpful to schedule moring exercise as a part of regular school activities in order to encourage adolescents to go to sleep before midnight. she also suggestd that the school times to fit the biological clocks of adolescents.</p> <p>and it may be worth cosidering changing school start times in greenhill high schools.</p> <p>in my opinion, i think we should start our daily acitivities earlier in the day. it is because, we can use our time more useful.and also sleeping is one of the most important things for teenager. for that, i believe we should start our dairly life earier in the day</p>	<p>farmers animals and said that would be a good way to reduce costs for the council and benefits many local farmers.</p> <p>there isa another person who have a unique idea. sarah case who is an environmental officer said that he is always shocked by how much food they throw away. even if it is hard to judge how much they will sell each day, he believe that we should try and finf a way to reuse some of this food. to solve the problem, he suggested to have a classs in schools to teach children about issue. he belive if children learn from a young age that wasting food is bad, they will grow up to teach thier own children the same. and that will also please parents, as their children will learn to throw away less food.</p> <p>in my opinion, we should introduce a special tax for resturants, households. even if the chlidren learn that food waste is bad, we cannot say that it will helps to reduce the food waste right now. for that, the city should have a special tax and that would be lead to reduce the food waste in greenfield.</p>
---	--

The two essays produced by ST030 are shown below. Although both essays only constitute a single paragraph, she has demonstrated her ability to synthesize the information from both source texts. Some information was also provided from the graphs. Although the writer did state which solution she thought was the best, she did not include the reason behind the choice. All of this is an indication of not identifying the purpose of the essay. Since the essay did not have a logical paragraph structure, the essay was also marked as at the B1 level by the two raters.

Table 82. Written outputs by ST030

Task 1	Task 2
<p>Recently, students' lack of sleep is worried in Greenhill. To deal with this problem, Mike Parker, the principal at North Greenhill High School suggested that teachers should educate students about the importance of sleep by holding a special session to discuss the issue to encourage students to change the daily routine after school. Also, he suggested that school start time should be later to give students extra sleep. Sarah Case, a school nurse, said that morning exercise at school might be helpful to encourage students to go to sleep earlier. She also believed that school times should be adjusted to fit the biological clocks of students. I think educating students about the importance of sleep by discussing is most effective. As the left figure shows, many students spent after school hours watching TVs and using cellphones and this may cause the lack of sleep. If they reconsider the importance by discussing the issue, they won't waste their time watching TVs and using cellphones without sleeping.</p>	<p>In Greenfield, the food waste in the city is worried because garbage-collection is a heavy and expensive work for the city. To address this issue, Mike Parker, the head of the city's waste collection unit, suggested encouraging restaurants to throw away less food by imposing them a special tax that would be calculated based on the amount of food waste. He also suggested that grocery stores should give farmers unwanted food to feed animals. Sarah Case, an environmental health officer, thought stores should try to find a way to reduce some of unsold food. In addition, she thought having classes in schools to teach children about food waste would be helpful to encourage them, their future children, and their parents to know discarding food is bad. I think grocery stores' giving farmers unwanted food is the best way to address this problem. According to the right figure, the amount of food waste from shops and restaurants occupies no less than 34%, so firstly we have to reduce food waste from them. Giving unwanted food is good for the council and shops as well as farmers because they can reduce costs for wasting food.</p>

The essays written by ST037 are shown below. The essays were marked as at the B2 level by the two raters. Again, both essays were similar in style. The participant later revealed that she had taken the real TEAP test in the winter and that she received the news that she has been successfully accepted by Sophia University. Although she was already familiar with the test

format, the gaze-plots reported earlier showed that she thoroughly followed the Task Instruction during the first and last 5 minutes of eye-recordings.

Table 83. Written outputs by ST037

Task 1	Task 2
<p>In the school in Greenhill, the average hours of sleep is gradually decreasing since 2000. Without enough sleep, it will be very hard for students to perform better. To improve this situation there are 3 solutions that have been suggested.</p> <p>First, changing daily routine after school hours. The principal at North Greenhill High School, thinks that to change students' daily routine after school hours can be the first step in dealing with the problem. In 2015, 22% students spent after school hours by doing homework, 31% of them spent after school hours by watching TV, 34% of them spent after school hours by using cellphone and 13% of them spent after school hours by doing club activities. This can be one of the reasons why students can't get enough time to sleep. So the principal suggested to make students realize the importance of sleeping and make them change their daily routine by themselves.</p> <p>Second, delaying the start of morning classes. The principal and Sarah who is working as a school nurse believe that adjusting the school times to fit the biological clocks of adolescents is more realistic and effective. Also, they think it can help children to get enough sleep and can make them more productive in normal classes.</p>	<p>In greenfield, the food waste problem became a serious issue. About 50 tons of food waste increase every 5 years since 2000. Not only the amount of the garbage but also the cost to collect and dispose it became a serious problem. To improve this situation there are 3 solutions that have been suggested.</p> <p>First, targeting restaurants and encourage them to throw away less food. According to a graph of sources of food waste, 34% food waste came from shops and restaurants. Mike Parker, the head of the city's waste collection unit thinks that targeting restaurants is the first step in dealing with the problem. He suggested introducing a special tax to restaurant when they throw away their food waste in order to decrease the amount of food waste which comes from restaurants.</p> <p>Second, using the unwanted food for a different purpose. Mike also suggested to use unwanted food to feed the animals since the food that is thrown away by grocery stores cannot be consumed by humans. He insists that this solution would be a good way to reduce costs and benefit many local farmers.</p> <p>Third, changing citizens' mind to wasting food. Sarah Case who is working as an</p>

<p>Third, suggest students to work out earlier in the day. The school nurse believes that exercise can help young people fall asleep faster and sleep more soundly. So she suggested to work out earlier in the day than in the evening in order to encourage adolescents to go to sleep before midnight.</p> <p>Take these things into consideration, i think delaying the start of morning classes would work the best. Since there are so many students who tend to stay up late doing homework and other activities, it is very realistic and effective way to ensure students' extra sleep. Hence, I think delaying the start of classes would be the best solution among these 3 solutions.</p>	<p>environmental health officer, suggested to make families in Greenfield to change their habits. She was shocked by the great amount of food they discard without even thinking. According to the graph of the sources of food waste, 25% of food waste came from households. So she want schools to teach children about food waste problem so that they can teach their own children when the grew up, too. This will also please parents, as their children will learn to throw away less food.</p> <p>Among these 3 solutions, I think changing citizens' mind to wasting foods would work the best. I think introducing a special tax to restaurants is also a good way to improve the situation in Greenfield but it's a temporary solution and it can't improve the situation completely in the future. So I think it's better to make citizens' realize that wasting food is a serious problem and it is a problem which comes from their habits. This solution will be helpful to decrease the amount of food waste in the future, too. Hence, I believe that changing citizens' mind to wasting food would work the best.</p>
---	--

It is evident from the written outputs that the students tend to produce similar essays reflecting their abilities at their language proficiency level, including the essay length, writing styles as well as the structure of the essays, as indicated by the number of paragraphs.

6.5 Conclusions

The findings of this qualitative analysis resonated with many of the findings from the eye-recording data analysis in the previous chapters, which explored the cognitive processes of

integrated writing process. Some selected cases were reviewed to compare the cognitive processes of integrated writing during the preparing-to-write stage among participants with different writing proficiencies. Gaze-plots during the first five minutes of recordings showed that the participants with a lower proficiency did not review the Task Instruction very carefully in order to identify the purpose of the essay, which was also validated through the survey analysis. The AOI switches of the participants also showed in which order they read the paragraphs from the source texts. The case study illustrated how some participants with higher marks read the source texts in order of the paragraphs, whereas some participants with lower marks skipped a paragraph or two. Although it is not definitive, the case studies showed that the average saccades, regressions and return sweeps could give some understandings of participants' reading behaviours in that, in some cases, the participants with higher ratings had the most saccades and return sweeps, while the less proficient students had more regressions.

The questionnaire responses of the four selected cases revealed that basic knowledge of academic writings affected the behaviours of the participants during the translating and writing stage. By reviewing these additional details about the essays, it was found that less proficient students tended to borrow longer strings of words from the source texts by copying the sentences from the original sources, while more successful students only borrowed some keywords and phrases from the source texts. While most participants were familiar with all the graph types included in the tests it appears that the difference in outcomes lies in the writing skills needed to translate the trends evident in the graphical information into English.

Finally, the gaze-plots from the last five minutes of the recordings of some individuals showed a difference in how individual participants behaved during the monitoring and revising stage of writing. The gaze plots show how each of those four participants reviewed and monitored their essay. Furthermore, the gaze plots and written outputs were compared between Task 1 and Task 2 and it was found that both essays exhibited a similar pattern.

Overall, this chapter has illustrated some selected individual cases for qualitative analyses using a combination of gaze-plot data for timed segments, written outputs and survey results from each cognitive writing process. Integrated writing processes are known to be complex, and higher proficiency in reading and writing does not always guarantee a successful result in the integrated essays. These individual cases showed how some interpretations of the eye-tracking data could be made to understand the cognitive processes of individual test-takers in depth.

Chapter 7. Findings from the Focus Group Discussions

7.1 Introduction

This mixed-methods study used an explanatory sequential design in which the analysis of quantitative data was followed up with some qualitative data analysis. The eye-tracking data explored overall trends of test-takers' reading-into-writing behaviours in association with eye-movement metrics, and the questionnaires provided possible explanations for test-takers' decision-making processes. Focus group discussions formed the final phase of the findings. The discussions were conducted with a group of participants who were willing to discuss and share their test-taking experiences. The meetings were scheduled within several days of the eye-tracking experiments and 24 students participated to discuss ten open-ended questions during the course of an hour (See Appendix E). Focus groups discussions were appropriate because participants were able to share their experiences with their peers. Exchanging their experiences and thoughts about the test was a positive learning experience for all. The eye-tracking information that was shared with the groups sparked extensive discussion. This chapter reports the comments and feedback from those participants.

7.2 Analysis of the Focus Group Discussions

7.2.1 Focus group discussions as a methodology

After the eye-recording tests had taken place, all participants were invited to join an open feedback in a focus group discussion. Since the eye-recordings were conducted right before the school spring holidays, it was hard to recruit all 42 of the students who had participated in eye-recordings. Some students had already graduated by the time all the eye-recordings were finished and had moved out of the town for a college education. Nevertheless, a total of 24 students were able to participate in the discussions.

The focus groups began with some samples of gaze plots taken from the series of recordings captured at 10-minute intervals in order to highlight how participants act differently at certain language levels. A clip of a video was played to show how the eye-fixation moved as time progressed in the video. There were a lot of “wows” from students. Then a series of questions were asked, and the groups discussed these one by one. The analysis that follows summarizes their feedback and comments on each topic, categorized by their writing proficiency level.

7.2.2 Participants’ inner thoughts revealed in focus group discussions

Only 24 participants were available to join the focus group discussions. The responses to the questions were categorized according to the essay ratings, using the TEAP criteria rubrics, as an attempt to highlight what kind of problems the students had faced at different language proficiency levels.

Table 84. Participants’ language level in Focus Group Discussions

Based on TEAP Ratings	Number of students
B2	1
B1	8
A2	15
A2 below	0
Total	24

1. Did you clearly understand the instructions for the test?

When the discussions began, some students seemed to question whether they had misunderstood the task requirements for the test. The researcher therefore asked one of the students to read the task instruction out loud in order to refresh the group’s memory and to clarify what they had been asked to do in terms of writing the essay. By the look on their faces the students seemed puzzled. The students were asked if they have taken any Eiken-Grade tests before. Many of them had taken Eiken-Grade Pre-2 tests or Grade 2 tests in the past, since over

two and a half million junior and high school students take these each year in Japan. This question was asked in order to see whether the students were familiar with a type of independent task which usually appears in the speaking section of the Eiken-Grade tests. In these tasks they are typically asked to state their opinions about a social topic (e.g., Do you think drinks in plastic bottles will be more popular in the future? Why or why not?). They knew they had to choose a side they agree with in order to state an opinion or a view with a few reasons to support their statement. They usually gave reasons based on their recollection of memory, knowledge and experience. Such independent tasks were also everyday tasks when they had a debate and discussion in classrooms, for which they sometimes prepared written scripts. Consequently, Japanese students were mostly familiar with an impromptu essay. They seemed to know little, however, about the integrated essay approach that appears in TOEFL iBT®, IELTS® and TEAP®.

A quick tour was therefore given to show the major differences between independent and integrated writing tasks. One student said, "I have taken the TEAP test before, so I understood the instruction very easily (B2)." Other students stated that they understood the instruction. Some students, however, admitted that, "I thought I understood, but now that I heard what you just explained, I am not sure if I entirely understood the instruction back then. Also, I did not know how to summarize the main points of discussions. (B1)" This echoed another student who said, "I understood the instruction only when I read the task instruction for Task 2 (B1)."

2. Which part of the tasks did you have the most difficulty with?

The student with a higher proficiency level (B2) said that giving the reasons behind the argument was a more difficult task than summarizing the main points about the solution. About half of the participants (B1-level participants (N=4) and A2-level students (N=8)), however, claimed that summarizing the main points was the most challenging work. There was one

exception in that one student said describing a situation by interpreting the graphic information was the most difficult. Then, their focus and attention shifted from stating their reasons for summarizing the task. After the task requirement was made clear, they suddenly all came to realize that they needed to synthesize the information from the given source texts. Without the focus group discussions, the participants would have been left misunderstanding the required task for the test.

3. To what extent do you think your typing skills affected your writing process?

There were a handful of students who said: "I do not think it (keyboard) has affected me so much" (N=4 from B1 N=3 from A2). As we took some time to discuss this in groups, however, some voices were heard such as, "It took longer to write an essay using the keyboard (A2)." "I had to pause a bit to recall the correct spelling of the word (because I am not used to doing this on the keyboard) (A2)." "It took me a lot more time (A2)." "It affected me a lot (A2)." "I had to keep looking down at the keyboard. So, I run out of my time (B1)." "I am not good at typing. So, I do not know to what extent it affected me (B1)." Participants' basic typing skills could have caused some delays and difficulties when trying to produce an essay using the keyboard; however, other voices from the focus group discussions also seemed to suggest that it has to do with the pressure to produce as many as 200 words within the 40-minute time constraint.

4. How did you decide which information to include?

This question uncovered rather more detail than had been evident in the original survey question. The table below summarizes the inputs according to the participants' language proficiency levels. Apparently, some students did not know the basic organization of an academic essay. They were often confused about the topic sentence which appears in every

paragraph with a thesis statement (main topic). This could be due to some cultural differences. In an academic essay written in the Japanese language, readers will often find the thesis statement in the conclusion, since this is considered the most important sentence of the essay. Based on the feedback comments, it was evident that some students looked for main ideas in the last paragraph (which is the conclusion) or in the final sentence of each paragraph.

Table 85. Summary of feedback and comments in the Group Discussions on Question #4.

TEAP Essay Marked by raters	Feedback and comments shared by the participants of the Focus Groups Discussions on how they decided which information to include,
B2	<ul style="list-style-type: none"> • Looked for keywords that appeared many times in the source texts.
B1	<ul style="list-style-type: none"> • I did not think anything very deeply. • Graphs.... • I wrote in order of the task requirement. • I read the last paragraph. • While I was reading the texts, I sort of decided what to include. I looked for keywords that appeared in the text. • I borrowed some words that some people said in the texts.
A2	<ul style="list-style-type: none"> • I just began writing without much thinking or planning. • I thought where to look as I read the texts. But I do not think I did not give much thoughts about it. • I do not remember much details, but I looked for a topic sentence or main idea from the text. • I looked at the last sentence in each of the paragraphs. • What the main character has said about and graphs • After I had read the text, I began writing • The parts in which some general statements were written • At that time, I did not think much about anything • The opinions that main characters have said. • The title of graph • I thought about it while I was reading • I wrote some words that seemed to be important from the text.

When asserting their own point of view as a writer, Japanese students often use vague or hedging language. For example, they tend to start a sentence with phrases like "I wonder...." and "I think.....". This was true of many of the participants who began their essay by stating, "I think..." or "I believe...." Hirose (2003) has reported that "(t)his tendency to postpone one's main point until a later position has been found not only in student texts but also in those written by professional writers. (Hirose, 2003. p.182)." Listening to their discussions, it became clear that they intended to write an independent essay. Indeed, there were abundant reasons to believe that they wrote their piece in that fashion because culturally that was the norm for how they would write an academic essay in Japanese. Such cultural differences hindered their abilities to write a standard academic essay to some extent. Whether their essay had several

paragraphs or just one single paragraph, it was not written in a way that most English speakers would easily understand due to a lack of cohesiveness. This is not only an issue for Japanese English learners, and indeed various scholars have undertaken empirical studies on the topic of L1 influence on L2 production (e.g., Kubota, 1998, Hirose, 2003).

5. How did you plan when you started to write your essay?

According to the survey results, only three students claimed that they outlined what to write before they began to write an essay, and only twelve students had thought about how many paragraphs there should be in their essay. Through the discussions, most of the students said that they began writing an essay without allowing much time to plan and to think about its organization. The summary table below shows that this was, unfortunately, the case across different writing levels. There was some evidence, however, that a few of the B2 and B1 participants had some consciousness, even though they were not directly making any outlines before they started writing an essay. For example, a B2 level student recalled what steps she had taken, such as “by summarizing the information from graphs and organizing my thoughts in my mind” and “by thinking about the situations, then summing up the points and my opinions.” On the other hand, many of the participants with A2 level writing proficiency said: “I will make outlines (in the future).” In other words, prior to learning from the focus group discussion they had not thought about how to plan an essay before they began writing it.

Table 86. Summary of feedback and comments in the Group Discussion on Question #5.

TEAP Essay Marked by raters	Feedback and comments shared by the participants of the Focus Groups Discussions on how planned when they began to write the essay.
B2	<ul style="list-style-type: none"> ▪ Looked for keywords that appeared many times in the source texts. ▪ By summarizing the graphic information and organizing my thoughts in my mind.
B1	<ul style="list-style-type: none"> • I looked for the main idea. • By thinking about the situations, then summarizing the points and my opinions. • I did not have any planning, but I began to include some information from graphs and my opinions about the issue. • I tried to look for some relevant information and began writing about these points. • I did not have any planning. I just began by following the direction of the task instruction.
A2	<ul style="list-style-type: none"> • I began with Mike Parker's opinions. • I just began writing my essay. • I did not think about my planning. • I thought about the main story by picking up some keywords. • I first realized the overall situation and thought about the flow of the essay; then I began writing. • I used my hand gestures while I was thinking. • First, I chose my opinion about the topic. • I did not make an outline and had no plan before writing. • Without giving much thought to it, I began producing my essay. • I did not think of anything but just began writing my essay. • I made sure that a sentence follows in a logical order. • I wrote facts first and then my opinions. • I will make an outline next time. • I will make an outline of some keywords. • I will make some outlines. • I will describe situations (using) graphs, and then summarize points. • I want to describe the situation and then summarize the main points.

6. What did you often do while you were writing your essay?

This question opened a broad range of responses from the focus group participants. When the responses were organized according to the different levels of writing proficiencies, it became apparent that the responses were affected by their writing proficiencies. For example, most

responses from the B1 level students had to do with mechanical corrections and monitoring, such as spellings. Also, at A2 level, they heavily depended on the keywords in the source texts, which led them to copy sentences with relevant information. On the other hand, B1 level students said they worried about keeping up with time, whether they had enough words, and about avoiding the repetition of the same keywords in their essay. Also, there were some signs that they were worried about organizing thoughts while writing.

As mentioned earlier, Weir and Shaw (2007) described some features of cognitive processing across the examination suits in CEFR levels (See Section 2.3, Table 2). According to their description, writers at B1 level showed some evidence of “monitoring and revision of vocabulary, grammar, and basic organization.” At B2 level, writers showed some evidence of “monitoring and revision of styles and contents.” At C level, writers demonstrate “rhetorical and organizational analysis, evaluation, planning and monitoring and revision of style and content.” The descriptions here matched with the descriptions of the cognitive processing proposed by Weir and Shaw to a great extent. If this were the case, I believe this focus-group question validates that the cognitive processes vary among test-takers with different writing proficiencies.

Table 87. Summary of feedback and comments in the Group Discussion on Question #6.

TEAP Essay Marked by raters	Feedback and comments shared by the participants of the Focus Groups Discussions on what they often did while writing the essay.
B2	<ul style="list-style-type: none"> • I counted how many words I have written. I was worried about how much time left for writing the essay, but I checked my sentences to see if there were any spelling mistakes.
B1	<ul style="list-style-type: none"> • I often went back to the keywords in the text. • I looked for the solutions while writing. • I copied from the text but I made some changes to the sentence. I used some words that were closer in meaning. • I entirely copied some words and sentences. • I like to think by using my hand (body-gesture) while thinking. • I used some sentences from the texts.
A2	<ul style="list-style-type: none"> • I re-read my essay and tried to avoid the use of the same word. • I went back to the main idea, like the issue and solutions. • I reviewed my essay to see if I have written something wrong. • I paused once a while. I could not really organize my thoughts while I was writing. • I looked at the text while I was writing. • I actually copied some sentences and borrowed some words from the texts. Now I know. Don't copy and paste! • I counted how many words I have written. • I re-read some parts of the text where I could find some keywords and the solutions. • I reviewed the text by thinking whether I could have related to my experience. • I reviewed a part of significance and I also copied these sentences. • I reviewed by going back to some keywords. • I used my hand to think while writing. • I monitored the spelling of words. • I copied some sentences from the text. • I found some keywords and thought about the solutions. I also copied some sentences. • I literally copied the sentences from the text. • Find problem and solutions by finding the keywords

As shown in Table 88, many of the B1 level participants on the TEAP test seem to have engaged in micro-level processes (i.e., mechanical monitoring such as grammar and spelling), while the B2 student showed some evidence of macro-level (rhetorical and organizational) monitoring and revision processes. A2 level students failed to engage in a thorough monitoring and revising process.

7. What did you do after you finished writing your essay?

This question was asked to validate whether the test-takers had gone back to review the Task Instruction to make sure they had fulfilled all the required tasks, and whether the evidence shown in the eye-recordings matched with their claim of what they did during the last five minutes of the test. Most of the participants stated that they had spent some time going back to their written outputs in the last five minutes, but none of them mentioned that they went back to the Task Instruction for monitoring and revising.

Table 88. Summary of feedback and comments in the Group Discussion on Question #7.

TEAP Essay Marked by raters	Feedback and comments shared by the participants of the Focus Groups Discussions on what they did after they finished writing the essay.
B2	<ul style="list-style-type: none"> • I checked subject-verb agreement for missing "S." Also, I tried to revise some sentences that would be easier (for readers) to understand.
B1	<ul style="list-style-type: none"> • I tried to review my essay but had little time to do so. • Reviewed. • Last 10 minutes for reviewing • Until very last minute, I tried to review and think of the idea. • I did not do any reviewing. • I stopped now and then when it was convenient • I didn't have the time to finish all my writing. I wish I would be able to review my essay by finishing early.
A2	<ul style="list-style-type: none"> • Reviewed. • I spent the last two minutes reviewing my essay. • I reviewed • I stretched a bit but I did my spelling check and revised the verb tense in sentences. • I re-read the whole essay and counted my words in the essay • Re-read my essay • Reviewed the entire essay I wrote • I did not have the time to review my essay for correction. • I added some sentences. I also checked my spellings. • Reviewing is very important.

8. What kind of strategies did you use while taking the test?

The question was about what kind of strategies they used during the test. The responses also included what they did not do. The responses were coded into five types of strategies: a) structure of an essay, b) paraphrasing, c) use of synonyms, d) spelling and vocabulary and finally e) time movement. The table below summarizes which strategies each level of learner was concerned with.

Table 89. Summary of feedback and comments in the Group Discussion on Question #8.

TEAP Essay Marked by raters	Feedback and comments shared by the participants of the Focus Groups Discussions on what kind of strategies they used while taking the test.
B2	<ul style="list-style-type: none"> • I tried to avoid using the same phrases. For example, instead of "Taking these things," I wrote "Among these three solutions."
B1	<ul style="list-style-type: none"> • I need to learn how to paraphrase. • I don't have any. • I tried to avoid repetition of the same word and also tried to manage the time. • I thought I tried to write in my own words as much as possible. In reality, however, I wrote my essay without giving much thought to it. • I tried to avoid the same word as much as I could. • Time management • There wasn't anything particularly I did.
A2	<ul style="list-style-type: none"> • Time allocation for each paragraph. • Time management: planning a framework of an essay is very important. • I tried to keep my eye on the time, but it did not go too well. • I tried to shorten a paragraph of the text into sentences. I did not think much. • Time management • Nothing really. • To stretch my essay, I avoided a contraction (I'd = I would). • I was very conscious about an organizational structure with paragraphs. • extending the word counts by avoiding a complex sentence • I tried to copy a sentence from an important part of the essay, but that was a mistake I made. • Write both a summative and an opinion in one essay. • I hope to paraphrase next time and use a synonym word--not the same word all the time. • I saved last few minutes for reviewing and revising my essay. • I gave it to all my best using my brain and thinking in my hand.

Some A2 level students expressed that they would have approached their writing differently if they had referred to the question, such as “I hope to paraphrase next time.” The types of test-taking strategies that most B1 level students were particularly concerned with focused on paraphrasing and use of synonyms. Such skills are necessary for achieving the B2-level essays. B2-level students were more concerned about the overall structure of an essay and the time management skills. In fact, less than half of the B2 and B1 level learners mentioned time management. Nonetheless, time management skills are essential if the writers were to write some logically developed paragraphs.

9. What were you mostly concerned about your essay?

After asking some specific questions about a certain stage of the writing tasks, this open question was asked to reveal the participants’ overall experience of taking the test. They were concerned about many things, ranging from writing skills to reading comprehensions for the test. The table below summarizes the comments made by the participants, organized by different writing proficiency levels.

Table 90. Summary of feedback and comments in the Group Discussion on Question #9.

TEAP Essay Marked by raters	Feedback and comments shared by the participants of the Focus Groups Discussions on what they were mostly concerned about the essay.
B2	<ul style="list-style-type: none"> • Verb tense because I sometimes used a mixture of present and past verb in the sentence.
B1	<ul style="list-style-type: none"> • I was concerned about my grammar and also the reasoning for the choice I've made. • How to develop paragraphs-introduction, bodies, and conclusion • I was concerned where in the essay I should describe the information from graphs. • How to develop paragraphs-introduction, bodies, and conclusion • I could not think of any good phrase to be used in my essay. • Time management • I did not have the time to finish writing my essay. • I was worried about my grammar mistakes and how much time I was left. • I was worried whether my essay had a coherence in the paragraphs.
A2	<ul style="list-style-type: none"> • I was concerned about the verb tense and in which order to write the essay. • My skill to write an essay--what is an essay--I kept wondering about it. • I did not have enough time to finish my essay. I did not have enough word count for my essay. • I do not think my essay was written in a cohesive and logical manner. • Nothing particular • Verb tense and how many paragraphs there should be in the essay. • I was not quite confident how to write things I meant to say. I wish I knew how to write a better sentence. • Copy and paste • Verb tense • Whether I clearly understood texts and wrote my essay without a grammar mistake • Whether I wrote my essay according to the task requirement. • I did not understand what to write for my essay. • I was concerned about not being able to summarize the points well. • I also could not write enough words for my essay. • Time and word counts

At A2 level, some students were not sure if they had clearly understood what they were supposed to write. Also, many of them were not confident about the grammar and were worried about making mistakes. At B1 level, students were mostly concerned about the lack of phrases and writing an essay with an adequate length. The B2 level student was worried about which verb tense to use in the essay.

10. What do you think you would do differently if you had to take the same test again?

The last question in the focus groups was about understanding what the participants would do differently if they were faced with the same kind of test in the future. Table 91 summarizes the participants’ responses organized by their language proficiency levels in writing.

Table 91. Summary of feedback and comments in the Group Discussion on Question #10.

TEAP Essay Marked by raters	Feedback and comments shared by the participants of the Focus Groups Discussions on what they would do differently if they had to take the same test again.
B2	<ul style="list-style-type: none"> • I understood I need to make an outline before I start composing an essay.
B1	<ul style="list-style-type: none"> • I need to be more cautious of where I got the ideas from. Also, I need to read the texts by paying my attentions to 5W1H. • I understood I need to make an outline before I start composing an essay. • I need to make outlines by thinking about an overall structure. • I want to enhance the usage of my vocabulary • I learned about a paragraph writing I also want to apply six things I learned today. • I will first skim the entire text and check keywords. • Use synonyms to avoid the repetitive words, be careful about the tense. Not borrowing the sentences from the source text and I will write my opinions. • First, I will think about an essay comprised of three paragraphs, and I will use my words instead of copying the original source.

A2	<ul style="list-style-type: none"> • I will make sure I understood the task requirement and then skim the source text. I will make a list of important things and summarize them in my words. • I will need to review the task instruction more carefully. I will not copy from the original sentences and paraphrase them instead. I will reduce the run-on-sentence errors. • I will not ever copy and paste the original sentences. I learned the importance of the forms of academic writing. • I will apply what I learned today in our discussion like read the task instruction more carefully and review my texts after I finish writing my essay. • I wish to make an outline next time I write an essay. • I want to pay more attention to an organizational structure of each paragraph. Especially, I want to pay more attention to the discourse marker, like “AND” and “But” (→Furthermore, However). • I want to improve paragraph writing skills and would also like to increase the number of vocabularies. • I will more carefully read the source texts thoroughly and pay more attention to run-on sentences. • Time management and paraphrasing • I will try to avoid repeating the same word. I will write my essay in the order of Introduction-Body-Conclusion • I will make an overall structure that is easier for the readers to understand. • I will make an outline before writing my essay next time. • I will answer what is written in the task instruction and answer to each task accordingly. • I will be careful how to develop an overall structure of the essay. • Time management. I will need to review whether problems and solutions have been stated.
----	---

After the discussions, the students seemed to understand some of their weakness and what could be done to improve their score on the TEAP writing Task B test. The participants also understood that the TEAP writing test looks not only at the grammatical and lexical range and accuracy, but also at the main idea and cohesion and cohesiveness. The participants seemed to understand that using synonyms to avoid repetition of the same word can help build lexical range, and paraphrasing sentences rather than copying them is essential for a cohesive and coherent essay (Shi, 2012). Most of these students, however, did not know how paraphrasing could avoid “plagiarism” until we discussed the issue. For some, it was the first time they had

heard the term “plagiarism.” Furthermore, they seemed to understand that making outlines was helpful for the logical development of the paragraphs in an essay. This is particularly the case among the students at higher writing proficiency levels.

7.2.3 Summary of focus group discussions

Overall, the participants in the group discussions could get some useful advice by sharing their experiences with each other. It is important to acknowledge that one of the most distinctive features of focus-group discussions is the dynamics that such groups can create. Group discussions have the advantage of providing information on a range of issues from a spectrum of perspectives and opinions of individuals (Rabiee, 2004). This study also benefited from the range of data generated by the participants from different levels of writing proficiencies.

Focus group discussions collected not only the data from the participants about their test-taking experiences, but they also helped to reveal some information about their knowledge of academic writing. The aim of the focus group discussions was to investigate what the students did before, during and after the TEAP writing test. Skilled writers (i.e. B2 and B1 level students) were able to describe what they did or did not do during the test to unskilled writers during the focus group discussions. The video clips, and some gaze-plots of eye-tracking data, were very helpful in reminding them what they did. Some of the less successful writers, however, did not remember much about what they did or why they did it.

As noted earlier when reporting the results of the eye-tracking analysis, some students at higher writing proficiency levels did go back to the Task Instruction, although some of them were not conscious of what they had done. The video clips were shown to illustrate the difference in how the last five minutes was spent between high and low proficiency learners. The participants in the focus groups could clearly see how successful test-takers went back to the Task

Instruction during the monitoring and revising processes. The scientific evidence provided here had a great impact on the participants in terms of helping them to see the differences in techniques applied by higher and lower proficiency writers.

Through the focus groups discussions, it became clearer that some of these students did not know what they were supposed to write. In the discussion, some of them reported that they did not know how to borrow ideas and synthesize information. The discussions also revealed that the students with lower proficiency levels were more worried about making spelling and grammar mistakes rather than building cohesive paragraphs for the essay.

The dynamics of focus group discussions encouraged both successful and less successful writers to share their experiences with the process of integrated writing and gave the opportunity to compare how much they knew in advance of the test about the basics of academic writing. In this regard, focus group discussions contributed tremendously to getting deeper insights into the student's knowledge of academic writing techniques. The focus group discussions were very beneficial for understanding participants' perspectives and experiences of this study.

Chapter 8. Discussion

8.1 Introduction

The current study was conducted to investigate what affects the cognitive processes of integrated writing when test-takers used information from multiple source texts and graphs as prompts. This study used a mixed-methods approach with quantitative and qualitative data, including standard tests to describe participants' language proficiency backgrounds, eye-tracking experiments, cognitive-processing questionnaires, and focus group discussions.

In the first phase of data collection, the Aptis test was conducted to measure the participants' reading and writing proficiencies. Only moderate correlations were found between their reading and writing scores ($r=.561$, $p<.001$). The findings from this study also confirm the literature, in that Flashive and Bailey (1993) argued that only moderate correlations were found between reading and writing among L2 learners. In a sense, the participants who had higher reading scores could not automatically be assumed to have higher scores in writing. Thus, it became all the more important to investigate the complexity of L2 learners' cognitive process of integrated writing.

In the second phase of data collection, the *Tobii TX300* eye-tracker was used to collect eye-movement data while the participants took two sets of integrated writing tests modelled on the paper-based TEAP Task B test. Task 1 and Task 2 were designed to be similar in format and language features except they were written on a different topic and had different types of information from graphs (i.e., Line Graph in Task 1 and Bar Graph in Task 2). The eye-tracking metrics of the participants were examined in association with their reading and writing abilities to run some statistical analyses, such as the Wilcoxon Signed-ranks tests, the Mann Whitney-U tests and the Multiple Regression tests. The Wilcoxon signed-ranks tests were performed to identify statistically significant differences between Task 1 and Task 2 in terms of how the

participants viewed different AOIs. The researcher attempted to find any differences between the different types of source texts and information from graphs. The Mann Whitney-U tests were performed between the upper (B2 above) and the lower (B1 below) intermediate groups of the students in both reading and writing in order to investigate the effects of language proficiencies in integrated writing. Finally, the Multiple Regression tests were conducted to identify whether any variables in eye-movement metrics can be predicted statistically from reading and writing scores.

In addition, the analytic scores of the essays produced during the eye-tracking experiments were rated by two raters for further analysis. Using the five criteria from the TEAP Task B Integrated Writing (i.e., Main Idea, Coherence, Cohesion, Lexical Range & Accuracy, Grammatical Range & Accuracy), Pearson Correlation tests were conducted using the eye-tracking measurements to explore whether the eye-movement metrics could predict any of the scores. This analysis was carried out to help understand what kind of information eye-tracking metrics contain and how they could help understand the test-takers' behaviours and decision-making process while synthesizing information from multiple source texts and graphs into writing.

In the third stage, the survey results were analysed to explain the circumstances systematically in each stage of the cognitive process of writing. Using the model of the Socio-Cognitive Framework for Language Testing proposed by Shaw & Weir (2007) as a theoretical basis, some individual cases were described in three stages of writing, namely the preparing-to-write stage, the Translating and Writing stage and the Monitoring and Revising stage. The answers to cognitive-processing questionnaires attempted to understand participants' views of the source texts, information from graphs, their familiarity with the graph types used, their test-taking strategies, and their knowledge of academic writing.

Finally, the findings from focus groups discussions helped to account for the reasoning behind the numbers. The participants volunteered to discuss open-ended questions in small groups to share their experiences and learn from each other about how they did on the integrated writing test. The feedback and comments were categorized according to their essay ratings in CEFR (using the TEAP rating rubrics) to help deepen the understanding of an L2 writer's cognitive process of integrated writing.

The study explored how test-takers employ cognitive processes while completing tasks using the TEAP Writing Task B (which uses multiple source texts and graphs). The study attempted to validate one type of integrated writing task (reading into writing.) This study validates the earlier work of Chan (2013) which examined the context and cognitive validity of reading-into-writing test tasks in comparison with real-life academic writing performance. Her study concluded that reading-into-writing tasks during the tests were similar to the cognitive process of writing as demonstrated in the real-life academic writing task. According to Chan, the test scores may “be able to predict the performance in the target context better at high and low levels than at the medium-level” (p.297). The results of this study have also shown the difference between higher and lower intermediate groups of participants in terms of how they process the information during the first ten minutes.

In addition, this study also reported only a moderate association between reading and writing scores by analysing the Aptis scores and the TEAP writing scales, indicating the complexity of the integrated writing tasks among the L2 learners as shown in the eye-movements data (AOI switches, regression, saccades, and written outputs). These quantitative findings were triangulated, by data from the questionnaire and focus group discussions.

The table below is an executive summary of what affects integrated writing using the TEAP integrated writing tests.

Table 92. Executive Summary

What affects Integrated Writing Processes	Eye-tracker	Written Outputs and Questionnaires	Focus Groups Discussions
Source Texts	-Task instruction -Expeditious reading	-Source text borrowing -Knowledge of academic writing	-Identifying the purpose -Organizational structure
Information from graphs	-Position of the graphs -Graph types	-Familiarity with the graphs -Ability to summarize the graph trends in English	-Ambiguity of the task requirement
Test-Taking Strategies	-Identifying the purpose -Time management	-Preparing to write process -Avoiding source text borrowing	-Language use -Types of essays
Language Proficiencies	-Headings of the source texts -Task instruction -Essay compositions -Monitoring process	-Identifying the purpose -Organizational structure -Language use	-Paraphrasing -Paragraph writing -Coherence

This chapter will revisit each of the five research questions to discuss the elements that could help determine what affects the cognitive processes of integrated writing. The contents of the chapter are organized according to each research question such as (1) Effects of source texts, (2) Effects of information from graphs, (3) Effects of types of graphs, (4) Effects of test-taking strategies, and finally (5) Effects of language proficiencies.

8.2 Research Question (1) Effects of Source Texts

Writing ability using information from source texts is one of the fundamental skills for academic success. To find the effects of source texts on L2 writers in the context of language assessment, it was important to understand the role of reading abilities in writing. In other words, how reading proficiency influenced the way that the test-takers could synthesize information from the source texts. Based on the findings reported in the previous chapters, this section has three components: (1) Eye-tracking, (2) Cognitive-processing questionnaires and written outputs, and (3) Focus Groups Discussions.

8.2.1 Evidence from Eye-tracking data and analysis

Plakans (2008) revealed that her participants showed a recursive approach during the process of integrated writing, largely because they needed to go back to the source texts to formulate ideas. While this was found to be mostly true, the eye-tracking data in this study showed strong evidence that the participants engaged with the source texts most during the first ten minutes.

Also, the Wilcoxon tests suggested that the participants tended to rely upon the information from the first few paragraphs (Paragraph 1 and Paragraph 2 in the Newsletter article). Mann-Whitney U tests were conducted to understand the patterns found in these reading behaviours. The results indicated that language proficiencies did affect the way students view those texts. Particularly, a statistically significant difference was found between the upper and lower intermediate writing groups in respect to the extent of focus on the Task Instruction during the first ten minutes of recording. The multiple regression tests were able to find strong associations between the Total Visit Duration on Task Instruction for the first ten minutes, as well as for the 40 minutes of recordings for more proficient students.

In addition, four individual cases were investigated to explore the reading behaviours in more depth; comparing their average numbers of saccades, regressions, return sweeps, and the overall numbers. The literature review showed that the number of saccades is a good indication for expeditious reading. Saccades of fewer than 7-9 letters have been reported as a sign of reading difficulties in the previous literature (e.g., Rayner, 1998, Bax 2013). Although, in the current study, the number of saccades were only illustrated based on a small sample, the participants who had successfully completed the integrated writing tasks in the TEAP tests tended to have more average regression and return sweeps. According to previous studies (e.g., Bax, 2013; Bax & Weir, 2012; Brunfaut & McCray, 2015) the larger the number of average regressions the more the reader is likely to be having difficulties. In integrated writing, however,

the successful test-takers may be required to go back and forth between the texts to synthesize information from the source texts as quickly as possible. It can therefore be said that having more regressions and return sweeps in the context of integrated writing may not necessarily be a sign of unsuccessful behaviours, as is evident in the cases reported in this study.

8.2.2 Evidence from Surveys and Written outputs

Text borrowing is a concern for rating integrated writing essays. Weigle and Parker (2012) investigated the effects of source text borrowing in a language assessment context for integrated writing, asserting that only a few differences were found regarding the amount of and the types of borrowing between higher and lower proficiency groups of students. Cumming et al. (2005), however, argued that textual borrowing differed depending on the types of topics and the students' language proficiencies. The current study also found through the cognitive-processing questionnaires that the participants with less writing proficiency tended to copy the sentences directly from the source texts or borrow some words and phrases from the source texts.

As has previously been suggested by Plakans (2008), the outcomes of integrated writing tasks showed a substantial difference in respect to students' experience in academic writing. In other words, the performance mirrors students' knowledge or prior experience in academic writing. The four sample written outputs were extracted to indicate their verbatim source use as well as the parts of sentences which showed their attempts to give credit to the person or information in the source texts as a reference. From the written outputs, it became clear that the number of words lifted from the source texts also differed between the participants and they seemed to be influenced by the language proficiency levels.

In addition, the number of words produced in the essay had a close relationship with the number of paragraphs for the essay. The differences in the use of source texts could be the result of the

difference in knowledge and familiarity with academic writing. The survey results and written outputs shed light on the use of source texts among the participants at different writing proficiencies in a way that complements the eye-recording data analysis.

8.2.3 Focus Group Discussions

According to Yu (2009), the extent of familiarity with the discourse types of source text affected students' abilities to summarize the information in the source texts. The questionnaires revealed how participants decided what information to include from the source texts. As reported earlier, the comments from some participants revealed that some students looked for the most important views or the opinions of the authors in the conclusion of the texts, as is often the case in Japanese essays. This showed how these students did not know where to find the relevant information in English essays. Had they known the basic organizational structure of an English academic essay (i.e., main thesis, topic sentence, so forth), this would not have happened. The group discussions were a useful vehicle in understanding some of the cultural differences in and students' knowledge about academic writing in English.

8.3. Research Question (2) Effects of Information from Graphs and (3)

Types of Graphs

There are two types of graphs used for the TEAP writing test, typically a line graph and a pie chart. In designing the two TEAP sample tests for eye-recordings, different types of graphs were presented in order to study the effects of graph types. In Task 1, a line graph titled, "Average Hours of Sleep Among High-School Students in Greenhill" was placed in the upper left-hand corner, and a pie chart titled, "How Students Spent after School Hours in 2015" in the upper right-hand corner. Task 2 included a pie chart titled, "Sources of Food Waste" in the upper left-hand corner, and a bar graph, "Food Waste in Greenfield (collected as garbage)" in the upper right-hand corner.

Wilcoxon tests were performed on the eye-tracking data for 40 minutes of recording to find any differences between the types of graph in the tests in respect to the Total Visit Duration. The survey questions sought to rate each type of graph in order of difficulty on a scale of 1 to 5. The focus group discussions were also analysed to explain the eye-tracking data and the questionnaire results.

8.3.1 Evidence from Eye-Tracking data and analysis

According to Yu et al. (2011), different types of graphs affect students' cognitive processes of writing. Particularly, tables require higher cognitive demands than line graphs. In this study, pie charts, line graphs and bar graphs were compared. The longer fixation duration could mean either that the participants were trying to get as much information as possible or that they were having difficulty with the information. The Ratio of Total Fixation Duration revealed that the most difficult type of graphs seemed to be the line graph. As reported in the findings, less proficient students tended to rely on the pie charts as a main source of information.

What was unexplained from the eye-tracking analysis was the difference between the bar and the line graphs. Given the similarities of the information presented in those two graphs, it was surprising as to why the participants viewed the line graphs longer than the bar graphs. There were a few reasons that could possibly explain the trends. One was that the participants simply found more useful information in the line graph than the bar graph. Another was that the participants found it cognitively more challenging to find the information from the line graphs (and thus viewed the graph longer). Another reason, however, could be that the students were struggling with the English words or phrases that could be used to describe the trends illustrated in a certain graph type. These questions could not simply be answered by analysing the eye-tracking data. In this regard, the cognitive-processing questionnaires and the focus group discussion were very helpful.

8.3.2 Evidence from Surveys and Written outputs

There were five questions asked in the survey related to the use of information from graphs. The participants had been invited to rate the degree of difficulty they experienced in understanding certain information from the graphs, including: 1) Titles, 2) Values on units, 3) X and Y-axis, 4) Interpreting information from the graphs, and 5) Summarizing trends from the graphs. The survey results showed there was no significant difference in respect to the familiarity with the various types of graphs. Thus, there was no difference among the participants in interpreting the information from the graphs. There seems, however, to be some differences in their language abilities to summarize the graph trends in English according to the survey results.

The results from the eye-tracking data and survey could be cross-examined by referring to the written outputs of the participants. For example, Extract 3 from ST053 (A2-below level) had no reference to information from graphs, while Extract 1 from ST030 (B1 level) showed that her essay referred to the information from the pie chart. From the pie chart in Task 1, the test-takers were able to describe possible causes that led to the lack of sleep among the teenagers in Greenhill. Extract 2 from ST037 (B2 level) referred successfully to both the line graph and the pie chart for information. Yet, the description of the information from the pie chart was more extensive than the information taken from the line graph. The only mention of the line graph was “the average hours of sleep is gradually decreasing since 2000” (Extract 2, Line 1). By looking at the language used for describing the line graph, it was easy to grasp that the students did not have much experience in writing about information from graphs. In addition to the use of the gerund to describe the trends in the line graph (e.g., decreasing, increasing, rising, dropping), they had to be able to use an appropriate tense for matching the time frame (e.g., from year X to year Y, between year X and Y, since year X). None of the participants was able to describe the trends in detail. For example, someone with higher proficiency would

be able to describe how much of a drop or increase had happened by using some phrases like “dropped in half” or “two-fold increase.” Such lack of confidence impacted their ability to refer to the information from the line graph. Instead, most participants relied heavily on the information from the pie chart to describe the leading causes of the problems in the essay.

8.3.3 Focus Groups Discussions

Initially, the participants overwhelmingly seemed to agree that the most difficulty was with “summarizing the main points about solutions that have been suggested”. This was stated in the task instruction and the participants knew that the goal of the essay was to summarize the main points about the solutions. The second popular answer was “giving the reasons behind your argument.” This was also stated in the task instruction as “say which of the solutions you think would work the best based on the reasons given.” Most students who participated in the focus groups discussions did not know the expectations of the raters for placing the two graphs as prompts because the task instruction did not say to describe the situation using the information from graphs; it was assumed that the test-takers would incorporate the information from the graphs. From the participants’ perspectives, it was their choice as to how much, if any, information to include from the graphs. The participants were quite surprised when the sample answer was handed to them and they were walked through some of the information taken from the graphs. For more clarity, it might be better for test-developers, such as the Eiken Foundation Japan to add a few words in the original sentence in the task instruction to avoid ambiguity on this matter. The task instruction needs to clearly transmit the idea that the information from graphs is available for a good reason to evaluate test-takers’ abilities to synthesize the information from both graphs.

8.4. Research Question (4) Effects of Test-Taking Strategies

The current study witnessed different test-taking strategies applied by the participants. To investigate what kinds of strategies were used and whether the use of strategies differed between the groups of participants with higher and lower writing proficiencies, gaze plots of some individual students for the first five minutes and the last five minutes of eye-recording data were compared. Due to the use of a range of strategies in planning, organizing, translating, monitoring and revising, it was hypothesized that major differences in the use of test-taking strategies would be observed during the first five and last five minutes of timed testing conditions. Next, survey questions were asked about what participants did before, during, and after their writing. The test-takers' answers revealed their behaviours and writing processes with respect to different types of strategies taken during the different stages. Finally, some of the key points were discussed further in the focus group discussions in order to scrutinize the rationale behind their actions. Given the complexity of the cognitive processes involved in integrated writing with multiple source texts and graphs, the discussions in this section will focus on whether proficient and less proficient test-takers apply different strategies when undertaking integrated writing tasks using multiple texts and graphs.

8.4.1 Evidence from Eye-Tracking data and analysis

According to Cohen (2006), one of the key test-taking strategies is for test-takers to organize the features of the language as necessary. Likewise, Yang and Plakans (2012) argued that the use of a discourse synthesis strategy, namely selecting, connecting and organizing skills, had a positive impact on test performance. To do this, the test-takers first needed to identify the test instructions, or what they were required to do for their essay.

As reported earlier, the Total Fixation Duration on the Task Instruction indicated a statistically significant difference between the upper and lower intermediate groups. Viewing the Task

Instruction for longer, however, did not necessarily indicate how carefully the test-takers have read and understood the task requirements in full. Some participants with less writing proficiency did not read the task instruction at all during the first five minutes of recording during Task 1. Similarly, the Gaze-plot of the Task Instruction during the first five minutes of recording showed, in some cases, that the participants who were less successful in writing did not go back to the Task Instruction.

Spending the first several minutes of fixation count, order and duration on the Task Instruction may be an indication of successful writing in TEAP Writing Test. It was only possible then to tell whether an individual participant had successfully engaged with the Task Instruction by looking at the gaze plots for how they engaged during the first five minutes. The same thing can also be done by reviewing the video clips. This confirmed that it was an important strategy for any test-taker to carefully review the required tasks and identify what kind of information they needed to be looking for before they began reading the source texts and information from the graphs.

8.4.2 Evidence from Surveys and Written Outputs

The participants were asked several questions concerning the use of test-taking strategies which may describe their behaviours before they began writing the essay. Such questions included whether they made an outline before writing, whether they thought about how many paragraphs the essay should consist of, whether they thought about what verb tense form to use, and whether they thought about which pronoun forms to use. Although no statistical significance was found between students at B2 or above and B1 or below in answering the questionnaires, the samples of written outputs found that the multi-paragraph essays were produced by more proficient students.

The researcher found the questionnaire responses were occasionally unreliable unless questions were carefully designed to allow triangulation. For example, approximately half of the participants claimed they thought about which verb tense to use, and the survey result from Q8_7 revealed that the participants at the B1 or below level mostly used the present tense while the participants at the B2 and above level claimed they have used the past tense. These self-reported results, however, were not necessarily reflected in the written outputs themselves. When the participants used the past tense in sentences, they were often directly borrowed from the source texts. The biggest problem arising from this was the subject-verb agreement. For less proficient writers, the subject-verb agreement was very difficult, as evident in the written outputs.

It is widely agreed that integrated writing requires more skills in terms of transforming knowledge by referring to multiple sources of information (Weigle, 2002). Thus, a key part of the test-taking strategies in integrated writing involves the use of source texts. In this regard, as Shi (2004) reported, different types of textual borrowings exist in essays, and the textual borrowings in test conditions pose significant issues for raters in terms of whether and how much textual borrowing should be allowed.

The current study looked at the written outputs for verbatim source use. Although this was found at all language levels, it was more prominent at A2 and below levels, not only in terms of the number of source borrowings but also the strings of words that were borrowed from the source texts. As shown in the examples, their essays were mainly patchy works copying the entire sentence, word-for-word, from some of the paragraphs. At B1 and B2 levels, some words and phrases were borrowed from the source texts, but it was different in the sense that the essay seemed to be developed following a logical order with cohesive language. From this study, it became clearer that a good use of source text depended on how well these sources were connected rather than how many words were borrowed from the source text.

8.4.3 Focus Group Discussions

The participants during the group discussions discussed what kind of strategies they used during the two tests. The participants then ended up sharing the information about what they were mostly concerned about. By surveying the trends, it was evident that the kind of test-taking strategies adopted by the participants with higher writing proficiencies were quite different from those used by students with lower writing proficiencies.

As previously reviewed, there were two distinct test-taking strategies applied by the students: test-wiseness strategies that have to do with the test-takers' prior knowledge about the test, and test-taking strategies that have to do with the use of language by the test-takers (Cohen, 2006.) Test-takers' strategies were affected by their prior knowledge of academic writing (e.g. organizational structure, time management) and their language usage (e.g. paraphrasing, use of synonyms, spelling and vocabulary). The current study was able to report that more proficient writers were more concerned about test-wiseness strategies and less proficient writers were more concerned about the test-taking strategies or the language usage.

Also, the test type affected the use of test-taking strategies among the test-takers. Two types of essays are an independent test that tests the knowledge-telling skills, and an integrated test, which tests skills for knowledge-transferring from the source information. Given the complexity of integrated writing tests, the subject has recently been rigorously studied by some researchers (see 2.4.2, e.g. Yu, Rea-Dickins, & Kiely, 2012; Yang, 2012; Xu & Wu, 2012). The focus group discussions in the present study also confirm that students with lower proficiency struggled to synthesize information from the source texts and graphs.

8.5 Research Question (5) Effects of Language Proficiencies

Finally, the current study made an attempt to explore the roles of language proficiency in integrated writing. The Mann-Whitney U test was applied between the B1 and below group (lower intermediate) and the B2 and above group (upper intermediate), based on the Aptis reading and writing scores. Also, the essays were marked using the TEAP rating rubrics. Finally, the comments and feedback in the focus groups discussions were classified according to the students' writing levels to find overall trends at each language level.

8.5.1 Evidence from Eye-Tracking data and analysis

There is growing interest among researchers in the process of reading into writing. The previous literature has suggested that being a better reader often implies better writing skills in an L1 context (e.g., Elley, 1991; McGinley 1992; and Wagner & Stanovich, 1996). More recent studies, however, have revealed this is not always the case for L2 learners (e.g., Flashive & Bailey, 1993; Grabe 2003).

Just as in these recent studies, the current research did not find any statistically significant differences between the upper and lower proficiency levels in respect to the groups of combined AOIs (i.e., Areas of Interests such as Task Instruction, Bar/Line Graph, Pie Chart, Newspaper Article and Letter-to-the Editor) in the first ten minutes, last ten minutes and all 40 minutes of the eye-tracking recordings. The findings implied only weak associations between the eye-tracking metrics, reading and writing scores. The multiple regression tests for the individual AOIs, however, found some statistically significant differences. For example, the current study found that a focus on the sub-headings of the source texts (i.e., Newspaper Article, Letter-to-The-Editor) could be a good indicator of a better performance in writing. It is hard to define what makes a "better" writer, but the eye-tracking data showed some correlations between the eye-tracking metrics (e.g., Total Fixation Duration) and the writing scores.

Although no strong relationship was found in the participants' reading and writing scores, the Mann-Whitney U tests indicated that the language proficiency levels did play a role in how long they viewed the Essay Composition area. Based on the results, it can be interpreted that the longer they spent on the essay composition, the more words they were likely to produce for the essay. This could be one indication of a better writer (Shanhan & Tirenay, 1990; McGinely, 1992; Wagner & Stanovich, 1996). However, quantitative analysis alone would not explain why spending more time in essay composition is an indicator for a successful writer.

8.5.2 Evidence from Surveys and Written outputs

In the survey, the participants were also asked whether they thought about what verb tense to use before writing (Q8-5), when they decided which verb to use (Q8-6), which tense they mostly used (Q8-6), whether they thought about what pronoun form to use for writing (Q8-7), and which pronoun forms they mostly used (Q8-8). The results indicated that only higher-level students thought about what verb tense to use before writing, but most other students did not. The most popular verb tense among the participants was the present tense. Although most students claimed that they thought about what pronoun forms to use, they could not seem to agree which personal pronouns they used most. When the written outputs of some students were closely examined, it was evident that less proficient participants (A2 and below A2 levels) tended to use the combination of present verb tense and first-person pronoun forms. It is evident, therefore, that there were different language uses among participants with different writing proficiencies and that this had an impact on the integrated writing tasks. These findings would be difficult to arrive at unless the essays and the survey results were examined closely.

8.5.3 Focus Groups Discussions

(1) Paraphrasing (A2 and A2 below): The last question in the focus group discussions best describes the participants' views on what role their language proficiency level played in their

approach to integrated writing tasks. The question was what they would do differently if they had to take the same test again. From the discussion, the keyword that most participants claimed to have learned as new knowledge was “paraphrasing.” Those at or below A2 level, in discussing what they would do differently, mentioned how they would in future paraphrase the source texts by commenting such as “summarize (the source texts) in my own words”, “use to avoid the repetitive words”, “Not borrowing the sentences from the source text”, and “Avoid repeating the same words”. Prior to the focus group discussions, the participants at the lower level had not realized that they needed to acquire paraphrasing skills.

(2) Paragraph writing (B1): For the same question, “paragraph writing” received attention among the B1 level students. Some participants voiced opinions such as “I want to improve paragraph writing skills” and “I learned about a paragraph writing.” Also, they were more concerned about the use of language, such as using the word “However” instead of “But”, and “Moreover” instead of “And.” The participants did not know previously that some colloquial language was not appropriate for academic writing. At a sentence level, they mentioned that they wanted to improve grammar and avoid “run-on sentences” which was one of the common mistakes that was brought to their attention in the focus groups discussions.

(3) Coherence (B2): The participants who had the highest writing proficiency showed more concerns about how to make their essay more coherent. “I learned the importance of the forms of academic writing” and “I understood I need to make an outline before (I) start composing an essay.” Moreover, she was more aware to give a proper credit to a source, such as “I need to be more cautious of where I got the idea from.” Equally, she also became aware that monitoring and revising are an essential process for a better essay.

Although many points were reviewed and discussed in the focus groups, some of the feedback comments were different and reflected their language levels. The discussions were useful in

understanding what concerns the participants had about integrated writing at different language levels.

8.6 Conclusions

This chapter has focused on how the current research can shed light on what affects the integrated writing process. Since the current findings are based on a small number of participants (N=38), the findings cannot be generalized. Nonetheless, this thesis has provided comprehensive results concerning the effects of (1) source texts, (2) graphs, (3) graph types, (4) test-taking strategies and (5) the roles of language proficiency in integrated writing. The findings of this study support the idea that a higher reading proficiency level cannot always guarantee higher performance in integrated writing in L2 contexts. Integrated writing revealed recursive visits to source texts and graphs in order to synthesize the information contained therein. One of the achievements was that while previous studies suggested that a higher number of average regressions was an indication of the test-taker having difficulties, in integrated writing, successful test-takers may be required to go back and forth over the texts in order to synthesize the information from the source texts as quickly as possible. Taken together, the results would seem to suggest that return sweeps and/or regressions may not necessarily be an indication of an unsuccessful performance in integrated writing. During the first ten minutes of the integrated writing processes (i.e., task representation, macro-planning and micro-planning) there were more statistically significant differences among the eye-metrics than during the last ten minutes, or the overall 40 minutes. These results suggest, therefore, that the first ten minutes are crucial for identifying the patterns or behaviours of the test-takers' performance in integrated writing.

Apart from these data analyses using the eye-tracking data, the cognitive-processing questionnaires and focus group discussions helped to gain an understanding of the decision-

making processes of the test-takers in many ways. The quality of the test-takers' essays depended on how they were able to connect ideas cohesively. Borrowing from the source texts was often an area of concern in determining the writing performance of the test-takers. The results from the cognitive-processing questionnaires showed that the test-takers' use of source texts varied according to their language proficiency levels. While more proficient students were able to borrow ideas and phrases and use them coherently, less proficient students tended to copy sentences and were not able to connect their ideas successfully.

In the current study, the graph types did not pose any difference in the level of cognitive challenge, but the eye-tracking data indicated some differences in the duration of time test-takers viewed each graph. With the cognitive-processing questionnaires and focus discussion groups, it became clear that the students lacked the English vocabulary or phrases to describe the trends in graphs. Without the qualitative analysis, it would have been difficult to gain meaningful insights into the academic knowledge required by the targeted students, such as for essay structures and English vocabulary to describe graphs and charts. Although there were several limitations in this research method, overall this thesis has been able to highlight the complexity of the issues.

Chapter 9 Conclusions

9.1 Introduction

The current research study attempted to gain insights into the cognitive strategies of L2 writers during reading-into-writing tasks. While many previous studies have focused mainly on source texts, written products and language proficiency, research on the cognitive processes of integrated writing tasks is still scarce. This study employed quantitative data to explore participants' cognitive processes of integrated writing tasks when multiple texts and information from graphs were used as prompts. The research instruments for this mixed-methods study included language assessments (i.e. the Aptis® and the TEAP®), the Tobii Eye-tracker for eye-tracking and keystroke analysis, a survey, and follow-up group discussions. Comparing cognitive patterns between more successful and less successful writers during the tasks using the data helped to provide an in-depth understanding of the effects of the source texts and information from graphs. The findings will be significant not only for students and educators but also for test developers, by providing information about what constitutes L2 integrated writing tasks. This section describes what kind of implications the present study has for each stakeholder.

9.1.1 Implications from the lack of task familiarity and from the administration of the eye-tracking experiment

Through the questionnaire and focus group discussions, findings revealed students' lack of familiarity with the TEAP Writing Task B. Although a sample test was given to the participants days in advance, with only one exception, the students did not have any prior experience of taking the TEAP writing test. This might have affected the results because of the condition that the participants were not familiar with the task. This was acknowledged as a part of the research limitation. Also, the administration of the eye-tracking experiment raised a concern after the

data were collected. Task 1 and Task 2 were not randomized to the participants. Therefore, the students performed better in the second round of similar tasks. A by-product of this non-randomized experiment was the finding of the familiarity effect. The participants quickly learned that the task instruction was almost identical for Task 1 and Task 2, and they were therefore able to go straight into the source texts and graphs rather than spending time on the task instruction to comprehend the task format. Since in typical test settings students would practise sample test(s) before taking the actual test, the findings from Task 2 could be said to give more accurate results regarding test performance than Task 1.

9.1.2 Implications for students with low reading proficiency

First and foremost, it is critical for students to be familiarized with the test instructions and the required tasks for each section in advance. Students who have lower reading proficiency would most probably benefit the most from understanding the task requirement and the purpose of the essay in advance. By understanding what kind of essay is required for writing tasks (e.g. an integrated essay or an independent essay), the student will be able to determine what pronoun form and verb tense to use.

Secondly, this study unveiled that the less successful participants did not read the source texts in order and jump straight to writing the essay. It is therefore important for them to develop expeditious reading skills, which may be a different type of reading than that which the students were accustomed to using for reading comprehension tests (i.e. careful reading). It is important that the students can skim quickly through the texts and identify where the main points while understanding the content of the reading.

Thirdly, the less able students often showed longer strings of words or sentences copied directly from the source texts. If the information had to be synthesized and reported, it would usually mean that students need to know how to paraphrase the sentences in their words. The simplest

way to do this is to replace a word with synonyms. Also, the sentence structure can be changed. By combining two sentences with a use of a conjunction, for example, the essay could become more logical and coherent. When the less proficient students reported information they often omitted the source of this information, giving the impression that it was their own opinion, which created a confusion in respect to whether the essay was an independent essay. Understanding the basic academic knowledge of how to paraphrase sentences and cite sources would be a key area of improvement for the less successful writers.

Moreover, the less able students had challenges in describing the information from graphs. Knowing different types of graphical information requires different words and phrases for describing trends, and therefore prior training on phrases to describe a trend in information from graphs would be essential, particularly line graphs. Such prior knowledge would certainly have a positive impact on the details that the students would be able to include in their essays.

Finally, the essay length and the number of paragraphs was a statistically significant factor between less and more successful participants. Participants in this study who had lower writing proficiency did not seem to know what constitutes a paragraph. Making an outline with a keyword later developed as a topic sentence for a paragraph would be helpful in guiding them to write a multi-paragraph essay.

9.1.3 Implications for students with higher writing proficiency

It is of utmost importance for any test-taker to understand the required task by reviewing the Task Instruction accurately. Most test-takers in this study have shown that they carefully reviewed the task and understood the aim of the essay, yet the questionnaire and focus groups discussions found that only a few students were strategically engaged in pre-writing processes such as making an outline before writing. To effectively manage their time in timed tests, it is

recommended that the students make a list of required tasks and an overall organizational structure for the essay.

Students in this group were more concerned about their essay being coherent and logically sound. When they were asked what the most challenging task was, more successful participants answered, “giving the reasons for the choice he or she made.” For those students, improving the lexical knowledge on signposting or discourse markers would have a positive impact on the coherence of the essay.

Finally, those students with high proficiency showed some signs that they were engaged in monitoring and revising processes. Pauses at the end of each paragraph also suggested that the test-taker was reviewing and monitoring their written outputs. When the essay is complete, there should be more attention paid to the overall written output, going over some points and making a quick change if this could be done in the last few minutes or so.

9.1.4 Implications for teachers

The day has come when teaching a sentence grammar and structure by translating between two languages cannot ensure readiness for academic writing. Now that integrated writing has become a norm for language assessment, teachers should focus on developing reading skills in combination with paraphrasing skills especially since L2 students often face difficulties in writing tasks that require inferencing (Grabe & Zhang, 2013).

The days are also gone where successful writing can be taught by telling students to memorize a frequently-used word or phrases in index cards. Whereas single sentence compositions were often the focus of writing class in middle and high schools in Japan, these days, the concept of writing in paragraphs has become more widely acknowledged by English teachers. In most cases, however, writing instructions focus on independent writing on chosen topics. Even when students are trained to write an excellent independent essay, this study showed that this does

not guarantee that they would produce an excellent integrated essay as the latter is much more cognitively challenging.

Paraphrasing, which is a fundamental skill for synthesizing information for integrated writing, is often not used in English composition instructions. Teachers often prefer doing an exercise with an answer key or teaching mechanics such as grammar or translation. When it comes to paraphrasing, there are many ways to convey the same information. Having a sample sentence, or two, showing different levels of language proficiency would probably give students an idea that the skilled writer can produce a better sentence instead of one “good” answer. To cope with the increasing demands of teaching academic writing skills, teachers themselves also need to be well equipped with some additional training and resources.

Finally, the results of this study may offer an insight into how teachers should approach their students differently. In particular, this study examined what influences integrated writing tasks between more and less skilled writers. The study found that students face different types of challenges depending on their reading and writing proficiency levels, as reported in the findings and discussion chapters. For example, the use of first person pronouns in the integrated tasks was found among the less successful test-takers, and could be an indication that they needed to understand the task requirements for the essays. In such cases, teachers would need to make sure students understand and identify the purpose of the essay and work on an appropriate use of language, given the context.

9.1.5 Implications for test-developers

The TEAP test measures the levels of students between A2 and B2 learners. TEAP Task B, which is an integrated writing that uses multiple texts and information from graphs, can accurately measure what the test developer has intended at these language proficiency levels. In my opinion, however, it is still unclear whether the integrated writing tasks with multiple

texts and graphs can really differentiate the level of writing skills at the lowest end of the spectrum. This is because (a) most students at the Below A2 level demonstrated limited abilities to monitor and revise their written outputs, (b) the number of words produced were limited, and (c) they were not able to fulfil the required tasks of describing and summarizing the main points in the source texts and information from graphs. Most critically, it was difficult to distinguish between the lower level students when they borrowed and copied some phrases and sentences directly from the source texts, as often happened. Perhaps, this was why some of the tests use a combination of both independent and integrated writings for assessment.

In the TEAP Test, however, both Task A and B are integrated writing. In my opinion, Task A which is a summative task of a paragraph, should be replaced with an independent writing task on an academic topic. Since a summative task is already an integral part of Task B, as students are required to summarize the main points of discussion, it can be said that the work in Task A is redundant. Introducing an independent writing task as a first essay would help distinguish the writing abilities of participants working at level A2 and below. If Task A needs to remain as it is, another option to consider is to add a short independent writing task before Task A.

Test developers assume that potential candidates (test-takers) are familiar with the test formats before taking such high-stakes tests as those used for college admission processes. In this experiment, the participants, except for one student, had not taken the TEAP test before the eye-recording experiments, although they were given a sample question in advance. This highlighted how less skilled learners failed to identify the purpose of the essay and ended up writing another type of essay. For the same reason, even some students who had higher proficiency sometimes did not include information from graphs in their essay. It could be that the Task Instruction does not explicitly state that test-takers are expected to describe the situation “by referring to the information on the graphs.” By adding such a phrase, it would

become clearer to test-takers that they must describe the information from graphs and that there is a reason they were given this information.

9.1.6 Implications for researchers

In addressing what affects the cognitive processes of writing, the present study employed an eye-tracker, which gave an insight into what kind of eye-metrics can be helpful for us to understand reading-into-writing processes. Given that few publications are yet available for the use of eye-tracking technology in the context of understanding the cognitive processes of integrated writing among L2 learners, this study may provide some sense of urgency that more studies need to be conducted to validate the findings.

In essence, the eye-tracker with *Tobii Studio* is a powerful tool for a non-engineering researcher like myself, with a user-friendly interface. Careful planning of the experiments and some technical assistance or training are essential, however, for successful experimental outcomes using this powerful tool. Given the cost of the lease, it is still not a research instrument that anyone can have access to immediately, but hopefully, with some technological advances, it will become more accessible to language researchers in the near future.

The current study only looked at what affects integrated writing processes under a timed testing condition using the TEAP Task B as a model. Although the present study examined different types of information from graphs, it did not discuss various kinds of source texts. In other words, the types of essay and kind of language produced depend on the task requirements and source information. If other types of documents were given as a source text instead of a newspaper article which described a problem and solution, it could have shown other effects as well. Thus, future studies may be needed to investigate the effects of different types and genres of source texts during integrated processes of writing.

9.2 Limitations

Just as with any other study, the current study encompasses several limitations. To reduce these limitations for future research, three main concerns are reported with regards to (1) Sampling, (2) Research Instruments, and (3) Research Procedures and Analysis.

9.2.1 Sampling

The participants were purposively chosen from four local high schools in western parts of Japan. Although an attempt was made to include both public and private schools, those schools that participated in this research had a strong English curriculum, and the teachers who were involved had ensured that the participating students were comfortable taking the experiments on the computer screen (for eye-recordings) and with some keyboard typing experiences. Consequently, the participants mostly had B1 and B2 level reading and writing proficiencies. Ideally, the participants would have an equal number of A, B1, B2, and C (1)-level students.

The eye-tracking experiments alone took approximately 120 minutes for each participant, during which time each participant had to take a similar test for 40 minutes. The tests could only be conducted one at a time, so the experiment was only possible during the spring break or after school hours when a classroom was not in use. Due to the length of the experiment, the participants looked overwhelmingly tired after one test. Some sweets (candies, drinks) were provided during a five-minute break. The survey questions were filled immediately after the two tests. By then, the participant looked exhausted, and there was no room for conducting the stimulated-recall interview as initially predicted. In an ideal situation, however, it is best to carry out the interviews when memories are fresh. Given the time limitations and the conditions; however, this study chose to include focus groups discussions a few weeks after the experiments.

9.2.2 Research instruments

Before using the eye-tracker, each participant spent five minutes carrying out the calibrations needed to make sure that gazes were captured on screen. This was done automatically by having the participants follow a red dot on the screen. Some participants who normally wear contact lenses came with eyeglasses. In one case, a participant forgot to bring their own eyeglasses and had to take the test without the glasses he said he would normally need. As a result, the gaze weight percentages were much lower than hoped for. This could have been prevented by announcing in advance that they should wear contact lenses if they usually wore them.

9.2.3 Research procedures and analysis

According to the technical staff at *Tobii Studio*, most of their clients had used the eye-tracker on a much smaller scale (N=6) and for a much shorter time (10-20 minutes each). These researchers come from different disciplinary backgrounds (psychology, clinical researchers, etc.) who had their eye-tracking experiment tightly controlled even to the extent of a size and type. The aim of the current experiment was to study eye-recording findings in the application. Just as in the real test, therefore, there were different font sizes for different purposes. In calculating saccades and regressions based on how many pixels was worth each letter font, it was not as precise as it should have been. Also, the lines between the sentences were not wide enough. These conditions would be very difficult to fulfil when testing something like a real test. To minimize the effects of conditions on eye-recordings for analysis, one needs to be aware that paying attention to these details can be crucial for some experiments depending on the purpose. Some experience of using the eye-tracker with data analysis, however, gives a better sense of how to pay attention to these details when conducting research with this system.

Also, there was an issue of familiarity with regards to the task prompt. Most of the participants had not taken the TEAP test before. Although a sample test was shown, the participants did not seem fully confident of what they were supposed to do. This might have affected the results.

Not long ago, when this research was still at the proposal stage, the research plan was presented at a language assessment conference. At a poster presentation, some experts who have done similar research using the eye-tracker had commented that this study was very ambitious, and others mentioned the full scope of this study might be worth two doctoral studies. Having conducted this study for my thesis, now it is comprehensible why they made these comments. This thesis made me humble thinking what hard work it must have been for these researchers to pioneer their studies in applied linguistics. With this inspiration, my endeavour will continue.

References

- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 2). *Language teaching*, 35(2), 79-113.
- Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, 18(3), 191-208.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford and New York: Oxford University Press.
- Bachman, L.F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453-476.
- Bax, S. (2013a). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, 30(4), 441-465.
- Bax, S. (2013b). Readers' cognitive processes during IELTS reading tests: evidence from eye tracking. British Council, ELT Research Papers 13-06.
- Bax, S., & Weir, C. J. (2012). Investigating learners' cognitive processes during a computer-based CAE Reading test. *Research Notes*, (47), 3-14. Cambridge ESOL.
- Bereiter, C., Burtis, P. J., & Scardamalia, M. (1988). Cognitive operations in constructing main points in written composition. *Journal of memory and language*, 27(3), 261-278.
- Blanchard, H. E., Pollatsek, A., & Rayner, K. (1989). The acquisition of parafoveal word information in reading. *Perception & Psychophysics*, 46(1), 85-94
- Bourke, B. (2014). Positionality: Reflecting on the research process. *The Qualitative Report*, 19(33), 1-9.
- Braine, G. (1989). Writing in science and technology: An analysis of assignments from ten undergraduate courses. *English for Specific Purposes*, 8, 3-15.
- Bridges, G. (2010). Demonstrating cognitive validity of IELTS academic writing task 1. *Research Notes*, 42, 24-33.
- British Council (2015). *Aptis*. Retrieved from <http://www.britishcouncil.org/Aptis>
- Brunfaut, T., & Révész, A. (2015). The Role of Task and Listener Characteristics in Second Language Listening. *TESOL Quarterly*, 49(1), 141-168.

- Brunfaut, T. & McCray, G. (2015). Looking into test-takers' cognitive processes whilst completing reading tasks: a mixed-method eye-tracking and stimulated recall study. ARAGs Research Reports Online, Vol. AR/2015/001. London: The British Council.
- Carpenter, P. A., & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2), 75.
- Carson, J. (2001). A task analysis of reading and writing in academic contexts. In: D. Belcher & A. Hirvela (Eds.), *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 246–270). Ann Arbor, MI: University of Michigan Press
- Carswell, C., Emery, C., & Lonon, A. M. (1993). Stimulus complexity and information integration in the spontaneous interpretations of line graphs. *Applied Cognitive Psychology*, 7(4), 341-357.
- Chan, S. H. C. (2013) 'Establishing the Validity of Reading-into-writing Test Tasks for the UK Academic Context'. Unpublished PhD thesis. University of Bedfordshire.
- Chan, S. H., Wu, R. Y., & Weir, C. J. (2013). Examining the context cognitive validity of the GEPT Advanced Writing Task 1: A comparison with real-life academic writing tasks. (LTTC-CRELLA Collaboration Project-RG03). Taipei: The Language Training and Testing Center.
- Chen, Y.-S., & Su, S.-W. (2012). A genre-based approach to teaching EFL summary writing. *ELT journal*, 66(2), 184-192.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, 6(4), 284.
- Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307-331.
- Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education*. Routledge. London
- Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.): Sage publications. London
- Creswell, J. H., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publishing.
- Crotty, M. (1998). *The foundations of social science: Meaning and perspective in the research process*. Thousand Oaks, CA: Sage Publishing.
- Cumming, A., Rebuffot, J., & Ledwell, M. (1989). Reading and summarizing challenging texts in first and second languages. *Reading and writing*, 1(3), 201-219.

- Cumming, A. (2001). Learning to write in a second language: Two decades of research. *IJES, International Journal of English Studies*, 1(2), 1-23.
- Cumming A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005) Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10 (1), 5–43.
- Cumming, Lai, & Cho. (2016). Students' writing from sources for academic purposes: A synthesis of recent research. *Journal of English for Academic Purposes*, 23, 47-58.
- Delaney, Y. A. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, 7(3), 140-150.
- Eiken Foundation of Japan, http://www.eiken.or.jp/teap/construct/rating_crit.html
- Eiken Foundation of Japan. Press Release on TEAP, 2016. The Eiken Foundation of Japan. (2016). Retrieved from <https://www.eiken.or.jp/teap/info/2016/pdf/20161122>
- Elley, W. B. (1991). Acquiring literacy in a second language: The effect of book-based programs. *Language learning*, 41(3), 375-411.
- Emig, J. (1971). The composing processes of twelfth graders. IL: National Council of Teachers of English
- Emig, J. (1983). *The web of meaning: Essays on writing, teaching, learning, and thinking*. NJ: Boynton / Cook Publication.
- Field, J. (2004). *Psycholinguistics: The Key Concepts*, London: Routledge.
- Flahive, D., & Bailey, N. (1993). Exploring reading/writing relationships in adult second language learners. In J. Carson & L. Leki (Eds.), *Reading in the composition classroom: Second language perspectives*, 128-140. Boston: Heinle & Heinle.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication*, 32(4), 365-387.
- Folch-Lyon, E., & Trost, J. F. (1981). Conducting focus group sessions. *Studies in family planning*, 443-449.
- Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In *International Conference on Theory and Application of Diagrams* 18-30. Springer, Berlin, Heidelberg.
- Frenck-Mestre, C. (2002). An on-line look at sentence processing in the second language. *Advances in Psychology*, 134 (9), 217-236.

- Frenc-Mestre, C. (2005). Ambiguities and anomalies: What can eye-movements and event-related potentials reveal about second language sentence processing? *Handbook of bilingualism*, 268-284.
- Frenc-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: a review of methodologies and experimental findings. *Second Language Research*, 21(2), 175-198.
- Gebril, A. M. (2006). Independent and integrated academic writing tasks: A study in generalizability and test method (Doctoral dissertation, University of Iowa).
- Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Language Testing*, 26(4), 507-531.
- Gebril, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: A generalizability analysis. *Assessing Writing*, 15(2), 100-117.
- Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 7(1), 47-84.
- Grabe, W., & Kaplan, R. (1996). *Theory and practice of writing*: New York: Longman.
- Grabe, W. (2001). Notes toward a theory of second language writing. In T. Silva & P. K. Matsuda (Eds.), *On second language writing*, 39-57.
- Grabe, W. (2003). Reading and writing relations: Second language perspectives on research and practice. In B Kroll (Ed.), *Exploring the dynamics of second language writing* (242-259). Cambridge: Cambridge University Press.
- Grabe, W., & Zhang, C. (2013). Reading and writing together: A critical component of English for academic purposes teaching and learning. *TESOL Journal*, 4(1), 9-24.
- Grosvenor, T. (2000). *Qualitative research in the transport sector* (No. E-C008). *Proceedings of an International Conference on Transport Survey Quality and Innovation, 1997 Conference* (IIA1-IIA17). Grainau; Germany.
- Green, A. (2014) "The Test of English for Academic Purposes (TEAP) Impact Study: Report 1-Preliminary Questionnaires to Japanese High School Students and Teachers." *Tokyo: Eiken Foundation of Japan*.
- Hayes, J. R. (2012). Modeling and Remodeling Writing. *Written Communication*, 29(3), 369-388.
- Hirose, K. (2003). Comparing L1 and L2 organizational patterns in the argumentative writing of Japanese EFL students. *Journal of second language writing*, 12(2), 181-209.

- Horowitz, D. (1986). What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20 (3), 445-462.
- Houser, C., & Thornton, P. (2004). Japanese college students' typing speed on mobile devices. In *Wireless and Mobile Technologies in Education, 2004. Proceedings. The 2nd IEEE International Workshop on* (pp. 129-133). IEEE.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- Hyland, T. A. (2009). Drawing a line in the sand: Identifying the border zone between self and other in EL1 and EL2 citation practices. *Assessing Writing*, 14(1), 62-74.
- IELTS Research Report, Test-Takers Performance (2013). *Mean band scores for the most frequent places of origin (Academic)* Retrieved from http://www.ielts.org/researchers/analysis_of_test_data/test_taker_performance_2013.aspx
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field methods*, 18(1), 3-20.
- Jacobson, J. Z., & Dodwell, P. C. (1979). Saccadic eye movements during reading. *Brain and Language*, 8(3), 303-314.
- Kaplan, R. B. (1966). Cultural thought patterns in inter-cultural education. *Language learning*, 16(1-2), 1-20.
- Khalifa, H., & Weir, C. J. (2009). *Examining writing: Research and practice in assessing second language writing*. Cambridge: Cambridge University Press
- Kim, S., Dong, Z., Xian, H., & Yi, J. (2012). Does an Eye-tracker Tell the Truth about Visualizations? Findings while Investigating Visualizations for Decision Making. *IEEE Transactions on Visualizations and Computer Graphics*, 18(12), 2421-2430.
- Kitzinger, J. (1995). Qualitative research. Introducing focus groups. *BMJ: British medical journal*, 311(7000), 299.
- Koizumi, R., & Nakamura, K. (2016). Factor structure of the Test of English for Academic Purposes (TEAP®) test in relation to the TOEFL iBT® test. *Language Testing in Asia*, 6(1)
- Kubota, R. (1998). An investigation of Japanese and English L1 essay organization: Differences and similarities. *Canadian modern language review*, 54(4), 475-508.
- Kunnan, A. J. 1996. Connecting fairness with validation in assessment. In A. Huhta, V. Kohonen, L. Kurki-Suonio and S. Luoma (eds.) *Current developments and alternatives in language assessment*. Jyvaskyla, Finland: University of Jyvaskyla. 85-105.

- Leki, I. & Carson J. (1997). Completely Different Worlds: EAP and the Writing Experiences of ESL Students in University Courses. *TESOL Quarterly*, 31(1), 39-69.
- Lewkowicz, J. (1994). Writing from Sources: Does Source Material Help or Hinder Students' Performance? presented at International Language in Education Conference, Hong Kong, 1993.
- Matsuda, P. K. (2003). Second language writing in the twentieth century: A situated historical perspective. In B Kroll (Ed.), *Exploring the dynamics of second language writing*, 1, 15-34. Cambridge: Cambridge University Press
- McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, 35(1), 51-73.
- McCulloch, S. (2013). Investigating the reading-to-write processes and source use of L2 postgraduate students in real-life academic tasks: An exploratory study. *Journal of English for Academic Purposes*, 12(2), 136-147.
- McKinley, W. (1992). The role of reading and writing while composing from sources. *Reading Research Quarterly*, 27(3) 227-248.
- McNamara, T. 1998. Policy and social considerations in language assessment. In W. Grabe, *et al.* (eds.) *Annual Review of Applied Linguistics*, 18. *Foundations of second language teaching*. New York: Cambridge University Press. 304-319
- Messick, S. (1989). *Validity*. In R. L. Linn (Ed.), *Educational measurement*. 13-103. New York: Macmillan.
- Mickan, P., Slater, S., & Gibson, C. (2000). Study of response validity of the IELTS writing subtest. *IELTS Reports 3*. British Council, 29-48.
- Nakamura, K. (2014). Examination of Possible Consequences of a New Test within the Context of University Entrance Exam Reform in Japan. Paper Presented at The 36th International Language Testing Association at University of Amsterdam.
- Nishikawa, M. (2015). “*TEAP Writing Teachers’ Manual*” Retrieved from https://www.eiken.or.jp/teap/info/2015/0901_01.html.
- Oshima, K., & Muramatsu, Y. (2015). Current situation and issues related to ICT utilization in primary and secondary education. *Fujitsu Scientific and Technical Journal*, 51(1), 3-8.
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13(2), 111-129.
- Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, 8(4), 252-266.

- Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17(1), 18-34.
- Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217-230.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26-56.
- Power, M. K., & Gendron, Y. (2015). Qualitative research in auditing: A methodological roadmap. *Auditing: A Journal of Practice & Theory*, 34(2), 147-165.
- Eiken Foundation of Japan. Press Release on TEAP, 2016. The Eiken Foundation of Japan. (2016). Retrieved from <https://www.eiken.or.jp/teap/info/2016/pdf/20161122>
- Rabiee, F. (2004). Focus-group interview and data analysis. *Proceedings of the nutrition society*, 63(04), 655-660.
- Ransdell, S., Levy, C. M., & Kellogg, R. T. (2002). The structure of writing processes as revealed by secondary task demands. *L1-Educational Studies in Language and Literature*, 2(2), 141-163.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.
- Révész, A. (2011). Task Complexity, Focus on L2 Constructions, and Individual Differences: A Classroom-Based Study. *The Modern Language Journal*, 95(s1), 162-181.
- Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning*, 62(1), 93-132.
- Révész, A., & Han, Z. (2006). Task content familiarity, task type and efficacy of recasts. *Language Awareness*, 15(3), 160-179.
- Ritchie, J. (2003). *Qualitative Research Practice: A Guide for Social Science Students and Researchers* (J. J. Ritchie and J. Lewis Eds.). London: SAGE Publication, Ltd.
- Ritchie, J & Lewis, J & Elam, G. (2003). Designing and selecting samples. *Qualitative research practice: A Guide for Social Science Students and Researchers 2* (J. Ritchie and J. Lewis Eds.). London: SAGE Publication, Ltd.
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenburg (Ed.), *Advances in applied psycholinguistics*, (Vol.2). 142-175. Cambridge: Cambridge University Press.

- Shanahan, T., & Tierney, R. J. (1990). Reading-writing connections: The relations among three perspectives. In *National Reading Conference Yearbook*, 39, 13-34. National Reading Conference.
- Shaw, S. D., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing* (Vol. 26). Cambridge: Cambridge University Press.
- Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21(2), 171-200.
- Shi, L. (2006). Cultural backgrounds and textual appropriation. *Language Awareness*, 15(4), 264-282.
- Shi, L. (2012). Rewriting and paraphrasing source texts in second language writing, *Journal of Second Language Writing* 21 (2), 134-148.
- Somekh, B., & Lewin, C. (2011). *Theory and methods in social research*: Sage.
- Spivey, N. N. (1990). Transforming Texts Constructive Processes in Reading and Writing. *Written communication*, 7(2), 256-287.
- Suvorov, R. (2015). The Use of Eye Tracking in Research on Video-Based Second Language (L2) Listening Assessment: A Comparison of Context Videos and Content Videos. *Language Testing*, 32(4), 463-483.
- Test and Score Data Summary for TOEFL iBT® Tests (2014). [Table 16] *TOEFL iBT Total and Section Score Means I All Examinees, Classified by Geographic Region and Native Country* Retrieved from https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf.
- Underwood, N. R., & Mc. Conkie, G. W. (1985). Perceptual span for letter distinctions during reading. *Reading Research Quarterly*, 20(2), 153-162.
- Underwood, P. (2010). A comparative analysis of MEXT English reading textbooks and Japan's National Center Test. *RELC Journal*, 41(2), 165-182.
- Wagner, R. K., & Stanovich, K. E. (1996). Expertise in reading. *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games*, 189-325. New York: Psychology Press.
- Watanabe, Y. (1996). Investigating washback in Japanese EFL classrooms: Problems and methodology. *Australian Review of Applied Linguistics*, 13, 208-239.
- Watanabe, Y. (2001). Read-to-write tasks for the assessment of second language academic writing skills: investigating text features and rater reactions. Unpublished doctoral dissertation, University of Hawaii.

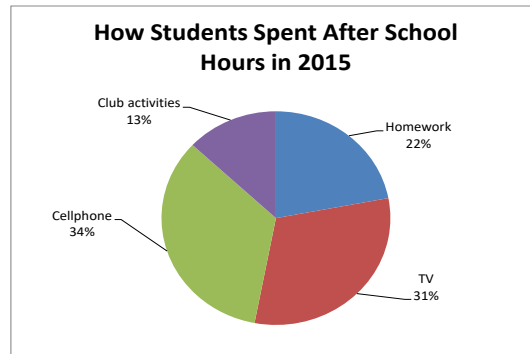
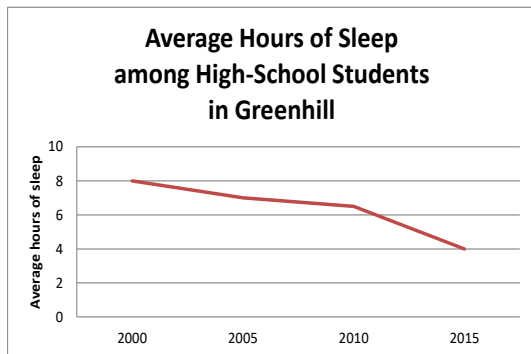
- Watanabe, Y. (2013). The National Center Test for University Admissions. *Language Testing*, 30(4), 565-573.
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing writing*, 9(1), 27-55.
- Weigle S. C. (2005) Second Language Writing Expertise. In: Johnson K. (eds) Expertise in Second Language Learning and Teaching. Palgrave Macmillan, London.
- Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21(2), 118-133.
- Weir, C J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.
- Weir, C. J. (2014). A Research Report on the Development of the Test of English for Academic Purposes (TEAP) Writing Test for Japanese University Entrants.
- Weir, C.J, Chan S. H.C., & Nakatsuhara F. (2014). Examining the Criterion-Related Validity of the GEPT Advanced Reading and Writing Test: Comparing GEPT with IELTS and Real-life Academic Performance. (LTTC-CRELLA Collaboration Project-RG01). Taipei: The Language Training and Testing Center.
- Xi, X. (2010). Aspects of performance on line graph description tasks: Influenced by graph familiarity and different task features. *Language Testing*, 27(1), 73-100.
- Xu, Y., & Wu, Z. (2012). Test-taking strategies for a high-stakes writing test: An exploratory study of 12 Chinese EFL learners. *Assessing Writing*, 17(3), 174-190.
- Yang, H.-C. (2012). Modeling the relationships between test-taking strategies and test performance on a graph-writing task: Implications for EAP. *English for Specific Purposes*, 31(3), 174-187.
- Yang, H. C. (2014). Toward a Model of Strategies and Summary Writing Performance. *Language Assessment Quarterly*, 11(4), 403-431.
- Yang, S.-N., & McConkie, G. W. (2001). Eye movements during reading: A theory of saccade initiation times. *Vision research*, 41(25), 3567-3585.
- Yang, H. C., & Plakans, L. (2012). Second Language Writers' Strategy Use and Performance on an Integrated Reading-Listening-Writing Task. *TESOL Quarterly*, 46(1), 80-103.
- Yu, G. (2005). *Towards a model of using summarization tasks as a measure of reading*

- comprehension*. Unpublished PhD thesis, University of Bristol, Bristol.
- Yu, G. (2008). Reading to summarize in English and Chinese: A tale of two languages? *Language Testing*, 25(4), 521-551.
- Yu, G. (2009). The shifting sands in the effects of source text summarizability on summary writing. *Assessing Writing*, 14(2), 116-137.
- Yu, G., He, L., & Isaacs, T. (2017). The cognitive processes of taking IELTS academic writing task one: From concurrent think aloud to eye-tracking with stimulated recall interview.
- Yu, G, Rea-Dickins, P & Kiely, R. (2011). The cognitive processes of taking IELTS academic writing task one. *IELTS Reports 11*. British Council, 373-449.
- Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly*, 10(1), 110-114.
- Yu, G., & Lin, S. W. (2014). A Comparability Study on the Cognitive Processes of Taking Graph based GEPT-Advanced and IELTS-Academic Writing Tasks. *LITC-GEPT Research Report RG-02*.
- Zhu, X., Li, X., Yu, G., Cheong, C. M., & Liao, X. (2016). Exploring the Relationships Between Independent Listening and Listening-Reading-Writing Tasks in Chinese Language Testing: Toward a Better Understanding of the Construct Underlying Integrated Writing Tasks. *Language Assessment Quarterly*, 13(3), 167-185.

Appendixes

Appendix A (Task 1) and Sample Essay (1)

Your teacher has asked you to write an essay for class using the information below. Describe the situation concerning schools in Greenhill and summarize the main points about the solutions that have been suggested. In your conclusion, say which of the solutions you think would work the best based on the reasons given. You should write about 200 words.



Education News

A new report found a worrying trend concerning teenagers in Greenhill. Mike Parker, the principal at North Greenhill High School, thinks that changing the daily routine after school is the first step in dealing with the problem. Parker talked about his ideas at a recent meeting with parents. "We need to educate our children about the importance of sleep," Parker said. "Average teenagers need about nine hours of sleep each night to feel well rested," he said. Parker wants to hold a special session for students to discuss this issue.

However, Parker suggested other steps are also necessary. For example, the government is considering changing school start times in high school. One reason is that high-school students tend to stay up late doing homework and other activities. "To be more realistic," said Parker, "this might be the only way to solve the problem. By delaying the start of morning classes by an hour, we can ensure that students gain extra sleep." Many teachers agree that students perform better when they are well-rested.

LETTER TO THE EDITOR

Dear Editor,

I am very concerned about the recent trend among adolescents regarding sleep. Based upon my many years of experience as a school nurse, I would like to offer some advice. First, exercise can help young people fall asleep faster and sleep more soundly. According to recent studies, it is better to work out earlier in the day than in the evening. So, it might be helpful to schedule morning exercise as a part of regular school activities in order to encourage adolescents to go to sleep before midnight.

I also believe the government should take action to resolve the situation. Some experts recommend adjusting the school times to fit the biological clocks of adolescents. Research has shown that students in classes with a later start time were twice as productive as those in normal classes. It may be worth considering changing school start times in Greenhill high schools.

I'm confident that parents and local educators will find effective ways to address this issue.

Sincerely,
Sarah Case

Task 1 [Model Essay Answer]

Teenagers in Greenhill are getting less sleep over the years. According to the statistics, the average hours of sleep among high-school students dropped by half from 8 hours to 4 hours between 2000 and 2015. The survey in 2015 indicates that 34 percent of after-school hours were spent on cell phones, followed by 31 percent on TV, 22 percent on homework and 15 percent on club activities.

Mike Parker, the Principal of North Greenhill High School, made two suggestions. First, he suggested educating his students and their parents about the importance of sleep. The second suggestion was to delay the school start times in high schools. This way, he said we can ensure the students get enough sleep every day.

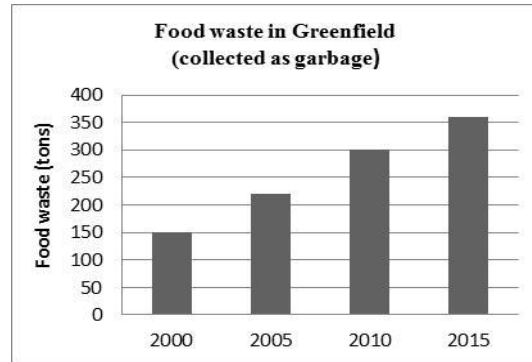
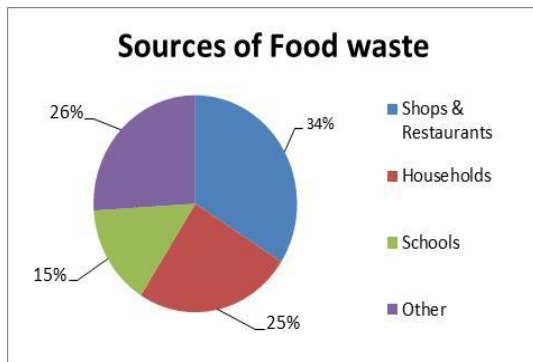
The idea of changing the school times to fit the “biological clocks of adolescents” was also echoed by Sarah Case, who is a school nurse. In addition, she said introducing morning exercises might help the students fall asleep faster at night and sleep more soundly.

In my opinion, having morning exercises would be the best solution. I believe delaying the start of morning classes would only cause the students to stay up much later than they already do. This is not a healthy way to solve the problem.

(207 words)

Appendix B (Task 2) and Sample Essay (2)

Your teacher has asked you to write an essay for class using the information below. Describe the situation concerning food waste in Greenfield and summarize the main points about the solutions that have been suggested. In your conclusion, say which of the solutions you think would work the best based on the reasons given. You should write about 200 words.



City Council News

Members of Greenfield City Council met yesterday to discuss the issue of food waste in the city. Mike Parker, the head of the city's waste collection unit expressed his concern. "The city's garbage-collection service has a very heavy workload," he said. He also pointed out that collecting and disposing of food waste has led to increased costs for the city. Parker thinks that targeting restaurants is the first step in dealing with the problem. "We need to encourage restaurants to throw away less food," he suggested, adding that "the only way to do this is to hit them where it hurts, in their pockets." To do this, he suggested introducing a special tax that would be calculated based on the amount of food waste a business produces.

Parker also put forward another idea. "The food that is thrown away by grocery stores cannot be consumed by humans," he said, "but it could be used for a different purpose." "I have a friend who is a farmer, and he told me he'd welcome the unwanted food to feed his animals." Parker argued that this would be a good way to reduce costs for the council and benefit many local farmers. The city council will discuss this issue further at its next meeting.

LETTER TO THE EDITOR

Dear Editor,

I am very concerned about this recent trend in Greenfield. In my work as an environmental health officer I sometimes visit local stores. I'm always shocked by how much unsold food they throw away. I know that it is difficult for them to judge how much they will sell each day, but I believe we should try and find a way to reuse some of this food.

I also believe families in Greenfield must change their habits. When I visit my friends' homes, I'm surprised by how much food they discard without even thinking. I suggest it would be a good idea to have classes in schools to teach children about this issue. If children learn from a young age that wasting food is bad, they will grow up to teach their own children the same. This will also please parents, as their children will learn to throw away less food. I read a recent report that showed that families can reduce their monthly spending by twenty percent simply by wasting less food.

I'm confident that the city council will find effective ways to address this issue, and I hope my ideas will be useful for them.

Sincerely,
Sarah Case

Task 2 [Model Essay Answer]

Greenfield city has a problem with food waste. The amount of food disposal in the city rose from 150 tons to 350 tons between 2000 and 2015. In 2015, the biggest proportion of such garbage came from shops and restaurants which accounted for 34 percent, followed by 25 percent from household and 15 percent from school.

The article introduced two possible solutions proposed by Mike Parker who is the head of the city's waste collection. First, he suggested imposing a special tax on the disposal of food waste in the business sector. His second suggestion was to utilize the food waste as animal feed.

Meanwhile, Sarah Case, an environmental health officer, proposed changing the habits of discarding food waste in households by educating children. She said that families can save up to 20 percent of monthly spending by reducing food waste.

In my opinion, encouraging industries to reduce the amount of food waste would be the best solution. If the city imposes a tax on shops and restaurants, this would probably bring an immediate result on the reduction of the food waste. On the other hand, educating children might take a long time to have a desirable effect.

(198 words)

Appendix C TEAP Writing test, Task B Scoring Rubrics

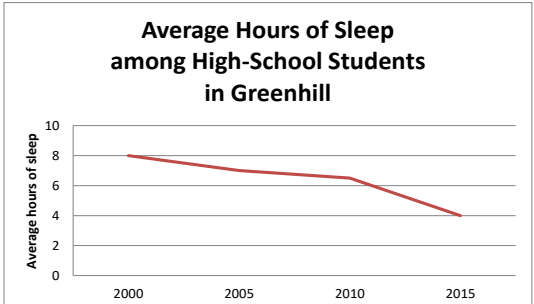
(http://www.eiken.or.jp/teap/construct/rating_crit.html)

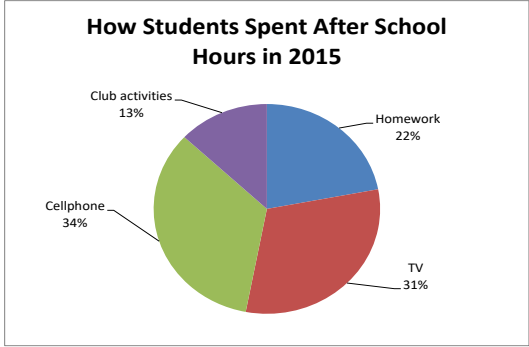
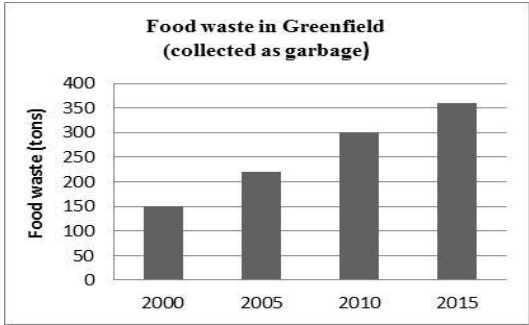
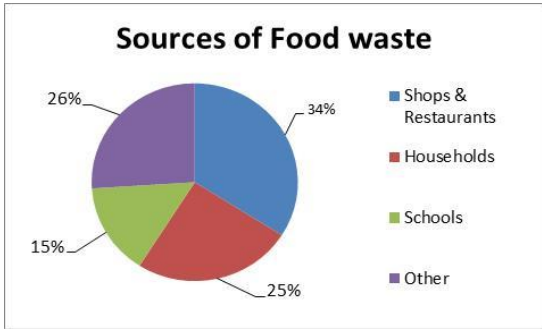
TEAP WRITING, TASK B

CEFR	Level Descriptors	
	Main Ideas	Synthesizes and evaluates information and arguments from all of the verbal and nonverbal input texts.
	Coherence	Organized as a coherent response to the task; organization of ideas within and across paragraphs is generally clear, though may be formulaic.
	Cohesion	Uses discourse markers and referential cohesive devices effectively to mark the relationship between sentences and link utterances into clear, coherent discourse.
B2	Lexical Range & Accuracy	Uses appropriate synonyms and alternative expressions to convey the main ideas.
	Grammatical Range & Accuracy	Uses a range of sentence structures appropriately; grammatical errors rarely occur and do not impede understanding of the message.
	Main Ideas	Provides a basic summary of some of the main points, bringing together information from more than one of the input texts.
	Coherence	Sentences and paragraphs are generally connected using discourse markers; use of referential cohesive devices (for example, pronominal reference) is mostly clear.
	Cohesion	Sentences are generally connected using discourse markers; use of referential cohesive devices is mostly clear.
B1	Lexical Range & Accuracy	Gives a basic description of the main ideas in the input texts, but tends to rely on the vocabulary supplied in the input texts. Some inappropriate vocabulary usage is evident.
	Grammatical Range & Accuracy	Grammatical errors occur frequently but tend to be associated with attempts at complex structures and do not impede communication of the message.
	Main Ideas	The response refers to some of the elements or points mentioned in one or more of the input texts (verbal and/or non-verbal), but does not synthesize these points or make clear how they are related.
	Coherence	No logical paragraph structure or some separation which is not appropriate; text consists of mainly unconnected sentences with no clear direction or progression across sentences.
	Cohesion	Uses conjunctions to link clauses within sentences, but generally does not mark clearly the relationship between sentences. Use of referential cohesive devices (for example, pronominal reference) is generally not clear.

	Lexical Range & Accuracy	Usage of paraphrasing and synonyms is extremely limited, and alternatives are not appropriate for the task. Errors and unnatural/inappropriate usage are common when reusing vocabulary from the input texts
A2	Grammatical Range & Accuracy	Grammatical errors occur systematically and may impede communication of the message
		Unrelated to task/topic
Below		Fewer than 50 words
A2		Copied directly from the input with little or no original language

Appendix D Survey on computer literacy, and graph familiarity

Computer Literacy											
1. Do you have access to a desktop or laptop computer at home? (Q1_1)	<p>Please check Yes or No.</p> <p><input type="checkbox"/>Yes <input type="checkbox"/>No</p>										
2. How often do you use computers at home? (Q1_2)	<p>Please circle one that applies.</p> <table border="1"> <thead> <tr> <th>Never</th> <th>1-2 times /semester</th> <th>1-2 times /month</th> <th>1-2 times /week</th> <th>3-4 times /week</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> </tr> </tbody> </table>	Never	1-2 times /semester	1-2 times /month	1-2 times /week	3-4 times /week	1	2	3	4	5
Never	1-2 times /semester	1-2 times /month	1-2 times /week	3-4 times /week							
1	2	3	4	5							
3. Do you have access to a desktop or laptop computer at school? (Q1_3)	<p>Please check Yes or No.</p> <p><input type="checkbox"/>Yes <input type="checkbox"/>No</p>										
4. How often do you use computers at school? (Q1-4)	<table border="1"> <thead> <tr> <th>Never</th> <th>1-2 times /semester</th> <th>1-2 times /month</th> <th>1-2 times /week</th> <th>3-4 times /week</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> </tr> </tbody> </table>	Never	1-2 times /semester	1-2 times /month	1-2 times /week	3-4 times /week	1	2	3	4	5
Never	1-2 times /semester	1-2 times /month	1-2 times /week	3-4 times /week							
1	2	3	4	5							
5. When did you learn to use a computer (with keyboard)? (Q1_5)	<p>When I was () years old.</p>										
6. When you use the keyboard, do you type Japanese characters or English characters? (Q1_6)	<p>Please circle one that applies.</p> <p><input type="checkbox"/> Japanese-Kana <input type="checkbox"/> English-Alphabet</p>										
7. How easy or difficult is it for you to use the keyboard to type English sentences? (Q1_7)	<table border="1"> <thead> <tr> <th>Very Difficult</th> <th>Difficult</th> <th>Neutral</th> <th>Easy</th> <th>Very Easy</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> </tr> </tbody> </table>	Very Difficult	Difficult	Neutral	Easy	Very Easy	1	2	3	4	5
Very Difficult	Difficult	Neutral	Easy	Very Easy							
1	2	3	4	5							
Graph Familiarity											
8. From the following four graphs below please rate each of the graphs in order of difficulty on a scale of 5, with # 1 being the most difficult to understand and 5 being the easiest.	<p>Please circle one that applies.</p>										
(Q2_1)	<table border="1"> <thead> <tr> <th>Very Difficult</th> <th>Difficult</th> <th>Neutral</th> <th>Easy</th> <th>Very Easy</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>2</td> <td>3</td> <td>4</td> <td>5</td> </tr> </tbody> </table>	Very Difficult	Difficult	Neutral	Easy	Very Easy	1	2	3	4	5
Very Difficult	Difficult	Neutral	Easy	Very Easy							
1	2	3	4	5							
 <p>Average Hours of Sleep among High-School Students in Greenhill</p> <table border="1"> <thead> <tr> <th>Year</th> <th>Average hours of sleep</th> </tr> </thead> <tbody> <tr> <td>2000</td> <td>8</td> </tr> <tr> <td>2005</td> <td>7</td> </tr> <tr> <td>2010</td> <td>6</td> </tr> <tr> <td>2015</td> <td>4</td> </tr> </tbody> </table>	Year	Average hours of sleep	2000	8	2005	7	2010	6	2015	4	
Year	Average hours of sleep										
2000	8										
2005	7										
2010	6										
2015	4										

Graph Familiarity					
Continued					
Please circle one that applies.					
(Q2_2)	Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
					
(Q2_3)	Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
					
(Q2_4)	Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
					
9. How would you rate your ability to interpret the graph information in English (Q2_5)	Very Weak 1	Weak 2	Average 3	Strong 4	Very Strong 5

Test-Taking Strategies							
[Task Representation]							
10. Did you understand the instructions on how to write your essay? (Q3_1)			Please check Yes or No. <input type="checkbox"/> Yes <input type="checkbox"/> No				
Please circle one that applies.							
11. How easy or difficult did you find it to fulfil the task requirement? (Q3_2)			Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
12. Which part of the task requirements did you find most challenging? (Q3_3)			Please check one that applies. <input type="checkbox"/> Describing the situation (1) <input type="checkbox"/> Summarizing the main points (2) <input type="checkbox"/> Stating which solution might work best (3) <input type="checkbox"/> Giving the reasons for the choice you made (4) <input type="checkbox"/> No difficulty at all (5)				
[Macro-Planning]							
13. Did you identify the purpose of the essay? (Q4_1)			Please check Yes or No. <input type="checkbox"/> Yes <input type="checkbox"/> No				
14. Did you think about which solutions would work the best? (Q4_2)			Please check Yes or No. <input type="checkbox"/> Yes <input type="checkbox"/> No				
[Micro-planning: Graph information]							
15. How easy or difficult was it for you to read the titles of the graphs? (Q5_1)			Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
16. How easy or difficult was it for you to read the values on the graphs? (Q5_2)			Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
17. How easy or difficult was it for you to read the units on the x- and y-axis on the graphs? (Q5_3)			Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
18. How easy or difficult was it for you to interpret the information of the graphs? (Q5_4)			Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5
19. How easy or difficult was it for you to summarize the main trends of the graphs in English? (Q5_5)			Very Difficult 1	Difficult 2	Neutral 3	Easy 4	Very Easy 5

<p>[Micro-planning: Source texts]</p> <p>20. Did you re-read the parts where you thought it was important to include in your essay? (Q6_1)</p>	<p>Please check one that applies.</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> I don't remember</p>
<p>21. How did you decide which information to include from the texts? (Q6_2)</p>	<p><input type="checkbox"/> Never thought about it (1)</p> <p><input type="checkbox"/> By re-reading the whole text once or twice (2)</p> <p><input type="checkbox"/> By going back to a specific paragraph of some importance (3)</p> <p><input type="checkbox"/> By scanning and looking for the keywords (4)</p> <p><input type="checkbox"/> By memory (5)</p>

Test-Taking Strategies					
Continued					
<p>22. How did you connect the ideas from the texts? (Q6_3)</p>	<p><input type="checkbox"/> Never thought about it (1)</p> <p><input type="checkbox"/> By copying the sentences from the source texts (2)</p> <p><input type="checkbox"/> By borrowing some words and phrases from the source texts (3)</p> <p><input type="checkbox"/> By referring to ideas from the source texts (4)</p> <p><input type="checkbox"/> By memory (5)</p>				
[Micro-planning: Response format]					
<p>23. Did you make an outline BEFORE writing your essay? (Q7_1)</p>	<p>Please check one that applies.</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> I don't remember</p>				
<p>24. Did you decide how many paragraphs there should be in your essay? (Q7_2)</p>	<p>Please check one that applies.</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> I don't remember</p>				
[Fulfilling the requirements]					
<p>25. How well do you think you have described the situation? (Q8_1)</p>	Very Poor 1	Poor 2	Average 3	Good 4	Very Good 5
<p>26. How well do you think you have summarized the main points? (Q8_2)</p>	Very Poor 1	Poor 2	Average 3	Good 4	Very Good 5

27. How well do you think you have stated which solution would work best? (Q8_3)	Very Poor 1	Poor 2	Average 3	Good 4	Very Good 5
28. How well do you think you have given the reasons for the choice made? (Q8_4)	Very Poor 1	Poor 2	Average 3	Good 4	Very Good 5
[Translating-use of language]	Please check one that applies.				
29. Did you think what verb tense form to be used before writing? (Q8_5)	<input type="checkbox"/> Yes <input type="checkbox"/> No				
30. When did you decide which verb to use? (Q8_6)	<input type="checkbox"/> Don't remember (1) <input type="checkbox"/> After I finished writing the essay (2) <input type="checkbox"/> When I was writing the essay (3) <input type="checkbox"/> Before I started writing the essay (4) <input type="checkbox"/> While I was reading the source texts (5)				
31. Which tense did you mostly use? (Q8_7)	<input type="checkbox"/> Don't remember (1) <input type="checkbox"/> Future tense (2) <input type="checkbox"/> Present tense (3) <input type="checkbox"/> Past tense (4) <input type="checkbox"/> Present Perfect tense (5)				
32. Did you think what pronoun form should be used for writing? (Q8_8)	Please check one that applies. <input type="checkbox"/> Yes <input type="checkbox"/> No				
33. Which pronoun forms did you mostly use? (Q8_9)	<input type="checkbox"/> Don't remember (1) <input type="checkbox"/> Proper nouns (2) <input type="checkbox"/> First person pronouns (I / We) (3) <input type="checkbox"/> Second person pronouns (You) (4) <input type="checkbox"/> Third person pronouns (It/She/He/They) (5)				
[Monitoring & Revising]					
Please scale how much you agree or disagree with the statement below					
34. I checked if my sentences were grammatically correct. (Q9_1)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5
35. I checked if my spelling was correct. (Q9_2)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5

36. I checked if I had connected the ideas from the graphs (Q9_3)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5
37. I checked if I put my ideas in a logical order. (Q9_4)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5
38. I checked if I had fulfilled the task requirements by going back to the instruction. (Q9_5)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5
39. I tried my best to avoid repeating the same word or expressions in the essay. (Q9_6)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5
40. I checked if my essay was an appropriate length. (Q9_7)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5
41. I used some sentences and phrases prepared in advance to be used in the essay. (Q9_8)	Strongly Disagree 1	Disagree 2	Neither Agree or Disagree 3	Agree 4	Strongly Agree 5

Appendix E Questions for Focus Group Discussions

The following questions will be discussed in groups.

- a. Did you clearly understand the instruction of the test?
- b. Which part of the tasks did you have the most difficulty with?
 - Reading the texts
 - Understanding the graphs
 - Describing the situation by interpreting the graph information
 - Summarizing the main points about solutions that have been suggested
 - Choosing the best solution
 - Giving the reasons behind your argument
- c. To what extent do you think your typing skills of the keyboard affected your writing process?
- d. How did you decide which information to include?
- e. How did you plan when you start to write your essay?
- f. What did you often do while you are writing?
- g. What did you do after you finished writing your essay?
- h. What kind of strategies did you use while taking the test?
- i. What were you mostly concerned about your essay?
- j. What do you think you would do differently if you had to take the same test again?

Appendix F Feedback Sheet for Focus Group Discussions

FEEDBACK SHEET

お名前： _____ 学校名： _____

1) Did you clearly understand the instruction of the test?

TEAPのライティング・テストは、いわゆる自由記述のエッセイ方式ではなく、与えられた資料を基に論証文を書くエッセイです。まず、何を書かなければならないのか、しっかり指示文を正しく理解する必要があります。これは事前にサンプルテストをしっかり読んで対応することができると思います。

2) Which part of the tasks did you have the most difficulty with?

- ① Reading the texts
- ② Understanding the graphs
- ③ Describing the situation by interpreting the graph information
- ④ Summarizing the main points about solutions that have been suggested
- ⑤ Choosing the best solution
- ⑥ Giving the reasons behind your argument

エッセイを書く工程でどのタスク（課題）が一番難しいとかがえられたでしょうか？

- ① グラフに掲示してある内容は大意を理解することができたでしょうか？
- ② グラフの情報はうまく説明することができましたか？
- ③ 記事と投稿（書簡）の長文がよく読み取れましたか？
- ④ 主な解決策について、要約することがうまくできたでしょうか。
- ⑤ どの解決策が一番良いか書けたでしょうか。
- ⑥ 何故その解決策を選んだのかについて根拠を示すことができましたか？

3) To what extent do you think your typing skills of the keyboard affected your writing process?

パソコンのキーボードのタイピングの速さなどがエッセイを書くうえでどのくらい影響したとお考えでしょうか？

内容についてもう少し詳しく見ていきましょう。

4) How did you decide which information to include?

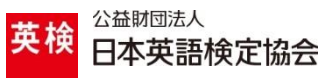
問題中、どの部分の情報をエッセイに含むかについて、どのように決めましたか？

5) How did you plan when you start to write your essay?

エッセイを書き始めるにあたり、どのように計画して書き始めましたか？

- 6) What did you often do while you are writing?
エッセイを書いている最中に、どんな事をよくやりましたか？
- 7) What did you do after you finished writing your essay?
エッセイを書き終えた後、何をしましたか？
- 8) What kind of strategies did you use while taking the test?
テスト戦略的にどんな手法を使いましたか？
- 9) What were you mostly concerned about your essay?
エッセイ作文で一番気がかりになったこととは？
- 10) What do you think you would do differently if you had to take the same test again?
次回もし同じテストを受けなければならなかったとして、どのような異なった方法で取り組みますか？

Appendix G Consent Form



Graduate School of Education

Consent Form

Project Title: An Investigation of Test-takers' Cognitive Processes during Integrated Writing Tasks Which Use Multiple Texts and Graphs as Prompts

1. I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions.
2. I agree to the arrangement described in the Information Sheet in so far as they relate to my participation.
3. I understand that my participation is voluntary and that I am free to withdraw from the project at any point during data collection without giving reason. If I withdrew after the data collection, the data which I had provided would be destroyed immediately in a secure manner.
4. I also understand that all personal information provided will remain confidential and no information will identify me will be made publicly available.
5. I have received a copy of this Consent Form and of the accompanying Information Sheet.
6. I agree to take part in the above study.

Name of the Participant:

Signed:

Date:

Name of the Guardian:

Signed:

Date:

Name of the Researcher: Mikako Nishikawa

Signed:

Date:

Appendix H Information Sheet



公益財団法人
日本英語検定協会



Researcher:

Mikako Nishikawa

E-mail: mikako.nishikawa@bristol.ac.uk

Phone: +81-(0)80-5328-0117

Supervisor

Dr. Guoxing Yu

Phone: +44-117-3314436

E-mail: Guoxing.Yu@bristol.ac.uk

Graduate School of Education

35 Berkeley Square

University of Bristol

Bristol, BS8 1JA, United Kingdom

Information Sheet

Project title: An Investigation of Test-takers' Cognitive Processes during Integrated Writing Tasks Which Use Multiple Texts and Graphs as Prompts

Purpose and Benefits:

Traditionally, English language proficiency has been measured with four separate skills: reading, listening, writing and speaking. Nowadays, integrated writing tasks for assessing the academic writing skills of second language (L2) learners have become more common. While there has been a rise in the number of publications on integrated writing, little has been studied regarding the use of graphic information in integrated writing. This study aims to identify some of the cognitive patterns which are involved in integrated writing when test-takers use information from multiple source texts and graphs as prompts using the Test of English for Academic Purposes (TEAP).

Procedures:

- (1) You will take a language proficiency test, called the APTIS test for reading and writing modules at school.

The official scores of the test will be reported to you a few weeks after the test date.

- (2) You will take two patterns of the TEAP Sample Test (similar to Task B of the TEAP writing test) consecutively on the same day in March. While you take the tests, eye-movements will be recorded one test-taker at a time. This will take 100 minutes all together. The appointment will be arranged by your English teacher at school.
- (3) Immediately upon finishing the tests, you will be asked to answer the survey.
- (4) A follow-up focus group interview will be arranged a few weeks after the TEAP sample tests in order to discuss your experience of taking the test. You will receive written feedback within 3 months after the experiment.

Anonymity and Confidentiality:

All personal information will remain confidential, and no information you provided for the research that could identify you will be publicly available. To ensure your anonymity, the researcher will use pseudonyms for data collection, including the test results, survey responses and responses in the focus group discussion.

Data Storage and Data Protection:

The data will be stored securely in the researcher's computer drive in the protected university server (University of Bristol) until at least five years after the research has been completed and then destroyed in a secure manner when it is no longer needed. Furthermore, the data will be used for research purpose only

Participants' Rights:

You are entitled to withdraw at any point during the data collection sessions and within one week after the data collection session; once you withdraw, the provided data will be destroyed in a secure manner. The test results of the study will be provided to you.

Further Information:

If you have further questions about the study, please feel free to contact the researcher or the researcher's supervisor; contact details are provided above.

Name of the Researcher: Mikako Nishikawa

Signed:

Date:

Appendix I Information Sheet (Japanese)



公益財団法人

日本英語検定協会

同意書



研究課題名：『統合型ライティングにおける認知プロセスの解明』

1. はじめに

従来、英語運用能力は技能別（リーディング、リスニング、スピーキング、ライティング）のテストで測られてきました。近年、よりアカデミックな場面で必要とされるライティング・スキルを測定する目的で、ライティングテストに統合型の問題（Integrated Writing Test）が問われるようになって参りました。（例：TOEFL iBT, IELTS）

日本英語検定協会が実施する TEAP (Test of English for Academic Purposes) も統合型のテストの代表であり、大学入試の外部試験活用などで 2014 年度より導入されております。この TEAP Task B においては、複数のパッセージと図表の情報から論証文（200 語）を書く問題が出題されていますが、この統合型ライティングに関する認知プロセスの解明に出来るだけ多くの現役高校生の皆様にご協力いただければと考えております。

2. 目的、方法について

- ① 統合ライティングでは、個々の技能（リーディングとリスニング）で、どの程度の能力を測ることが出来ているのか明確にするために、ブリティッシュ・カウンシルが運営するアプティス（APTIS）というテスト（所要時間 2 時間）を受験いただきます。12 月—1 月の間に実施を予定しています。スコアは約 2 週間でお手元に届きます。
- ② 平成 28 年度 3 月に TEAP Task B の模擬テスト 2 種（40 分 x 2 回）をパソコン上で一人ずつ受験いただきます。これはスコアを計測するものではなく、被験者が画面のどの部分を読み取っているかを眼球測定装置（PC 上のカメラ）を使い測定するものです。キーボード入力された文字情報とあわせて、被験者の認知プロセスの動向をデータ解析いたします。実験の実施時間・場所については、●●学校（英語科）の●●教諭にご相談のうえ、改めて通知をさせていただきます。

3. 予想される効果、効用について

英語 4 技能の中で日本人学習者が最も苦手とし、上達に学習時間がかかるのはライティングだと言われています。大学受験の英語 4 技能の要件化に伴い、アカデミック・ライティングはもはや避けては通れない英語スキルの一つとなりつつあります。本研究にご賛同いただきました生徒さまへは、データ解析に基づき傾向をフィードバックするほか、後日「TEAP アカデミック・ライティング講座」への受講を無償提供させていただきます。予定にしております。

4. 同意について

研究への参加は自由意志であり、研究に同意した後でも、自由意志によって同意を撤回でき、またそれらのことによって不利益を受けることはありません。

5. コストについて

テスト費用は無料です。

ただし、万が一なんらかの事情にて実施会場が学校以外の場所で行われた場合の交通費は実費となります。

6. 個人情報の保護

本研究で得られたデータを学会や論文で発表することはありますが、個人の名前を特定されるようなことは一切ありません。

7. 本研究に関する質問に対するお問い合わせ先

公益財団法人・日本英語検定協会 教育事業部 研究員 西川美香子

大阪市北区曽根崎新地 1-3-16 京富ビル 4階 (06) 7670-8085 [Mail] m-nishikawa@eiken.or.jp

Ph.D Candidate, Graduate School of Education (Applied Linguistics) University of Bristol,
mikako.nishikawa@bristol.ac.uk

Appendix J GSoE Research Ethic Form

It is important for members of the Graduate School of Education, as a community of researchers, to consider the ethical issues that arise, or may arise, in any research they propose to conduct. Increasingly, we are also accountable to external bodies to demonstrate that research proposals have had a degree of scrutiny. *This form must therefore be completed for each piece of research carried out by members of the School, both staff and students*

The GSoE's process is designed to be supportive and educative. If you are preparing to submit a research proposal, you need to do the following:

1. **Arrange a meeting with a fellow researcher**

The purpose of the meeting is to discuss ethical aspects of your proposed research, so you need to meet with someone with relevant research experience. A list of prompts for your discussion is given below. Not all these headings will be relevant for any particular proposal.

2. **Complete the form on the back of this sheet**

The form is designed to act as a record of your discussion and any decisions you make.

3. **Upload a copy of this form and any other documents (e.g. information sheets, consent forms) to the online ethics tool at: <https://dbms.ilt.bris.ac.uk/red/ethics-online-tool/applications>.**

Please note: Following the upload you will need to answer ALL the questions on the ethics online survey and submit for approval by your supervisor (see the flowchart and user guides on the GSoE Ethics Homepage).

If you have any questions or queries, please contact the ethics co-ordinators at: gsoe-ethics@bristol.ac.uk

Name(s): Mikako Nishikawa

Proposed research project: An Investigation of Test-takers' Cognitive Processes during Integrated Writing Tasks Which Use Multiple Texts and Graphs as Prompts

Proposed funder(s): Eiken Foundation of Japan

Discussant for the ethics meeting: Mrs. Faizulizami Osmin, a second year Ph.D. student at Graduate School of Education, University of Bristol

Name of supervisor: Dr. Guoxing Yu

Has your supervisor seen this submitted draft of your ethics application? Yes

Abstract

Integrated writing tasks for assessing the academic writing skills of second language (L2) learners have become more common than traditional writing-only tests in recent years (e.g. Plakans, 2008; Gerbril, 2010.) While there have seen a rise in the number of publications on integrated writing, little has been studied for the use of graphic information in integrated writing (e.g., Yu, Rea-Dickins & Kiely, 2012; Yang, 2012.) This study aims to identify some of the cognitive patterns which are involved in integrated writing when test-takers use information from multiple source texts and graphs as prompts using the Test for English as Academic Purposes (TEAP). To achieve this aim, it will analyze how test-takers (n=40) incorporate information from the multiple texts and the graphs by testing two different task features within subject. The study will employ a mixed methods approach which consists of four phases: (1) describing participants' language proficiency in reading and writing, (2) understanding the cognitive processes of integrated writing by measuring eye movements, (3) triangulating the findings through the use of surveys, and finally (4) follow-up group discussions on the test-takers experience of the test. In the first phase, the quantitative data will be collected in order to understand the language proficiency of the participants, using the APTIS test. In the second phase, data from eye-movement recording (Bax, 2013) will be analyzed in order to identify some of the key variables that affect reading-into-writing processes and to compare two types of tasks with different graphic information by referring to Weir's cognitive process of writing (Weir, 2014), which was modeled after Field (2004.) Lastly, both the quantitative and qualitative data collected from the survey and the group discussions will be used to explore participants' test-taking strategies (e.g. Cohen, 2006; Plakans, 2009; Yang 2012; Xu and Wu, 2012) in an attempt to explain the association between eye-movement and the effect of texts and graphs among the participants (Yu & Lin, 2014.) The findings of this study will also help identify similarities and differences in the cognitive process of integrated writing between skilled and unskilled L2 learners.

Ethical issues discussed and decisions taken:

Graduate School of Education (GSoE) ensures all research students to follow ethical guidelines of British Educational Research Association (BERA.)

We discussed ethical issues that should be taken into consideration using the prompts in the guideline. Here is the summary of what we have discussed on the issues and actions to be taken.

1. Researcher access/ exit

It is important to think through how the research project affects the participants. In this research, high-school teachers act as a gateway for the recommendation of students who will participate in the project. Although participants will join the study on a volunteer basis, they might be concerned how their grades or report cards might be affected from the participation in the study.

After the discussion, I decided to clearly indicate that the study will not affect teachers' appraisals on students' grade reports as written on the information sheet. The high-school teachers will confirm this by telling his/her students. After the completion of all data collection and initial analysis, I plan to visit the high school to thank the students and teachers that were involved in my study.

2. Information given to participants

It is also important to consider who will sign the contract and be informed of this project. In the proposal, the initial plan was to only have the participants and their parents sign the consent form, informing the research purpose, procedure, participants' rights, feedback and contact address. Through our discussion, we thought it would be safe to have a teacher sign on a separate consent form, as a gatekeeper must be aware of the research intent.

My decision after the discussion was to prepare the consent in English and Japanese which will be distributed to participants via high school teachers. This will help to answer any questions that the participants might have when I meet them face-to-face, thus I will be able to thoroughly explain what are required of the participants.

3. Participants right of withdrawal

Participants will be fully informed about their rights to withdraw from the study at any time of the data collection. However, it might be uncomfortable for a participant to tell his/her teacher to discontinue participation in the study. This is a sensible issue to be dealt with from the point of the participants' views.

Therefore, I will provide my contact e-mail address so that the participant can inform me through an e-mail straight away without worrying about further inquiry. I will also clearly indicate that the participants do not have to explain the reasons to withdraw from their participation of the study and that their grades or report cards will not be affected. If they wish, they can also contact directly to his/her teacher to stop participating in the study as well.

4. Informed consent

By agreeing to participate in the research project, the participants will be fully informed what they are required to do. The area of my concern was the cost of study. The cost to take the test will be given free of charge, but the transportation expenses (local bus or train) will not be covered if high-school teachers decide to arrange the venue for data collection outside of the school. The possibility of this is clearly stated on a consent form and will be fully informed to all the participants.

After the discussion, I decided to inform the participants this in advance so they can foresee whether they can afford the extra time participating in the study outside of the school. We also talked about setting up a time frame until when they can request to withdraw from the study after the data collection (e.g. they can inform the researcher to withdraw until one week after the data collection).

5. Complaints procedure

Participants need to know who to contact when they have any concerns about the conduct of the research. Although the researcher's name and contact are already stated, the participants may feel they need to contact someone other than the researcher.

Therefore, I decided to include not only my contact address as a first contact person but also my supervisor's contact information in case the participants wish to make a direct comment or complaint about the study.

6. Safety and well-being of participants/ researchers

In our discussion, an issue was raised "what is the safety procedure if the participants have to go to a different destination on weekends to participate in the data collection?" Since the high-school teachers will choose the venue of the data collection, the students should be familiar with the place in advance. Even so, it was felt that there are needs of a safety regulation given to the students beforehand.

To solve this issue, I decided to include a mobile phone number for emergency contact if the experiments take place on weekends. Should emergency happen, she/he will contact the student's high-school teacher and follow the safety regulation of that school.

7. Anonymity/ confidentiality

Data collected from the participants will not identify the participants and will be kept confidential. It is clearly stated in the information sheet as well as in the consent form. In this study, I also plan to use pseudonyms names by coding each student with numbers and

alphabets. I will clearly explain to the participants about their rights to anonymity and confidentiality when the researcher meets face to face.

8. Data collection

This study collects both quantitative data (i.e. language tests, eye-movement recordings on computer screen, keystroke inputs on a keyboard, surveys) as well as qualitative data (i.e. stimulated-recall, focus group interviews.) The participants might be concerned about eye-tracking for it is their first time to participate in this type of study.

It was recommended that I inform the participants about the data collection procedures in details. They will be monitored during the test in case they find the process either too uncomfortable or raise any questions about technical difficulty.

9. Data analysis

As mentioned above, this study uses a mixed methods approach. Some of the quantitative results will be triangulated by qualitative data in this study. I will ask the participants to look at the transcription of the interviews for accuracy of the data recorded.

10. Data storage

I will follow the UK Data Protection Act to ensure personal information safely stored in the University network. I will use a laptop with password protection.

11. Data Protection Act

By following the UK Data Protection Act, the consent form clearly states that the data collected will be used for the purpose of research only and that they will be destroyed in a secure manner upon the completion of the research.

12. Feedback

In this study, participants will receive some feedback on the test results in score report. Other oral feedbacks are also available through focussed group discussions. The participants might be worried to whom these results are reported to (e.g. high-school teachers.)

I will inform the participants fully in advance that the score reports will be sent in a sealed envelope via high-school teachers. The teachers will not have an access to individual scores although they will receive a summary report.

13. Responsibilities to colleagues/ academic community

This will be done in the form of a summary report. Also, some of the findings will be publically available through published journals.

14. Reporting of research

It is important to credit their contribution; therefore, I will acknowledge this in the doctoral thesis of this study.

By typing your name here, this is equivalent to a signature.

Signed: Mikako Nishikawa (Researcher)

Signed: Faizulizami Osmin (Discussant)

Date: November 18, 2015

Appendix K Focus Groups Discussions
FEEDBACK 回答一覧（平成 28 年 3 月 9 日実施分）

Q1_Did you clearly understand the instruction of the test?

－TEAP を受けたことがあったので理解しやすかったです

Q2_Which part of the tasks did you have the most difficulty with?

6) Giving the reasons behind your argument (n=4/6)

Q3_To what extent do you think your typing skills of the keyboard affected your writing processes?

－時間が足りなかった

－時間が足りなかった

－つづりを思い出すのにつまったこともあった。

－下を見ないと打てない時もあるって時間が足りないことがあった

－タイピングに慣れていなくて、時間が足りなかった

Q4_How did you decide which information to include?

－Keyword を探した

－重要そうな単語を入れた

－何度も出ているキーワードを見つけたりしました

－指示文に自分の考えを一つ選んでとあったので、どの意見に賛同するかをきめて自分のエッセイに引用した

－文中に出てくる人の言葉を借りました

－何回も書かれているキーワード

Q5_How did you plan when you start to write your essay?

－Mike Parker の意見から書き始めた

－特に計画を立てなかったのですが、グラフの情報や意見を書き始めようと思いました

－自分の意見を先にきめる

－グラフの情報をまとめて頭を整理しながら取り掛かった

－事実→意見の順に書いた

－Key Point を探して、大きなポイントを3つくらいにしました

Q6_What did you often do while you are writing?

－手で（振り回しながら）考えた

- ー見直し、内容に関して自分にも当てはまることがないか考えた
- ースペルが間違っていないかチェックしながらやりました（時間も気にしながら）
- ー文字数のチェック
- ー単語のスペルの見直し

Q7_ What did you do after you finished writing your essay?

- ー文全体の見直し
- ーキリがいいところで書くのを止めてその後見直しました
- ー自制やスペルのチェック、見直し、少し伸びをした
- ー三人称のsとかを見直しました。できるだけ分かりやすい文章にするために手直しました
- ーつけ足し、スペルミスのチェック
- ーぎりぎりまで考えて、直したりした

Q8_ What kind of strategies did you use while taking the test?

- ー同じ言葉を避けようとした。時間配分
- ー単語数、増やそうとした。ややこしい文章にせず、単純な文にした
- ー同じ言い回しは何度も使わないようにしました。Taking these things → Among these three solutions に変えるなど
- ー文字数を稼ぐため、略を使うのをやめた (I'd = I would)
- ーパラグラフライティングの構成を意識
- ー同じ言葉をできるだけひかえました
- ー要約&自分の意見をとりあえず書く。一つのエッセイを書き上げる

Q9_ What were you mostly concerned about your essay?

- ーちゃんと文を正しく理解できているか、文法が間違っていないか
- ー文法間違いや時間配分がきになりました
- ー言いたい文が英語にできず、別の言い回しにしたが自信が持てない
- ー時制（現在形が急に過去形になってしまうことがあったので
- ーどのように書けばいいかあまりわからなかった
- ー上手い言い回しが思いつかない！！

Q10_ What do you think you would do differently if you had to take the same test again?

- ーもっと分かりやすい文章の構成を考えてから書き上げる
- ー書き始める前にだいたいのアウトラインを決めようと思います

- パラグラフライティングの構成の見直しをして、単語や使える言い回しを増やして挑む
- 最初にアウトラインを作ってからエッセイに取り掛かりたいです
- 聞かれているタスクをライティング欄に書き出して出題された答えをきちんと答えられるようにする
- もっと色んな単語を使ってライティングしたい

FEEDBACK 回答一覧（平成 28 年 4 月 12 日実施分）

Q1_Did you clearly understand the instruction of the test?

- 2 つ目がわからなかった
- わかった！と思っていましたが、分かっていなかった！と素直にわかりました
- Yes
- 分かっていたが、どのように要約するのかがわからず、悩んだ
- No, I didn't

Q2_Which part of the tasks did you have the most difficulty with?

- 4) Summarizing the main points about solutions that have been suggested (n=10)
- 6) Giving the reasons behind your argument (n=6)

Q3_To what extent do you think your typing skills of the keyboard affected your writing processes?

- あまり影響していないと思う
- Little
- そんなに影響しなかったと思います
- 少し
- 遅かったがエッセイを書くのに特に影響されていない
- タイプの方が少し遅かったと思う
- 普段、結構使いたおしているの、キーボードによる影響はほとんどなかったと思います
- かなりキーボードは苦手なのでどうだったかなあ
- Maybe a little
- 遅かったけど、書くのに影響されてないです
- かなり影響した
- 関係ないと思う
- あまり影響していないと思う
- とても時間がかかってしまった

Q4_How did you decide which information to include?

- 問題文で問われた順に書く
- 登場人物が言っていたこと、グラフ
- 今回は、あまり何も考えていませんでした・・・
- 読みながら考えた、でも深く考えなかった
- 読みながら考えた
- エッセイを読みながら
- 詳しくは覚えていませんが、問題文中に書かれている中で主題になりそうなところを探して決めた。
- とりあえず書き始めてしまった。
- 読んでから書き始めた
- 深く考えられなかった
- 状況説明
- 段落の最後のへん
- 一般論としてかかっている部分
- グラフ
- 文中に登場した人物が述べている意見をとりいれるようにした
- 各段落の最後の部分

Q5_How did you plan when you start to write your essay?

- 状況把握、要約、自分の意見
- 重要場部分をかけるようにそこを探した
- あまり計画しなかった
- 特に何もなし、問題文が求めた順に書いた
- 特に気にせず書き始めた
- とりあえず書き始めてしまった
- あらすじを考える、キーワードを拾う
- 状況を説明して流を考えていった
- 特に計画していない
- アウトラインを作る（ようにしたい）
- 重要な部分を探す
- 文の前後が矛盾しないようにした
- あまり考えていなかった

Q6_ What did you often do while you are writing?

- キーワードに立ち返る
- 本文のコピー
- キーワードがあった箇所に戻って見直した
- 間違っただけを書き直した
- コピー
- まるうつし
- 問題文を少し見た
- 立ち止まってしまった。頭でまとまっていなかった
- 同じ単語を極力さける。文章を見返す
- 文章をそのまま使った
- 本文そのままコピー
- そのままコピーではなく、近い意味の単語を書いた
- キーワードや解決策が書かれているところに戻って読み返す（ようにしたい）
- 解決策を探す
- 大事なところを読み返す 本文を書き写した
- 資料で重要そうなところを抜き出した。

Q7_ What did you do after you finished writing your essay?

- 見直す
- 読み返すようにする（したい）
- 時間がなくて書き終わらなかった。見直しができるようになりたい！！
- 少し見直しをした、でも時間が足りなかった
- 見直し
- 見直ししなかった
- 全て読み返した。語数を数える。
- 2分ほど見はしました。
- 見直し
- 見直し
- 見直した
- 見直し
- 時間ぎりぎりだった
- 見直しが大切（見直すようにしたい）
- 見直す
- 時間がなかった

—見直し

Q8_ What kind of strategies did you use while taking the test?

—同じ内容の言い換え

—時間配分をする（したい）

—なるべく自分の古語場で書こうと思ったけど、あまり何も考えずに適当にやってしまいました。

—時間配分に気を付けようとおもった、でもうまくいかなかった

—特にないです

—特になし

—気合いです。頭と手を必死に動かし続けた

—1つのパラグラフ内での分を短くする

—時間配分

—最後に見直した（時間配分）

—特に使っていない

—時間配分、パラグラフィティングに注意すべき

—特になし

—大事そうなところを丸写しした

—考えていなかった

Q9_ What were you mostly concerned about your essay?

—グラフの情報をどこでいれるか

—時制の一致

—自分の書いた文章がまとまっていたのかどうか気になった

—時間が足りず、文字数が少なかった

—コピーペースト

—時間

—特になし

—My skill to write an essay。エッセイが何か、終始それについて考えていました

—時間、文字数

—話がずれていた

—時制と段落のつくりかた

—パラグラフ

—時間が足りなかった

—時制、何をどの順で書くか

—うまく要約ができなかったこと。字数が稼げなかったこと

—文法、根拠があるかどうか

—問題文の指示に沿っているかどうか

Q10_What do you think you would do differently if you had to take the same test again?

—全体に目を通して、構成を下書きする

—同じ語をなるべく使わないようにして、Introduction, Body, Conclusion を考える

—まず、3段階構成を頭の中で意識してから、本文の言葉をなるべく変えて書くようにしたい。

—今日、聞いたことを実践してしっかり読んで、見直しをする

—時間配分とパラフレーズ

—パラグラフ分けなどをきっちりしたい。今回学習した6つのことの実践

—指示文からしっかり読む。本文をコピーせず、別に言い換える。Run-on sentence をなくす

—コピーアンドペーストはやらない。大切なのは形なのですね！（今の私にできることは）

—時間を気にしながら、大切な所（解決策、問題点）にもどりたい

—モット計画を立てて書き始める

—段落構成を意識して、特に And- But- といった風に、文章の初めを意識したい（And/But でははじめない！）

—類義語を使う、時制に気を遣う、本文をそのまま使わない、自分の意見を述べる

—まず全体を読んでから、キーワードを見つける

—やらなければならないことを読み取って、さーっと読む、重要なことを並べて要約し、自分の意見をかく

—構成をしっかりととらえて気を付ける

—どこから引用したのか、5W1Hを意識する

—まず全体の資料をしっかり読もうと思う。そして run-on sentence にも気を付ける

Appendix L Wilcoxon signed rank tests of differences in eye-movement measurements in Task 1 and Task 2 (All 40 minutes)

Dependent variable	Z	p-value
Time to First Fixation Essay Mean	-3.698b	.000
Time to First Fixation Graph Title1 Mean	-1.922b	.055
Time to First Fixation Graph Title2 Mean	-1.707c	.088
Time to First Fixation Letter-to-Editor Mean	-2.074b	.038
Time to First Fixation Line Graph Mean	-1.936b	.053
Time to First Fixation Paragraph1 Mean	-.795b	.426
Time to First Fixation Paragraph2 Mean	-1.922b	.055
Time to First Fixation Paragraph3 Mean	-1.326b	.185
Time to First Fixation Paragraph4 Mean	-1.129b	.259
Time to First Fixation Pie Chart Mean	-2.903b	.004
Time to First Fixation Sarah Case Mean	-.991b	.322
Time to First Fixation Task Instruction Mean	-1.667c	.096
First Fixation Duration Essay Mean	-1.045c	.296
First Fixation Duration Graph Title1 Mean	-1.045b	.296
First Fixation Duration Graph Title2 Mean	-1.752b	.080
First Fixation Duration Letter-to-Editor Mean	-.513b	.608
First Fixation Duration Line Graph Mean	-1.864b	.062
First Fixation Duration Newspaper Article Mean	-1.939b	.052
First Fixation Duration Paragraph1 Mean	-.036b	.971
First Fixation Duration Paragraph2 Mean	-.991b	.322
Fixation Duration Paragraph3 Mean	-1.548b	.122
Fixation Duration Paragraph4 Mean	-1.488b	.137
Fixation Duration Pie Chart Mean	-.904b	.366
Fixation Duration Sarah Case Mean	-.365b	.715
Fixation Duration Task Instruction Mean	-2.111c	.035
Total Fixation Duration Essay Mean	-2.168b	.030
Total Fixation Duration Graph Title1 Mean	-2.123b	.034
Total Fixation Duration Graph Title2 Mean	-2.499b	.012
Total Fixation Duration Letter-to-Editor Mean	-1.016b	.310
Total Fixation Duration Line Graph Mean	-2.581b	.010
Total Fixation Duration Newspaper Article Mean	-2.387b	.017

Total Fixation Duration Paragraph1 Mean	-4.837b	.000
Total Fixation Duration Paragraph2 Mean	-3.212b	.001
Total Fixation Duration Paragraph3 Mean	-1.593b	.111
Total Fixation Duration Paragraph4 Mean	-2.606b	.009
Total Fixation Duration Pie Chart Mean	-.222b	.824
Total Fixation Duration Sarah Case Mean	-.108c	.914
Total Fixation Duration Task Instruction Mean	-1.073b	.283
Fixation Count Essay N	-1.000b	.317
Fixation Count Essay Mean	-.293c	.769
Fixation Count Graph Title2 N	-1.732b	.083
Fixation Count Graph Title2 Mean	-.445c	.656
Fixation Count Graph Title N	-1.732b	.083
Fixation Count Graph Title Mean	-2.603c	.009
Fixation Count Letter-to-Editor N	-1.000b	.317
Fixation Count Letter-to-Editor Mean	-2.066c	.039
Fixation Count Bar Graph N	-1.414b	.157
Fixation Count Bar Graph Mean	-1.512c	.130
Fixation Count Newspaper Article N	-1.291b	.197
Fixation Count Newspaper Article Mean	-2.989b	.003
Fixation Count Paragraph1 N	-1.414b	.157
Fixation Count Paragraph1 Mean	-4.617b	.000
Fixation Count Paragraph2 N	-1.414b	.157
Fixation Count Paragraph2 Mean	-.390c	.697
Fixation Count Paragraph3 N	-1.414b	.157
Fixation Count Paragraph3 Mean	-3.196c	.001
Fixation Count Paragraph4 N	-1.732b	.083
Fixation Count Paragraph4 Mean	-4.006b	.000
Fixation Count Pie Chart N	-1.414b	.157
Fixation Count Pie Chart Mean	-.343c	.732
Fixation Count Sarah Case N	-.302b	.763
Fixation Count Sarah Case Mean	-.361b	.718
Fixation Count Task Instruction N	-1.414b	.157
Fixation Count Task Instruction Mean	-2.701c	.007
Visit Duration Essay N	-1.660b	.097
Visit Duration Essay Mean	-1.976c	.048

Visit Duration Essay Sum	-.430c	.667
Visit Duration Graph Title2 N	-.593c	.553
Visit Duration Graph Title2 Mean	-2.534b	.011
Visit Duration Graph Title2 Sum	-.148b	.882
Visit Duration Graph Title N	-2.227c	.026
Visit Duration Graph Title Mean	-.614c	.539
Visit Duration Graph Title Sum	-2.372c	.018
Visit Duration Letter-to-Editor N	-1.124c	.261
Visit Duration Letter-to-Editor Mean	-.401c	.688
Visit Duration Letter-to-Editor Sum	-1.712c	.087
Visit Duration Bar Graph N	-.949c	.343
Visit Duration Bar Graph Mean	-.161b	.872
Visit Duration Bar Graph Sum	-1.768c	.077
Visit Duration _Newspaper Article N	-2.528b	.011
Visit Duration Newspaper Article Mean	-2.815b	.005
Visit Duration Newspaper Article Sum	-3.422b	.001
Visit Duration Paragraph1 N	-2.712b	.007
Visit Duration Paragraph1 Mean	-2.635b	.008
Visit Duration Paragraph1 Sum	-4.852b	.000
Visit Duration Paragraph2 N	-.397c	.692
Visit Duration Paragraph2 Mean	-.524b	.600
Visit Duration Paragraph2 Sum	-.618c	.536
Visit Duration Paragraph3 N	-2.408c	.016
Visit Duration Paragraph3 Mean	-1.566c	.117
Visit Duration Paragraph3 Sum	-3.712c	.000
Visit Duration Paragraph4 N	-3.885b	.000
Visit Duration Paragraph4 Mean	-.516b	.606
Visit Duration Paragraph4 Sum	-3.804b	.000
Visit Duration Pie Chart N	-.350c	.727
Visit Duration Pie Chart Mean	-.477c	.633
Visit Duration Pie Chart Sum	-1.277c	.202
Visit Duration Sarah Case N	-.338c	.735
Visit Duration Sarah Case Mean	-.823b	.410
Visit Duration Sarah Case Sum	-.570b	.568
Visit Duration Task Instruction N	-3.099c	.002

Visit Duration Task Instruction Mean	-1.411b	.158
Visit Duration Task Instruction Sum	-2.997c	.003
Total Visit Duration Essay N	-1.414b	.157
Total Visit Duration Essay Mean	-.430c	.667
Total Visit Duration Graph Title2 N	-1.732b	.083
Total Visit Duration Graph Title2 Mean	-.148b	.882
Total Visit Duration Graph Title N	-1.732b	.083
Total Visit Duration _Graph Title Mean	-2.372c	.018
Total Visit Duration Letter-to-Editor N	-1.000b	.317
Total Visit Duration Letter-to-Editor Mean	-1.712c	.087
Total Visit Duration Bar Graph N	-1.414b	.157
Total Visit Duration Bar Graph Mean	-1.768c	.077
Total Visit Duration Newspaper Article N	-1.291b	.197
Total Visit Duration Newspaper Article Mean	-3.565b	.000
Total Visit Duration Paragraph1 N	-1.414b	.157
Total Visit Duration Paragraph1 Mean	-4.852b	.000
Total Visit Duration Paragraph2 N	-1.414b	.157
Total Visit Duration Paragraph2 Mean	-.618c	.536
Total Visit Duration Paragraph3 N	-1.414b	.157
Total Visit Duration Paragraph3 Mean	-3.712c	.000
Total Visit Duration Paragraph4 N	-1.732b	.083
Total Visit Duration Paragraph4 Mean	-3.804b	.000
Total Visit Duration Pie Chart N	-1.414b	.157
Total Visit Duration Pie Chart Mean	-1.277c	.202
Total Visit Duration Sarah Case N	-.302b	.763
Total Visit Duration Sarah Case Mean	-.570b	.568
Total Visit Duration Task Instruction N	-1.414b	.157
Total Visit Duration Task Instruction Mean	-2.997c	.003
Visit Count Essay N	-1.414b	.157
Visit Count Essay Mean	-1.660b	.097
Visit Count Graph Title2 N	-1.732b	.083
Visit Count Graph Title2 Mean	-.593c	.553
Visit Count Graph Title N	-1.732b	.083
Visit Count Graph Title Mean	-2.227c	.026
Visit Count Letter-to-Editor N	-1.000b	.317

Visit Count Letter-to-Editor Mean	-1.124c	.261
Visit Count Bar Graph N	-1.414b	.157
Visit Count Bar Graph Mean	-.949c	.343
Visit Count Newspaper Article N	-1.291b	.197
Visit Count Newspaper Article Mean	-2.715b	.007
Visit Count Paragraph1 N	-1.414b	.157
Visit Count Paragraph1 Mean	-2.712b	.007
Visit Count Paragraph2 N	-1.000b	.317
Visit Count Paragraph2 Mean	-.397c	.692
Visit Count Paragraph3 N	-1.414b	.157
Visit Count Paragraph3 Mean	-2.408c	.016
Visit Count Paragraph4 N	-1.732b	.083
Visit Count Paragraph4 Mean	-3.885b	.000
Visit Count Pie Chart N	-1.414b	.157
Visit Count Pie Chart Mean	-.350c	.727
Visit Count Sarah Case N	-.302b	.763
Visit Count Sarah Case Mean	-.338c	.735
Visit Count Task Instruction N	-1.414b	.157
Visit Count Task Instruction Mean	-3.099c	.002
Rate of Total Fixation Duration Essay Sum	-1.290b	.197
Rate of Total Fixation Duration Graph Title2 Sum	-.242b	.809
Rate of Total Fixation Duration Graph Title	-2.715b	.007
Rate of Total Fixation Duration Letter-to-Editor	-1.697b	.090
Rate of Total Fixation Duration Bar Graph	-1.438b	.150
Rate of Total Fixation Duration Newspaper Article	-2.470c	.013
Rate of Total Fixation Duration Paragraph1	-4.288c	.000
Rate of Total Fixation Duration Paragraph2	-.605b	.545
Rate of Total Fixation Duration Paragraph3	-3.447b	.001
Rate of Total Fixation Duration Paragraph4	-3.575c	.000
Rate of Total Fixation Duration Pie Chart	-1.653b	.098
Rate of Total Fixation Duration Sarah Case	-.991b	.322
Rate of Total Fixation Duration Instruction	-5.471b	.000

P-values below .05 are highlighted in **bold**

Appendix M . Wilcoxon signed rank tests of differences in eye-movement measurements in Task 1 and Task 2 (First 10 minutes)

Dependent variable	Z	p-value
Time to First Fixation Essay Mean	-3.810b	.000
Time to First Fixation Graph Title1 Mean	-.766b	.444
Time to First Fixation Graph Title2 Mean	-1.008c	.313
Time to First Fixation Letter-to-Editor Mean	-.804b	.421
Time to First Fixation Line Graph Mean	-2.285c	.022
Time to First Fixation Newspaper Article Mean	-2.293b	.022
Time to First Fixation Paragraph1 Mean	-1.144b	.252
Time to First Fixation Paragraph2 Mean	-3.096b	.002
Time to First Fixation Paragraph3 Mean	-1.703b	.089
Time to First Fixation Paragraph4 Mean	-1.994b	.046
Time to First Fixation Pie Chart Mean	-2.903b	.004
Time to First Fixation Sarah Case Mean	-.886c	.375
Time to First Fixation Task Instruction Mean	-2.043c	.041
First Fixation Duration Essay Mean	-1.132c	.258
First Fixation Duration Graph Title1 Mean	-.406c	.684
First Fixation Duration Graph Title2 Mean	-1.661b	.097
First Fixation Duration Letter-to-Editor Mean	-.863c	.388
First Fixation Duration Line Graph Mean	-1.864b	.062
First Fixation Duration Newspaper Article Mean	-1.831b	.067
First Fixation Duration Paragraph1 Mean	-.348b	.727
First Fixation Duration Paragraph2 Mean	-.784b	.433
First Fixation Duration Paragraph3 Mean	-1.487b	.137
First Fixation Duration Paragraph4 Mean	-1.404b	.160
First Fixation Duration Pie Char Mean	-1.277b	.202
First Fixation Duration Sarah Case Mean	-.243c	.808
First Fixation Duration Task Instruction Mean	-2.162c	.031
Fixation Duration Essay Mean	-.075c	.940
Fixation Duration Graph Title1 Mean	-.334c	.738
Fixation Duration Graph Title2 Mean	-2.573b	.010
Fixation Duration Letter-to-Editor Mean	-.147b	.883
Fixation Duration Line Graph Mean	-3.006b	.003

Fixation Duration Newspaper Article Mean	-2.235b	.025
Fixation Duration Paragraph1 Mean	-4.196b	.000
Fixation Duration Paragraph2 Mean	-3.224b	.001
Fixation Duration Paragraph3 Mean	-2.242b	.025
Fixation Duration Paragraph4 Mean	-2.438b	.015
Fixation Duration Pie Chart Mean	-.566b	.572
Fixation Duration Sarah Case Mean	-.278b	.781
Fixation Duration Task Instruction Mean	-.830b	.406
Total Fixation Duration Essay Mean	-.457c	.648
Total Fixation Duration GraphTitle1 Mean	-.020b	.984
Total Fixation Duration GraphTitle2 Mean	-2.637c	.008
Total Fixation Duration Letter-to-Editor Mean	-.748c	.454
Total Fixation Duration Line Graph Mean	-.988b	.323
Total Fixation Duration Newspaper Article Mean	-2.881b	.004
Total Fixation Duration Paragraph1 Mean	-4.960b	.000
Total Fixation Duration Paragraph2 Mean	-.551b	.582
Total Fixation Duration Paragraph3 Mean	-1.493c	.135
Total Fixation Duration Paragraph4 Mean	-3.647b	.000
Total Fixation Duration Pie Chart Mean	-.935c	.350
Total Fixation Duration Sarah Case Mean	-.852b	.394
Total Fixation Duration Task Instruction Mean	-1.532c	.125
Fixation Count Essay N	-.447b	.655
Fixation Count Essay Mean	-.833c	.405
Fixation Count Graph Title1 N	-.333c	.739
Fixation Count Graph Title1 Mean	-1.285c	.199
Fixation Count Graph Title2 N	-1.265b	.206
Fixation Count Graph Title2 Mean	-2.185c	.029
Fixation Count Letter-to-Editor N	-.243c	.808
Fixation Count Letter-to-Editor Mean	-1.031c	.302
Fixation Count Line Graph N	-1.414b	.157
Fixation Count Line Graph Mean	-.775b	.439
Fixation Count Newspaper Article N	-1.500b	.134
Fixation Count Newspaper Article Mean	-2.788b	.005
Fixation Count Paragraph1 N	-1.732b	.083
Fixation Count Paragraph1 Mean	-4.691b	.000

Fixation Count Paragraph2 N	-1.633b	.102
Fixation Count Paragraph2 Mean	-.319c	.750
Fixation Count Paragraph3 N	-.632b	.527
Fixation Count Paragraph3 Mean	-2.052c	.040
Fixation Count Paragraph4 N	-1.265b	.206
Fixation Count Paragraph4 Mean	-3.162b	.002
Fixation Count Pie Chart N	-1.414b	.157
Fixation Count Pie Chart Mean	-1.465c	.143
Fixation Count Sarah Case N	-.905c	.366
Fixation Count Sarah Case Mean	-.334b	.738
Fixation Count Task Instruction N	.000d	1.000
Fixation Count Task Instruction Mean	-1.821c	.069
Visit Duration Essay N	-1.173b	.241
Visit Duration Essay Mean	-1.835c	.067
Visit Duration Essay Sum	-.887c	.375
Visit Duration GraphTitle1 N	-2.015c	.044
Visit Duration GraphTitle1_Mean	-.502b	.615
Visit Duration Graph Title1 Sum	-1.122c	.262
Visit Duration Graph Title2 N	-.629c	.529
Visit Duration Graph Title2 Mean	-1.432c	.152
Visit Duration Graph Title2 Sum	-1.707c	.088
Visit Duration Letter-to-Editor N	-.462c	.644
Visit Duration Letter-to-Editor Mean	-.673c	.501
Visit Duration Letter-to-Editor Sum	-.860c	.390
Visit Duration Line Graph N	-2.061b	.039
Visit Duration Line Graph Mean	-.405b	.686
Visit Duration Line Graph Sum	-.718b	.473
Visit Duration Newspaper Article N	-2.540b	.011
Visit Duration Newspaper Article Mean	-2.656b	.008
Visit Duration Newspaper Article Sum	-2.930b	.003
Visit Duration Paragraph1 N	-2.869b	.004
Visit Duration Paragraph1 Mean	-4.315b	.000
Visit Duration Paragraph1 Sum	-4.678b	.000
Visit Duration Paragraph2 N	-1.085c	.278
Visit Duration Paragraph2 Mean	-2.023b	.043

Visit Duration Paragraph2 Sum	-.355c	.722
Visit Duration Paragraph3 N	-1.065c	.287
Visit Duration Paragraph3 Mean	-.747b	.455
Visit Duration Paragraph3 Sum	-1.982c	.048
Visit Duration Paragraph4 N	-2.697b	.007
Visit Duration Paragraph4 Mean	-.845b	.398
Visit Duration Paragraph4 Sum	-2.110b	.035
Visit Duration Pie Chart N	-2.314c	.021
Visit Duration Pie Chart Mean	-.081c	.936
Visit Duration Pie Chart Sum	-2.003c	.045
Visit Duration Sarah Case N	-.416c	.677
Visit Duration Sarah Case Mean	-.991b	.322
Visit Duration Sarah Case Sum	-.956b	.339
Visit Duration Task Instruction N	-2.911c	.004
Visit Duration Task Instruction Mean	-1.512b	.130
Visit Duration Task Instruction Sum	-2.225c	.026
Total Visit Duration Essay N	-.447b	.655
Total Visit Duration Essay Mean	-.887c	.375
Total Visit Duration GraphTitle1 N	-.333c	.739
Total Visit Duration GraphTitle1 Mean	-1.151c	.250
Total Visit Duration Graph Title2 N	-1.265b	.206
Total Visit Duration GraphTitle2 Mean	-1.707c	.088
Total Visit Duration Letter-to-Editor N	-.243c	.808
Total Visit Duration Letter-to-Editor Mean	-.860c	.390
Total Visit Duration Line Graph N	-1.414b	.157
Total Visit Duration Line Graph Mean	-.726b	.468
Total Visit Duration Newspaper Article N	-1.500b	.134
Total Visit Duration Newspaper Article Mean	-2.930b	.003
Total Visit Duration Paragraph1 N	-1.732b	.083
Total Visit Duration Paragraph1 Mean	-4.678b	.000
Total Visit Duration Paragraph2 N	-1.633b	.102
Total Visit Duration Paragraph2 Mean	-.355c	.722
Total Visit Duration Paragraph3 N	-.632b	.527
Total Visit Duration Paragraph3 Mean	-1.982c	.048
Total Visit Duration Paragraph4 N	-1.508b	.132

Total Visit Duration Paragraph4 Mean	-2.110b	.035
Total Visit Duration Pie Chart N	-1.414b	.157
Total Visit Duration Pie Chart Mean	-2.003c	.045
Total Visit Duration Sarah Case N	-.905c	.366
Total Visit Duration Sarah Case Mean	-.956b	.339
Total Visit Duration Task Instruction N	-.577b	.564
Total Visit Duration Task Instruction Mean	-2.324c	.020
Visit Count Essay N	-.447b	.655
Visit Count Essay Mean	-1.173b	.241
Visit Count GraphTitle1 N	-.333c	.739
Visit Count GraphTitle1 Mean	-2.015c	.044
Visit Count GraphTitle2 N	-1.265b	.206
Visit Count GraphTitle2 Mean	-.629c	.529
Visit Count Letter-to-Editor N	-.243c	.808
Visit Count Letter-to-Editor Mean	-1.039c	.299
Visit Count Line Graph N	-1.414b	.157
Visit Count Line Graph Mean	-2.061b	.039
Visit Count Newspaper Article N	-1.500b	.134
Visit Count Newspaper Article Mean	-2.540b	.011
Visit Count Paragraph1 N	-1.732b	.083
Visit Count Paragraph1 Mean	-2.869b	.004
Visit Count Paragraph2 N	-1.633b	.102
Visit Count Paragraph2 Mean	-1.085c	.278
Visit Count Paragraph3 N	-.632b	.527
Visit Count Paragraph3 Mean	-1.065c	.287
Visit Count Paragraph4 N	-1.508b	.132
Visit Count Paragraph4 Mean	-2.697b	.007
Visit Count Pie Chart N	-1.414b	.157
Visit Count Pie Chart Mean	-2.314c	.021
Visit Count Sarah Case N	-.905c	.366
Visit Count Sarah Case Mean	-.416c	.677
Visit Count Task Instruction N	-.577b	.564
Visit Count Task Instruction Mean	-2.911c	.004
Rate of Total Fixation Duration Essay	-1.989b	.047
Rate of Total Fixation Duration Graph Title2	-1.653b	.098

Rate of Total Fixation Duration Graph Title	-2.177b	.029
Rate of Total Fixation Duration Letter-to-Editor	-.598b	.550
Rate of Total Fixation Duration Bar Graph	-.565c	.572
Rate of Total Fixation Duration Newspaper Article	-1.646c	.100
Rate of Total Fixation Duration Paragraph1	-4.436c	.000
Rate of Total Fixation Duration Paragraph2	-.573b	.567
Rate of Total Fixation Duration Paragraph3	-1.898b	.058
Rate of Total Fixation Duration Paragraph4	-3.430c	.001
Rate of Total Fixation Duration Pie Chart	-2.460b	.014
Rate of Total Fixation Duration Sarah Case	-.295c	.768
Rate of Total Fixation Duration Task Instruction	-3.091b	.002

P-values below .05 are highlighted in **bold**

Appendix N Wilcoxon signed rank tests of differences in eye-movement measurements in Task 1 and Task 2 (Last 10 minutes)

Dependent variable	Z	p-value
Fixation Duration Essay Mean	-1.642b	.100
Fixation Duration Graph Title1 Mean	-1.297b	.195
Fixation Duration Graph Title2 Mean	-.107b	.915
Fixation Duration Letter-to-Editor Mean	-1.085c	.278
Fixation Duration Line Graph Mean	-.820c	.412
Fixation Duration Newspaper Article Mean	-.471b	.638
Fixation Duration Paragraph1 Mean	-1.698b	.089
Fixation Duration Paragraph2 Mean	-.597b	.551
Fixation Duration Paragraph3 Mean	-.009c	.993
Fixation Duration Paragraph4 Mean	-1.572b	.116
Fixation Duration Pie Chart Mean	-.777c	.437
Fixation Duration Sarah Case Mean	-.355b	.723
Fixation Duration Task Instruction Mean	-.474b	.636
Total Fixation Duration Essay Mean	-.529b	.597
Total Fixation Duration Graph Title2 Mean	-.972b	.331
Total Fixation Duration Graph Title Mean	-1.144c	.253
Total Fixation Duration Letter-to-Editor Mean	-1.224c	.221

Total Fixation Duration Bar Graph Mean	-1.228c	.219
Total Fixation Duration Newspaper Article Mean	-.510b	.610
Total Fixation Duration Paragraph1 Mean	-2.339b	.019
Total Fixation Duration Paragraph2 Mean	-.195b	.845
Total Fixation Duration Paragraph3 Mean	-1.854c	.064
Total Fixation Duration Paragraph4 Mean	-2.776b	.006
Total Fixation Duration Pie Chart Mean	-.067b	.946
Total Fixation Duration Sarah Case Mean	-.243b	.808
Total Fixation Duration Task Instruction Mean	-1.220c	.222
Fixation Count Essay N	-1.732b	.083
Fixation Count Essay Mean	-4.859b	.000
Fixation Count GraphTitle2 N	-2.111c	.035
Fixation Count GraphTitle2 Mean	-3.426c	.001
Fixation Count Graph Title N	-2.324c	.020
Fixation Count Graph Title Mean	-4.153c	.000
Fixation Count Letter-to-Editor N	-3.800c	.000
Fixation Count Letter-to-Editor Mean	-3.649c	.000
Fixation Count Bar Graph N	-2.236c	.025
Fixation Count Bar Graph Mean	-2.821c	.005
Fixation Count Newspaper Article N	-2.837c	.005
Fixation Count Newspaper Article Mean	-2.620c	.009
Fixation Count Paragraph1 N	-1.414b	.157
Fixation Count Paragraph1 Mean	-4.249c	.000
Fixation Count Paragraph2 N	-1.414b	.157
Fixation Count Paragraph2 Mean	-2.473c	.013
Fixation Count Paragraph3 N	-1.387c	.166
Fixation Count Paragraph3 Mean	-4.187c	.000
Fixation Count Paragraph4 N	-1.667b	.096
Fixation Count Paragraph4 Mean	-.988b	.323
Fixation Count Pie Chart N	-.577b	.564
Fixation Count Pie Chart Mean	-2.561c	.010
Fixation Count Sarah Case N	-.943c	.346
Fixation Count Sarah Case Mean	-.018b	.986

Fixation Count Task Instruction N	-3.051c	.002
Fixation Count Task Instruction Mean	-5.310c	.000
Visit Duration Essay N	-3.314b	.001
Visit Duration Essay Mean	-4.503b	.000
Visit Duration Essay Sum	-4.926b	.000
Visit Duration GraphTitle2 N	-3.366c	.001
Visit Duration Graph Title2 Mean	-1.298c	.194
Visit Duration Graph Title2 Sum	-3.133c	.002
Visit Duration Graph Title N	-4.195c	.000
Visit Duration Graph Title Mean	-3.370c	.001
Visit Duration Graph Title Sum	-4.061c	.000
Visit Duration Letter-to-Editor N	-3.412c	.001
Visit Duration Letter-to-Editor Mean	-3.268c	.001
Visit Duration Letter-to-Editor Sum	-3.220c	.001
Visit Duration Bar Graph N	-2.660c	.008
Visit Duration Bar Graph Mean	-1.263c	.207
Visit Duration Bar Graph Sum	-2.093c	.036
Visit Duration Newspaper Article N	-2.821c	.005
Visit Duration Newspaper Article Mean	-1.830c	.067
Visit Duration Newspaper Article Sum	-2.343c	.019
Visit Duration Paragraph1 _N	-3.566c	.000
Visit Duration Paragraph1 Mean	-3.405c	.001
Visit Duration Paragraph1 Sum	-4.186c	.000
Visit Duration Paragraph2 N	-1.241c	.215
Visit Duration Paragraph2 Sum	-2.372c	.018
Visit Duration Paragraph3 N	-3.125c	.002
Visit Duration Paragraph3 Mean	-2.806c	.005
Visit Duration Paragraph3 Sum	-4.051c	.000
Visit Duration Paragraph4 N	-2.084b	.037
Visit Duration Paragraph4 Mean	-.935c	.350
Visit Duration Paragraph4 Sum	-.943b	.346
Visit Duration Pie Chart N	-2.923c	.003
Visit Duration Pie Chart Mean	-1.012c	.312

Visit Duration Pie Chart Sum	-2.637c	.008
Visit Duration Sarah Case N	-.235c	.815
Visit Duration Sarah Case Mean	-.122b	.903
Visit Duration Sarah Case Sum	-.137b	.891
Visit Duration Task Instruction N	-5.215c	.000
Visit Duration Task Instruction Mean	-4.940c	.000
Visit Duration Task Instruction Sum	-5.317c	.000
Visit Count Essay N	-1.732b	.083
Visit Count Essay Mean	-3.314b	.001
Visit Count Graph Title2 N	-2.530c	.011
Visit Count Graph Title2 Mean	-3.366c	.001
Visit Count Graph Title N	-2.324c	.020
Visit Count Graph Title Mean	-4.195c	.000
Visit Count Letter-to-Editor N	-3.800c	.000
Visit Count Letter-to-Editor Mean	-3.412c	.001
Visit Count Bar Graph N	-1.633c	.102
Visit Count Bar Graph Mean	-2.660c	.008
Visit Count Newspaper Article N	-2.837c	.005
Visit Count Newspaper Article Mean	-2.821c	.005
Visit Count Paragraph1 N	-1.414b	.157
Visit Count Paragraph1 Mean	-3.566c	.000
Visit Count Paragraph2 N	-1.414b	.157
Visit Count Paragraph2 Mean	-1.241c	.215
Visit Count Paragraph3 N	-1.387c	.166
Visit Count Paragraph3 Mean	-3.125c	.002
Visit Count Paragraph4 N	-1.667b	.096
Visit Count Paragraph4 Mean	-2.084b	.037
Visit Count Pie Chart N	-.577b	.564
Visit Count Pie Chart Mean	-2.923c	.003
Visit Count Sarah Case N	-.943c	.346
Visit Count Sarah Case Mean	-.235c	.815
Visit Count Task Instruction N	-3.051c	.002
Visit Count Task Instruction Mean	-5.215c	.000

Rate of Total Fixation Duration Essay	-.941b	.347
Rate of Total Fixation Duration Graph Title2	-1.290b	.197
Rate of Total Fixation Duration Graph Title	-.491c	.623
Rate of Total Fixation Duration Letter-to-Editor	-1.223c	.221
Rate of Total Fixation Duration Bar Graph	-.419c	.675
Rate of Total Fixation Duration Newspaper Article	-.078b	.937
Rate of Total Fixation Duration Paragraph1	-2.473b	.013
Rate of Total Fixation Duration Paragraph2	-.605b	.545
Rate of Total Fixation Duration Paragraph3	-2.074c	.038
Rate of Total Fixation Duration Paragraph4	-4.624c	.000
Rate of Total Fixation Duration Pie Chart	-.067b	.946
Rate of Total Fixation Duration Sarah Case	-.112b	.911
Rate of Total Fixation Duration Task Instruction	-1.425c	.154
P-values below .05 are highlighted in bold		

Appendix O Mann-Whitney tests of difference between participants with lower and upper intermediate skills in reading and writing (40 minutes)

Dependent variable	Task 1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Time to First Fixation Essay Mean	-1.592	.111	-.773	.440	-2.354	.019	-1.407	.160
Time to First Fixation Graph Title1 Mean	-.312	.755	-.433	.665	-2.453	.014	-1.014	.310
Time to First Fixation Graph Title2 Mean	-.043	.966	-.392	.695	-1.361	.173	-2.326	.020
Time to First Fixation Letter-to-Editor Mean	-.686	.493	-.559	.576	-.455	.649	-.054	.957
Time to First Fixation Line Graph Mean	-.681	.496	-.514	.607	-.780	.435	-1.366	.172
Time to First Fixation Newspaper Article Mean	-.118	.906	-1.379	.168	-2.235	.025	-.219	.827
Time to First Fixation Paragraph1 Mean	-1.008	.313	-1.016	.310	-.128	.898	-.122	.903
Time to First Fixation Paragraph2 Mean	-.142	.887	-.068	.946	-1.815	.070	-1.812	.070
Time to First Fixation Paragraph3 Mean	-.071	.943	-.460	.645	-.170	.865	-.703	.482
Time to First Fixation Paragraph4 Mean	-.284	.777	-.433	.665	-.113	.910	-.595	.552
Time to First Fixation Pie Chart Mean	-.043	.966	-1.204	.229	-.171	.865	-.325	.745
Time to First Fixation Sarah Case Mean	-1.710	.087	-1.686	.092	-1.702	.089	-1.294	.196
Time to First Fixation Task Instruction Mean	-.085	.932	-.108	.914	-.468	.640	-.650	.516
First Fixation Duration Essay Mean	-.710	.478	-1.057	.291	-1.535	.125	-.597	.551
First Fixation Duration Graph Title1 Mean	-.639	.523	-.488	.626	-.625	.532	-.014	.989
First Fixation Duration Graph Title2 Mean	-.341	.733	-.136	.892	-.582	.560	-.542	.588
First Fixation Duration Letter-to-Editor Mean	-.686	.492	-.177	.859	-.185	.853	-.475	.635
First Fixation Duration Line Graph Mean	-.909	.363	-.542	.588	-.156	.876	-.014	.989

First Fixation Duration Newspaper Article Mean	-.205	.837	-1.692	.091	-1.778	.075	-.820	.412
First Fixation Duration Paragraph1 Mean	-.298	.766	-1.029	.303	-1.052	.293	-1.206	.228
First Fixation Duration Paragraph2 Mean	-.397	.691	-.135	.892	-.965	.335	-1.286	.198
First Fixation Duration Paragraph3 Mean	-.514	.607	-1.076	.282	-.028	.977	-1.031	.303
First Fixation Duration Paragraph4 Mean	-1.173	.241	-.633	.527	-2.060	.039	-2.291	.022
First Fixation Duration Pie Chart Mean	-.910	.363	-.146	.884	-.937	.349	-.528	.597
First Fixation Duration Sarah Case Mean	-1.899	.058	-1.327	.184	-.433	.665	-1.074	.283
First Fixation Duration Task Instruction Mean	-.866	.386	-1.125	.261	-.498	.618	-.991	.322
Total Fixation Duration Essay Mean	-.626	.532	-.326	.745	-.627	.531	-1.169	.242
Total Fixation Duration Graph Title1 Mean	-1.009	.313	-.420	.674	-.156	.876	-.014	.989
Total Fixation Duration Graph Title2 Mean	-.128	.898	-.679	.497	-.725	.468	-.095	.924
Total Fixation Duration Letter-to-Editor Mean	-.429	.668	-.286	.775	-.740	.459	-.081	.935
Total Fixation Duration Line Graph Mean	-.071	.943	-1.006	.314	-.573	.566	-1.709	.087
Total Fixation Duration Newspaper Article Mean	-.117	.907	-1.301	.193	-1.147	.252	-.684	.494
Total Fixation Duration Paragraph1 Mean	-.100	.921	-1.181	.238	-1.110	.267	-.557	.578
Total Fixation Duration Paragraph2 Mean	-.256	.798	-1.235	.217	-.755	.450	-1.222	.222
Total Fixation Duration Paragraph3 Mean	-.014	.989	-.014	.989	-.825	.409	-.434	.664
Total Fixation Duration Paragraph4 Mean	-.585	.559	-.789	.430	-1.794	.073	-.136	.892
Total Fixation Duration Pie Chart Mean	-.584	.559	-.122	.903	-.214	.831	-1.046	.295
Total Fixation Duration Sarah Case Mean	-1.349	.177	-.885	.376	-.260	.795	-.730	.466
Total Fixation Duration Task Instruction Mean	-.356	.722	0.000	1.000	-.157	.876	-.041	.967
Total Fixation Duration Essay Mean	-.737	.461	-.852	.394	-2.098	.036	-2.218	.027
Total Fixation Duration GraphTitle1 Mean	-.964	.335	-.568	.570	-.510	.610	-.487	.626

Total Fixation Duration GraphTitle2 Mean	-.539	.590	-.595	.552	-.879	.379	-.257	.797
Total Fixation Duration Letter-to-Editor Mean	-.700	.484	-.709	.478	-.967	.334	-.922	.356
Total Fixation Duration Line Graph Mean	-.028	.977	-.500	.617	-.454	.650	-1.217	.224
Total Fixation Duration Newspaper Article Mean	-.381	.703	-1.552	.121	-.086	.931	-.574	.566
Total Fixation Duration Paragraph1 Mean	-1.219	.223	-1.177	.239	-.312	.755	-.460	.646
Total Fixation Duration Paragraph2 Mean	-.028	.977	-.906	.365	-1.021	.307	-.027	.978
Total Fixation Duration Paragraph3 Mean	-.539	.590	-.514	.607	-.142	.887	-.487	.626
Total Fixation Duration Paragraph4 Mean	-.879	.379	-1.190	.234	-.766	.444	-.352	.725
Total Fixation Duration Pie Chart Mean	-1.106	.269	-.284	.776	-1.021	.307	-1.840	.066
Total Fixation Duration Sarah Case Mean	-1.811	.070	-1.313	.189	-.707	.480	-1.087	.277
Total Fixation Duration Task Instruction Mean	-.170	.865	-.771	.441	-.312	.755	-.460	.646
Fixation Count Essay N	-0.748	0.454	-0.975	0.646	0	1	0	0.33
Fixation Count Essay Mean	-0.879	0.379	-0.906	1	-2.268	0.023	-2.272	0.365
Fixation Count Graph Title N	-1.305	0.192	-0.593	0.023	0	1	0	0.553
Fixation Count Graph Title Mean	-0.709	0.478	-0.311	1	-0.71	0.478	-0.528	0.756
Fixation Count Graph Title2 N	-1.305	0.192	-0.593	0.598	0	1	0	0.553
Fixation Count Graph Title2 Mean	-0.809	0.419	-0.501	1	-0.979	0.328	-0.406	0.617
Fixation Count Letter-to-Editor N	-1.134	0.257	0	0.685	-0.474	0.636	-0.411	1
Fixation Count Letter-to-Editor Mean	-0.744	0.457	-0.778	0.681	-0.899	0.369	-1.034	0.437
Fixation Count Line Graph N	-1.051	0.293	0	0.301	0	1	0	1
Fixation Count Line Graph Mean	-0.156	0.876	-0.284	1	-0.681	0.496	-1.001	0.776
Fixation Count Newspaper Article N	-0.401	0.688	-1.275	0.317	-1.576	0.115	-0.35	0.202
Fixation Count Newspaper Article Mean	-0.575	0.565	-1.714	0.727	-0.101	0.92	-0.549	0.087

Fixation Count Paragraph1 N	-1.051	0.293	0	0.583	0	1	0	1
Fixation Count Paragraph1 Mean	-1.489	0.137	-1.123	1	-0.709	0.478	-0.46	0.262
Fixation Count Paragraph2 N	-1.051	0.293	0	0.646	0	1	0	1
Fixation Count Paragraph2 Mean	-0.099	0.921	-0.676	1	-0.808	0.419	-0.203	0.499
Fixation Count Paragraph3 N	-1.305	0.192	-0.593	0.839	-0.734	0.463	-1	0.553
Fixation Count Paragraph3 Mean	-0.567	0.571	-0.406	0.317	-0.44	0.66	-0.487	0.685
Fixation Count Paragraph4 N	-1.305	0.192	-0.593	0.626	0	1	0	0.553
Fixation Count Paragraph4 Mean	-0.979	0.328	-1.272	1	-0.581	0.561	-0.176	0.203
Fixation Count Pie Chart N	-1.051	0.293	0	0.86	0	1	0	1
Fixation Count Pie Chart Mean	-1.475	0.14	-0.243	1	-1.078	0.281	-1.664	0.808
Fixation Count Sarah Case N	-1.99	0.047	-1.964	0.096	-1.083	0.279	-1	0.05
Fixation Count Sarah Case Mean	-1.713	0.087	-1.177	0.317	-0.997	0.319	-1.268	0.239
Fixation Count Task Instruction N	-1.051	0.293	0	0.205	0	1	0	1
Fixation Count Task Instruction Mean	-0.014	0.989	-0.798	1	-0.425	0.671	-0.622	0.425
Visit Duration Essay N	-1.843	0.065	-1.339	0.534	-1.858	0.063	-1.677	0.181
Visit Duration Essay Mean	-0.142	0.887	-0.122	0.093	-0.156	0.876	-0.365	0.903
Visit Duration Essay Sum	-1.219	0.223	-1.069	0.715	-1.73	0.084	-1.434	0.285
Visit Duration Graph Title N	-0.966	0.334	-0.23	0.152	-1.292	0.196	-0.988	0.818
Visit Duration Graph Title Mean	-0.511	0.609	-0.731	0.323	-0.78	0.435	-1.245	0.465
Visit Duration Graph Title Sum	-0.993	0.321	-0.649	0.213	-0.425	0.671	-0.189	0.516
Visit Duration Graph Title2 N	-0.866	0.386	-0.515	0.85	-0.128	0.898	-0.041	0.607
Visit Duration Graph Title2 Mean	-0.156	0.876	-0.69	0.968	-1.489	0.136	-0.379	0.49
Visit Duration Graph Title2 Sum	-0.539	0.59	-0.514	0.705	-1.191	0.234	-0.446	0.607

Visit Duration Letter-to-Editor N	-0.917	0.359	-0.438	0.655	-0.757	0.449	-1.131	0.662
Visit Duration Letter-to-Editor Mean	-0.243	0.808	-0.573	0.258	-0.768	0.443	-0.719	0.567
Visit Duration Letter-to-Editor Sum	-0.7	0.484	-0.709	0.472	-0.782	0.434	-1.058	0.478
Visit Duration Line Graph N	-0.411	0.681	-0.162	0.29	-1.404	0.16	-0.717	0.871
Visit Duration Line Graph Mean	-1.191	0.233	0	0.473	-0.737	0.461	-1.015	1
Visit Duration Line Graph Sum	-0.907	0.364	-0.122	0.31	-0.142	0.887	-0.947	0.903
Visit Duration Newspaper Article N	-0.273	0.785	-1.487	0.344	-0.302	0.762	-0.22	0.137
Visit Duration Newspaper Article Mean	-0.437	0.662	-1.071	0.826	-1.061	0.289	-1.258	0.284
Visit Duration Newspaper Article Sum	-0.256	0.798	-1.403	0.208	-0.186	0.852	-0.642	0.161
Visit Duration Paragraph1 N	-0.993	0.321	-0.528	0.521	-1.375	0.169	-0.744	0.598
Visit Duration Paragraph1 Mean	-0.993	0.321	-0.514	0.457	-1.829	0.067	-0.041	0.607
Visit Duration Paragraph1 Sum	-1.418	0.156	-0.987	0.968	-0.17	0.865	-0.325	0.323
Visit Duration Paragraph2 N	-0.397	0.691	-0.622	0.745	-0.468	0.64	-0.081	0.534
Visit Duration Paragraph2 Mean	-1.616	0.106	-0.23	0.935	-1.276	0.202	-0.784	0.818
Visit Duration Paragraph2 Sum	-1.021	0.307	-0.582	0.433	-1.333	0.183	-0.676	0.561
Visit Duration Paragraph3 N	-1.078	0.281	-0.717	0.499	-0.681	0.496	-0.663	0.473
Visit Duration Paragraph3 Mean	-1.049	0.294	-0.135	0.507	-0.468	0.64	-0.419	0.892
Visit Duration Paragraph3 Sum	0	1	-0.352	0.675	-0.142	0.887	-0.325	0.725
Visit Duration Paragraph4 N	-0.965	0.335	-1.272	0.745	-0.17	0.865	0	0.203
Visit Duration Paragraph4 Mean	-0.028	0.977	-0.189	1	-1.248	0.212	-0.135	0.85
Visit Duration Paragraph4 Sum	-0.482	0.63	-0.487	0.892	-1.163	0.245	-0.487	0.626
Visit Duration Pie Chart N	-1.773	0.076	-0.162	0.626	-0.823	0.411	-1.245	0.871
Visit Duration Pie Chart Mean	-0.071	0.943	-1.163	0.213	-0.709	0.478	-0.758	0.245

Visit Duration Pie Chart Sum	-1.078	0.281	-0.717	0.449	-0.227	0.821	-1.569	0.473
Visit Duration Sarah Case N	-1.393	0.163	-1.179	0.117	-0.594	0.553	-0.87	0.238
Visit Duration Sarah Case Mean	-1.492	0.136	-1.475	0.384	-1.443	0.149	-1.789	0.14
Visit Duration Sarah Case Sum	-1.956	0.05	-1.424	0.074	-0.895	0.371	-1.17	0.154
Visit Duration Task Instruction N	-0.085	0.932	-0.893	0.242	-0.355	0.723	-0.366	0.372
Visit Duration Task Instruction Mean	-0.397	0.691	-0.987	0.715	-0.17	0.865	-0.203	0.323
Visit Duration Task Instruction Sum	-0.482	0.63	-1.258	0.839	-0.227	0.821	-0.622	0.208
Total Visit Duration Essay N	-1.051	0.293	0	0.534	0	1	0	1
Total Visit Duration Essay Mean	-1.219	0.223	-1.069	1	-1.73	0.084	-1.434	0.285
Total Visit Duration Graph Title N	-1.305	0.192	-0.593	0.152	0	1	0	0.553
Total Visit Duration Graph Title Mean	-0.993	0.321	-0.649	1	-0.425	0.671	-0.189	0.516
Total Visit Duration Graph Title2 N	-1.305	0.192	-0.593	0.85	0	1	0	0.553
Total Visit Duration Graph Title2 Mean	-0.539	0.59	-0.514	1	-1.191	0.234	-0.446	0.607
Total Visit Duration Letter-to-Editor N	-1.134	0.257	0	0.655	-0.474	0.636	-0.411	1
Total Visit Duration Letter-to-Editor Mean	-0.7	0.484	-0.709	0.681	-0.782	0.434	-1.058	0.478
Total Visit Duration Line Graph N	-1.051	0.293	0	0.29	0	1	0	1
Total Visit Duration Line Graph Mean	-0.907	0.364	-0.122	1	-0.142	0.887	-0.947	0.903
Total Visit Duration Newspaper Article N	-0.401	0.688	-1.275	0.344	-1.576	0.115	-0.35	0.202
Total Visit Duration Newspaper Article Mean	-0.381	0.703	-1.552	0.727	-0.186	0.852	-0.642	0.121
Total Visit Duration Paragraph1 N	-1.051	0.293	0	0.521	0	1	0	1
Total Visit Duration Paragraph1 Mean	-1.418	0.156	-0.987	1	-0.17	0.865	-0.325	0.323
Total Visit Duration Paragraph2 N	-1.051	0.293	0	0.745	0	1	0	1
Total Visit Duration Paragraph2 Mean	-1.021	0.307	-0.582	1	-1.333	0.183	-0.676	0.561

Total Visit Duration Paragraph3 N	-1.305	0.192	-0.593	0.499	-0.734	0.463	-1	0.553
Total Visit Duration Paragraph3 Mean	0	1	-0.352	0.317	-0.142	0.887	-0.325	0.725
Total Visit Duration Paragraph4 N	-1.305	0.192	-0.593	0.745	0	1	0	0.553
Total Visit Duration Paragraph4 Mean	-0.482	0.63	-0.487	1	-1.163	0.245	-0.487	0.626
Total Visit Duration Pie Chart N	-1.051	0.293	0	0.626	0	1	0	1
Total Visit Duration Pie Chart Mean	-1.078	0.047	-0.717	1	-0.227	0.281	-1.569	0.473
Total Visit Duration Sarah Case N	-1.99	0.05	-1.964	0.117	-1.083	0.821	-1	0.05
Total Visit Duration Sarah Case Mean	-1.956	0.293	-1.424	0.317	-0.895	0.279	-1.17	0.154
Total Visit Duration Task Instruction N	-1.051	0.63	0	0.242	0	0.371	0	1
Total Visit Duration Task Instruction Mean	-0.482	0.293	-1.258	1	-0.227	1	-0.622	0.208
Visit Count Essay N	-1.051	0.065	0	0.534	0	0.821	0	1
Visit Count Essay Mean	-1.843	0.192	-1.339	1	-1.858	1	-1.677	0.181
Visit Count Graph Title N	-1.305	0.334	-0.593	0.093	0	0.063	0	0.553
Visit Count Graph Title Mean	-0.966	0.192	-0.23	1	-1.292	1	-0.988	0.818
Visit Count Graph Title2 N	-1.305	0.386	-0.593	0.323	0	0.196	0	0.553
Visit Count Graph Title2 Mean	-0.866	0.257	-0.515	1	-0.128	1	-0.041	0.607
Visit Count Letter-to-Editor N	-1.134	0.359	0	0.968	-0.474	0.898	-0.411	1
Visit Count Letter-to-Editor Mean	-0.917	0.293	-0.438	0.681	-0.757	0.636	-1.131	0.662
Visit Count Line Graph N	-1.051	0.681	0	0.258	0	0.449	0	1
Visit Count Line Graph Mean	-0.411	0.688	-0.162	1	-1.404	1	-0.717	0.871
Visit Count Newspaper Article N	-0.401	0.69	-1.275	0.473	-1.576	0.16	-0.35	0.202
Visit Count Newspaper Article Mean	-0.398	0.293	-1.633	0.727	-0.302	0.115	-0.22	0.102
Visit Count Paragraph1 N	-1.051	0.321	0	0.826	0	0.762	0	1

Visit Count Paragraph1 Mean	-0.993	0.454	-0.528	1	-1.375	1	-0.744	0.598
Visit Count Paragraph2 N	-0.748	0.691	-0.975	0.457	0	0.169	0	0.33
Visit Count Paragraph2 Mean	-0.397	0.192	-0.622	1	-0.468	1	-0.081	0.534
Visit Count Paragraph3 N	-1.305	0.281	-0.593	0.935	-0.734	0.64	-1	0.553
Visit Count Paragraph3 Mean	-1.078	0.192	-0.717	0.317	-0.681	0.463	-0.663	0.473
Visit Count Paragraph4 N	-1.305	0.335	-0.593	0.507	0	0.496	0	0.553
Visit Count Paragraph4 Mean	-0.965	0.293	0	1	-0.17	1	1.272	0.203
Visit Count Pie Chart N	-1.051	0.076	0	1	0	0.865	0	1
Visit Count Pie Chart Mean	-1.773	0.047	-0.162	1	-0.823	1	0	0.871
Visit Count Sarah Case N	-1.99	0.127	-1.964	0.213	-1.083	0.411	-1.245	0.05
Visit Count Sarah Case Mean	-1.528	0.293	-0.999	0.317	-0.594	0.279	-1	0.318
Visit Count Task Instruction N	-1.051	0.932	0	0.384	0	0.553	-0.87	1
Visit Count Task Instruction Mean	-0.085	0.454	-0.893	1	-0.355	1	0	0.372
Rate of Total Fixation Duration Essay	-0.879	.379	-0.446	.655	-2.467	.014	-1.839	.066
Rate of Total Fixation Duration Graph Title2	-0.822	.411	-0.189	.850	-0.369	.712	0.000	1.000
Rate of Total Fixation Duration Graph Title	-0.142	.887	-0.054	.957	-1.560	.119	-1.190	.234
Rate of Total Fixation Duration Letter-to-Editor	-0.829	.407	-0.368	.713	-1.080	.280	-0.922	.356
Rate of Total Fixation Duration Bar Graph	-0.170	.865	-0.311	.756	-0.681	.496	-0.298	.766
Rate of Total Fixation Duration Newspaper Article	-0.308	.758	-1.230	.219	-0.372	.710	-0.957	.339
Rate of Total Fixation Duration Paragraph1	-1.560	.119	-0.852	.394	-1.191	.234	-1.055	.291
Rate of Total Fixation Duration Paragraph2	-1.106	.269	-0.717	.473	-1.900	.057	-1.380	.168
Rate of Total Fixation Duration Paragraph3	-0.199	.843	-0.406	.685	-0.822	.411	-0.920	.358
Rate of Total Fixation Duration Paragraph4	-0.737	.461	-0.676	.499	-1.475	.140	-1.461	.144

Rate of Total Fixation Duration Pie Chart	-1.049	.294	-.771	.441	-.085	.932	-1.001	.317
Rate of Total Fixation Duration Sarah Case	-2.753	.006	-1.769	.077	-.635	.526	-.716	.474
Rate of Total Fixation Duration Task Instruction	-.624	.533	-.960	.337	-1.191	.234	-1.839	.066

Notes: P-values are exact (2-tailed). P-values below .05 are highlighted in **bold**. N=40 in all comparisons.

Appendix P Mann-Whitney tests of difference between participants with lower and upper intermediate skills in reading and writing (First 10 minutes)

Dependent variable	Task 1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Time to First Fixation Essay Mean	-.954	.340	-.299	.765	-1.560	.119	-.325	.745
Time to First Fixation Graph Title1 Mean	-.383	.702	-.961	.337	-.540	.589	-.515	.606
Time to First Fixation Graph Title2 Mean	-.484	.628	-.421	.674	-.170	.865	-.703	.482
Time to First Fixation Letter-to-Editor Mean	-.073	.942	-1.580	.114	-1.033	.302	-.816	.414
Time to First Fixation Line Graph Mean	-.681	.496	-.514	.607	-.780	.435	-1.366	.172
Time to First Fixation Newspaper Article Mean	-.552	.581	-.673	.501	-1.304	.192	-.304	.761
Time to First Fixation Paragraph1 Mean	-.640	.522	-.529	.597	-.128	.898	-.122	.903
Time to First Fixation Paragraph2 Mean	-.711	.477	-.570	.569	-1.787	.074	-1.299	.194
Time to First Fixation Paragraph3 Mean	-.427	.669	-.489	.625	-1.561	.119	-.298	.766
Time to First Fixation Paragraph4 Mean	-1.829	.067	-.750	.453	-.454	.650	-.027	.978
Time to First Fixation Pie Chart Mean	-.043	.966	-1.204	.229	-.171	.865	-.325	.745
Time to First Fixation Sarah Case Mean	-.599	.549	-.841	.400	-.499	.618	-.206	.836
Time to First Fixation Task Instruction Mean	-.085	.932	-.108	.914	-.099	.921	-.149	.882
First Fixation Duration Essay Mean	-.235	.815	-1.548	.122	-1.094	.274	-1.057	.290
First Fixation Duration Graph Title1 Mean	-.483	.629	-.556	.579	-1.763	.078	-1.397	.162
First Fixation Duration Graph Title2 Mean	-.883	.377	-.081	.935	-.170	.865	-1.572	.116
First Fixation Duration Letter-to-Editor Mean	-.249	.803	-.462	.644	-.694	.488	-.042	.966

First Fixation Duration Line Graph Mean	-.909	.363	-.542	.588	-.156	.876	-.014	.989
First Fixation Duration Newspaper Article Mean	-.091	.928	-1.070	.285	-1.102	.271	-1.217	.224
First Fixation Duration Paragraph1 Mean	-.582	.560	-.610	.542	-1.052	.293	-1.206	.228
First Fixation Duration Paragraph2 Mean	-.782	.434	-.190	.849	-.383	.702	-.433	.665
First Fixation Duration Paragraph3 Mean	-.699	.484	-1.212	.226	-.526	.599	-1.207	.228
First Fixation Duration Paragraph4 Mean	-.215	.830	-.709	.478	-1.507	.132	-1.695	.090
First Fixation Duration Pie Chart Mean	-1.251	.211	-.136	.892	-.937	.349	-.528	.597
First Fixation Duration Sarah Case Mean	-.362	.717	-.842	.400	-.949	.343	-.079	.937
First Fixation Duration Task Instruction Mean	-.866	.386	-1.125	.261	-.612	.541	-1.140	.254
Fixation Duration Essay Mean	-.398	.691	-1.166	.244	-1.747	.081	-1.897	.058
Fixation Duration Graph Title1 Mean	-.455	.649	-.217	.828	-1.649	.099	-1.546	.122
Fixation Duration Graph Title2 Mean	-.813	.416	-.734	.463	-.043	.966	-1.667	.095
Fixation Duration Letter-to-Editor Mean	-.249	.803	-.378	.706	-.863	.388	-.411	.681
Fixation Duration Line Graph Mean	-.654	.513	-.922	.357	-.515	.607	-1.064	.287
Fixation Duration Newspaper Article Mean	-.227	.820	-.607	.544	-.971	.332	-1.244	.213
Fixation Duration Paragraph1 Mean	-.384	.701	-1.140	.254	-1.835	.067	-.109	.914
Fixation Duration Paragraph2 Mean	-.313	.754	-.380	.704	-.711	.477	-.570	.569
Fixation Duration Paragraph3 Mean	-.043	.966	-.217	.828	-1.506	.132	-.258	.797
Fixation Duration Paragraph4 Mean	-.575	.565	-.014	.989	-.327	.744	-.434	.664
Fixation Duration Pie Chart Mean	-.015	.988	-.353	.724	-1.069	.285	-1.074	.283

Fixation Duration Sarah Case Mean	-.646	.518	-.797	.426	-.849	.396	-.095	.924
Fixation Duration Task Instruction Mean	-.456	.648	-.095	.924	-.171	.864	-.041	.967
Total Fixation Duration Essay Mean	-.709	.478	-.839	.402	-2.581	.010	-2.840	.005
Total Fixation Duration Graph Title1 Mean	-.511	.610	-.095	.925	-.468	.640	-.947	.344
Total Fixation Duration Graph Title2 Mean	-1.592	.111	-1.329	.184	-1.221	.222	-.866	.386
Total Fixation Duration Letter-to-Editor Mean	-.059	.953	-.923	.356	-.856	.392	-1.210	.226
Total Fixation Duration Line Graph Mean	-1.460	.144	-1.528	.126	-1.900	.057	-2.408	.016
Total Fixation Duration Newspaper Article Mean	-.637	.524	-.564	.573	-.232	.817	-1.134	.257
Total Fixation Duration Paragraph1 Mean	-1.843	.065	-1.055	.291	-.539	.590	-.352	.725
Total Fixation Duration Paragraph2 Mean	-.768	.443	-.461	.645	-.596	.551	-.433	.665
Total Fixation Duration Paragraph3 Mean	-.270	.787	-.163	.871	-1.220	.222	-.298	.766
Total Fixation Duration Paragraph4 Mean	-.743	.457	-.586	.558	-1.192	.233	-.379	.705
Total Fixation Duration Pie Chart Mean	-1.049	.294	-.473	.636	-1.475	.140	-1.758	.079
Total Fixation Duration Sarah Case Mean	-.788	.431	-.977	.329	-.632	.527	-.048	.962
Total Fixation Duration Task Instruction Mean	-.425	.671	-.933	.351	-.369	.712	-.568	.570
Fixation Count Essay N	-.062	.950	-.593	.553	-1.051	.293	-1.433	.152
Fixation Count Essay Mean	-.936	.349	-1.313	.189	-2.425	.015	-2.651	.008
Fixation Count Graph Title1 N	-.436	.663	-1.041	.298	-1.732	.083	-1.416	.157
Fixation Count Graph Title1 Mean	-.526	.599	-.095	.924	-1.053	.293	-.556	.578
Fixation Count Graph Title2 N	-1.249	.212	-.411	.681	-1.305	.192	-1.778	.075

Fixation Count Graph Title2 Mean	-1.664	.096	-1.289	.198	-.625	.532	-.894	.371
Fixation Count Letter-to-Editor N	-.401	.688	-.637	.524	-.695	.487	-.947	.343
Fixation Count Letter-to-Editor Mean	-.249	.803	-.840	.401	-.739	.460	-1.354	.176
Fixation Count Line Graph N	-.062	.950	-.593	.553	-.734	.463	-1.000	.317
Fixation Count Line Graph Mean	-1.407	.159	-1.885	.059	-2.199	.028	-2.449	.014
Fixation Count Newspaper Article N	-.655	.513	-.624	.532	-.755	.450	-.655	.513
Fixation Count Newspaper Article Mean	-.973	.331	-.435	.664	-.131	.896	-.972	.331
Fixation Count Paragraph1 N	-1.305	.192	-.593	.553	0.000	1.000	0.000	1.000
Fixation Count Paragraph1 Mean	-2.227	.026	-1.137	.256	-1.021	.307	-.311	.756
Fixation Count Paragraph2 N	-1.249	.212	-.411	.681	-.062	.950	-.593	.553
Fixation Count Paragraph2 Mean	-.753	.451	-.258	.797	-.284	.777	-.338	.735
Fixation Count Paragraph3 N	-1.249	.212	-.411	.681	-1.237	.216	-.472	.637
Fixation Count Paragraph3 Mean	-.427	.669	-.095	.924	-1.122	.262	-.501	.616
Fixation Count Paragraph4 N	-.602	.547	-.289	.773	-.247	.805	-.472	.637
Fixation Count Paragraph4 Mean	-.557	.577	-.491	.623	-1.037	.300	-.352	.725
Fixation Count Pie Chart N	-1.051	.293	0.000	1.000	0.000	1.000	0.000	1.000
Fixation Count Pie Chart Mean	-1.069	.285	-.675	.500	-1.390	.164	-1.624	.104
Fixation Count Sarah Case N	-.033	.974	-.947	.343	-.618	.537	0.000	1.000
Fixation Count Sarah Case Mean	-.634	.526	-1.118	.263	-.483	.629	-.016	.987
Fixation Count Task Instruction N	-.748	.454	-.975	.330	-.734	.463	-1.000	.317

Fixation Count Task Instruction Mean	-.326	.744	-1.042	.298	-.440	.660	-.745	.456
Visit Duration Essay N	-1.476	.140	-1.137	.256	-2.143	.032	-2.436	.015
Visit Duration Essay Mean	-.567	.571	-.541	.588	-1.631	.103	-1.786	.074
Visit Duration Essay Sum	-.993	.321	-1.353	.176	-2.070	.038	-2.029	.042
Visit Duration Graph Title1 N	-.725	.469	-.298	.765	-1.168	.243	-.842	.400
Visit Duration Graph Title1 Mean	-.326	.744	-.393	.695	-1.135	.256	-.623	.533
Visit Duration Graph Title1 Sum	-.610	.542	-.176	.860	-1.220	.222	-.839	.401
Visit Duration Graph Title2 N	-2.008	.045	-1.467	.142	-1.309	.190	-1.032	.302
Visit Duration Graph Title2 Mean	-.356	.722	-.515	.606	-.043	.966	-1.042	.297
Visit Duration Graph Title2 Sum	-1.592	.111	-1.424	.154	-.440	.660	-.961	.337
Visit Duration Letter-to-Editor N	-.678	.498	-.703	.482	-.653	.514	-1.387	.166
Visit Duration Letter-to-Editor Mean	-.513	.608	-.825	.409	-1.048	.295	-1.140	.254
Visit Duration Letter-to-Editor Sum	-.073	.942	-.937	.349	-.885	.376	-1.295	.195
Visit Duration Line Graph N	-1.194	.233	-.976	.329	-2.072	.038	-1.638	.101
Visit Duration Line Graph Mean	-.234	.815	-.998	.318	-.950	.342	-1.515	.130
Visit Duration Line Graph Sum	-.878	.380	-1.445	.148	-1.957	.050	-1.948	.051
Visit Duration Newspaper Article N	-.823	.411	-.567	.571	-.437	.662	-.779	.436
Visit Duration Newspaper Article Mean	-.516	.606	-.492	.623	-1.029	.303	-1.465	.143
Visit Duration Newspaper Article Sum	-.606	.544	-.593	.553	-.232	.817	-1.134	.257
Visit Duration Paragraph1 N	-1.589	.112	-.460	.645	-1.234	.217	-.311	.756

Visit Duration Paragraph1 Mean	-.766	.444	-.325	.745	-1.517	.129	-.717	.473
Visit Duration Paragraph1 Sum	-2.014	.044	-.649	.516	-.454	.650	-.162	.871
Visit Duration Paragraph2 N	-.754	.451	-.027	.978	-.428	.669	-.082	.935
Visit Duration Paragraph2 Mean	-.739	.460	-.488	.625	-.170	.865	-.487	.626
Visit Duration Paragraph2 Sum	-.426	.670	-.163	.871	-.425	.671	-.095	.925
Visit Duration Paragraph3 N	-.242	.809	-.109	.913	-.498	.619	-.922	.356
Visit Duration Paragraph3 Mean	-.441	.659	-.271	.786	-1.277	.201	-.663	.507
Visit Duration Paragraph3 Sum	-.128	.898	-.271	.786	-1.476	.140	-.812	.417
Visit Duration Paragraph4 N	-.415	.678	-.478	.632	-1.094	.274	-1.125	.260
Visit Duration Paragraph4 Mean	-.600	.548	-.273	.785	-.057	.955	-.447	.655
Visit Duration Paragraph4 Sum	-.857	.391	-.913	.361	-1.362	.173	-.758	.448
Visit Duration Pie Chart N	-1.248	.212	-.379	.705	-1.264	.206	-1.327	.184
Visit Duration Pie Chart Mean	-.766	.444	-.595	.552	-.610	.542	-.162	.871
Visit Duration Pie Chart Sum	-1.361	.173	-.122	.903	-1.361	.173	-1.325	.185
Visit Duration Sarah Case N	-.479	.632	-.975	.329	-.486	.627	-.016	.987
Visit Duration Sarah Case Mean	-.599	.549	-.917	.359	-.666	.506	-.238	.812
Visit Duration Sarah Case Sum	-.788	.431	-.992	.321	-.599	.549	-.111	.912
Visit Duration Task Instruction N	-.043	.966	-1.177	.239	-.625	.532	-.095	.924
Visit Duration Task Instruction Mean	-.468	.640	-1.015	.310	-.085	.932	-.027	.978
Visit Duration Task Instruction Sum	-.766	.444	-1.583	.114	-.085	.932	-.460	.646

Total Visit Duration Essay N	-.062	.950	-.593	.553	-1.051	.293	-1.433	.152
Total Visit Duration Essay Mean	-.993	.321	-1.353	.176	-2.070	.038	-2.029	.042
Total Visit Duration Graph Title1_N	-.436	.663	-1.041	.298	-1.732	.083	-1.416	.157
Total Visit Duration Graph Title1_Mean	-.439	.660	-.450	.653	-1.220	.222	-.839	.401
Total Visit Duration Graph Title2 N	-1.249	.212	-.411	.681	-1.305	.192	-1.778	.075
Total Visit Duration Graph Title2 Mean	-1.592	.111	-1.424	.154	-.440	.660	-.961	.337
Total Visit Duration Letter-to-Editor N	-.401	.688	-.637	.524	-.695	.487	-.947	.343
Total Visit Duration Letter-to-Editor Mean	-.073	.942	-.937	.349	-.885	.376	-1.295	.195
Total Visit Duration Line Graph N	-1.051	.293	0.000	1.000	0.000	1.000	0.000	1.000
Total Visit Duration Line Graph Mean	-1.054	.292	-1.124	.261	-1.957	.050	-1.948	.051
Total Visit Duration Newspaper Article N	-.655	.513	-.624	.532	-.755	.450	-.655	.513
Total Visit Duration Newspaper Article Mean	-.606	.544	-.593	.553	-.232	.817	-1.134	.257
Total Visit Duration Paragraph1 N	-1.305	.192	-.593	.553	0.000	1.000	0.000	1.000
Total Visit Duration Paragraph1 Mean	-2.014	.044	-.649	.516	-.454	.650	-.162	.871
Total Visit Duration Paragraph2 N	-1.249	.212	-.411	.681	-.062	.950	-.593	.553
Total Visit Duration Paragraph2 Mean	-.426	.670	-.163	.871	-.425	.671	-.095	.925
Total Visit Duration Paragraph3 N	-1.249	.212	-.411	.681	-1.237	.216	-.472	.637
Total Visit Duration Paragraph3 Mean	-.128	.898	-.271	.786	-1.476	.140	-.812	.417
Total Visit Duration Paragraph4 N	-.378	.705	0.000	1.000	-.247	.805	-.472	.637
Total Visit Duration Paragraph4 Mean	-.857	.391	-.913	.361	-1.362	.173	-.758	.448

Total Visit Duration Pie Chart N	-1.051	.293	0.000	1.000	0.000	1.000	0.000	1.000
Total Visit Duration Pie Chart Mean	-1.361	.173	-.122	.903	-1.361	.173	-1.325	.185
Total Visit Duration Sarah Case N	-.033	.974	-.947	.343	-.618	.537	0.000	1.000
Total Visit Duration Sarah Case Mean	-.788	.431	-.992	.321	-.599	.549	-.111	.912
Total Visit Duration Task Instruction N	-1.051	.293	0.000	1.000	-.734	.463	-1.000	.317
Total Visit Duration Task Instruction Mean	-.995	.320	-1.911	.056	-.085	.932	-.460	.646
Visit Count Essay N	-.062	.950	-.593	.553	-1.051	.293	-1.433	.152
Visit Count Essay Mean	-1.476	.140	-1.137	.256	-2.143	.032	-2.436	.015
Visit Count Graph Title1 N	-.436	.663	-1.041	.298	-1.732	.083	-1.416	.157
Visit Count Graph Title1 Mean	-.725	.469	-.298	.765	-1.168	.243	-.842	.400
Visit Count Graph Title2 N	-1.249	.212	-.411	.681	-1.305	.192	-1.778	.075
Visit Count Graph Title2 Mean	-2.008	.045	-1.467	.142	-1.309	.190	-1.032	.302
Visit Count Letter-to-Editor N	-.401	.688	-.637	.524	-.695	.487	-.947	.343
Visit Count Letter-to-Editor Mean	-.561	.575	-.539	.590	-1.226	.220	-.998	.318
Visit Count Line Graph N	-1.051	.293	0.000	1.000	0.000	1.000	0.000	1.000
Visit Count Line Graph Mean	-1.194	.233	-.976	.329	-2.072	.038	-1.638	.101
Visit Count Newspaper Article N	-.655	.513	-.624	.532	-.755	.450	-.655	.513
Visit Count Newspaper Article Mean	-.823	.411	-.567	.571	-.437	.662	-.779	.436
Visit Count Paragraph1 N	-1.305	.192	-.593	.553	0.000	1.000	0.000	1.000
Visit Count Paragraph1 Mean	-1.589	.112	-.460	.645	-1.234	.217	-.311	.756

Visit Count Paragraph2 N	-1.249	.212	-.411	.681	-.062	.950	-.593	.553
Visit Count Paragraph2 Mean	-.754	.451	-.027	.978	-.428	.669	-.082	.935
Visit Count Paragraph3 N	-1.249	.212	-.411	.681	-1.237	.216	-.472	.637
Visit Count Paragraph3 Mean	-.242	.809	-.109	.913	-.498	.619	-.922	.356
Visit Count Paragraph4 N	-.378	.705	0.000	1.000	-.247	.805	-.472	.637
Visit Count Paragraph4 Mean	-.415	.678	-.478	.632	-1.094	.274	-1.125	.260
Visit Count Pie Chart N	-1.051	.293	0.000	1.000	0.000	1.000	0.000	1.000
Visit Count Pie Chart Mean	-1.248	.212	-.379	.705	-1.264	.206	-1.327	.184
Visit Count Sarah Case N	-.033	.974	-.947	.343	-.618	.537	0.000	1.000
Visit Count Sarah Case Mean	-.479	.632	-.975	.329	-.486	.627	-.016	.987
Visit Count Task Instruction N	-1.051	.293	0.000	1.000	-.734	.463	-1.000	.317
Visit Count Task Instruction Mean	-.043	.966	-1.177	.239	-.625	.532	-.095	.924
Rate of Total Fixation Duration Essay	-.312	.755	-.514	.607	-2.581	.010	-2.489	.013
Rate of Total Fixation Duration Graph Title2	-.184	.854	-.771	.441	-1.079	.281	-.785	.432
Rate of Total Fixation Duration Graph Title	-2.161	.031	-1.329	.184	-.284	.777	-1.082	.279
Rate of Total Fixation Duration Letter-to-Editor	-.161	.872	-.853	.394	-.974	.330	-1.436	.151
Rate of Total Fixation Duration Bar Graph	-1.446	.148	-1.474	.140	-1.475	.140	-1.758	.079
Rate of Total Fixation Duration Newspaper Article	-.515	.606	-.333	.740	-.464	.643	-1.216	.224
Rate of Total Fixation Duration Paragraph1	-1.390	.165	-.433	.665	-.284	.777	-.730	.465
Rate of Total Fixation Duration Paragraph2	-.625	.532	-.054	.957	-.908	.364	-.162	.871

Rate of Total Fixation Duration Paragraph3	-0.085	.932	-.759	.448	-1.675	.094	-1.164	.244
Rate of Total Fixation Duration Paragraph4	-.800	.424	-.668	.504	-1.306	.192	-.921	.357
Rate of Total Fixation Duration Pie Chart	-.397	.691	-.149	.882	-.936	.349	-1.596	.110
Rate of Total Fixation Duration Sarah Case	-.945	.345	-1.112	.266	-.599	.549	-.016	.987
Rate of Total Fixation Duration Task Instruction	-1.106	.269	-2.151	.032	-.681	.496	-1.271	.204

Notes: P-values are exact (2-tailed). P-values below .05 are highlighted in **bold**. N=40 in all comparisons.

Appendix Q Mann-Whitney tests of difference between participants with lower and upper intermediate skills in reading and writing (Last 10 minutes)

Dependent variable	Task 1				Task 2			
	Aptis Reading		Aptis Writing		Aptis Reading		Aptis Writing	
	Z	p-value	Z	p-value	Z	p-value	Z	p-value
Fixation Duration Essay Mean	-0.796	.426	-0.054	.957	-0.270	.787	-0.163	.871
Fixation Duration Graph Title1 Mean	-0.162	.871	-0.746	.455	-0.989	.323	-0.533	.594
Fixation Duration Graph Title2 Mean	-1.600	.110	-0.392	.695	-0.279	.781	-0.965	.335
Fixation Duration Letter-to-Editor Mean	-0.448	.655	-0.854	.393	-0.321	.748	-0.518	.605
Fixation Duration Line Graph Mean	-0.498	.618	-0.548	.583	-0.768	.442	-0.774	.439
Fixation Duration Newspaper Article Mean	-1.220	.222	-2.859	.004	-0.557	.578	-0.347	.728
Fixation Duration Paragraph1 Mean	-0.367	.714	-1.649	.099	-0.611	.541	-0.732	.464
Fixation Duration Paragraph2 Mean	-1.043	.297	-1.522	.128	-0.838	.402	-0.745	.456
Fixation Duration Paragraph3 Mean	-1.057	.291	-1.362	.173	-0.029	.977	-0.920	.358
Fixation Duration Paragraph4 Mean	0.000	1.000	-0.783	.434	-1.038	.299	-0.624	.533
Fixation Duration Pie Chart Mean	-0.484	.628	-0.244	.807	-1.109	.268	-0.488	.625
Fixation Duration Sarah Case Mean	-0.157	.875	-1.101	.271	-0.187	.851	-1.121	.262
Fixation Duration Task Instruction Mean	-0.561	.575	-0.439	.661	-0.144	.885	-0.730	.465
Total Fixation Duration Essay Mean	-0.242	.809	-0.365	.715	-1.503	.133	-1.542	.123
Total Fixation Duration Graph Title1 Mean	-0.133	.894	-0.267	.789	-0.602	.547	-0.929	.353
Total Fixation Duration Graph Title2 Mean	-1.246	.213	-0.294	.769	-0.088	.930	-0.322	.748
Total Fixation Duration Letter-to-Editor Mean	-0.448	.655	-0.889	.374	-0.395	.693	-0.518	.605

Total Fixation Duration Line Graph Mean	-1.688	.091	-.650	.516	-.312	.755	-.420	.674
Total Fixation Duration Newspaper Article Mean	-1.113	.266	-2.859	.004	-.557	.578	-.306	.759
Total Fixation Duration Paragraph1 Mean	-.568	.570	-.514	.607	-.709	.478	-.703	.482
Total Fixation Duration Paragraph2 Mean	-.823	.410	-.162	.871	-.099	.921	-.392	.695
Total Fixation Duration Paragraph3 Mean	-.884	.377	-1.156	.248	-.302	.763	-.603	.546
Total Fixation Duration Paragraph4 Mean	-1.193	.233	-1.467	.142	-.341	.733	-.135	.892
Total Fixation Duration Pie Chart Mean	-.085	.932	-.325	.745	-.581	.561	-.879	.379
Total Fixation Duration Sarah Case Mean	-.210	.834	-1.034	.301	-.511	.609	-1.008	.314
Fixation Count Essay N	-1.181	.238	-.593	.553	-.723	1.000	0.000	1.000
Fixation Count Essay Mean	-.170	.865	-1.136	.256	-.581	.072	-1.745	.081
Fixation Count Graph Title2 N	-.655	.513	-1.041	.298	-1.148	.912	-1.049	.294
Fixation Count Graph Title2 Mean	-.085	.932	-.312	.755	-1.900	.730	-1.083	.279
Fixation Count Graph Title N	-.388	.698	-.411	.681	-.058	.688	0.000	1.000
Fixation Count Graph Title Mean	-.924	.355	-1.234	.217	-1.132	.848	0.000	1.000
Fixation Count Letter-to-Editor N	-.267	.789	-.637	.524	-.544	.805	-.472	.637
Fixation Count Letter-to-Editor Mean	-.029	.977	-.840	.401	-.575	.693	-.447	.655
Fixation Count Bar Graph N	-.707	.480	-.975	.330	-.674	.313	-.874	.382
Fixation Count Bar Graph Mean	-1.238	.216	-1.843	.065	-.147	.831	-.271	.786
Fixation Count Newspaper Article N	0.000	1.000	-1.249	.212	-.395	.636	-.411	.681
Fixation Count Newspaper Article Mean	-.410	.681	-.972	.331	-.345	.534	-.327	.744

Fixation Count Paragraph1 N	-.062	.950	-.593	.553	-.395	.173	-1.000	.317
Fixation Count Paragraph1 Mean	-1.262	.207	-.744	.457	-.313	.478	-.717	.473
Fixation Count Paragraph2 N	-.388	.698	-.411	.681	-.128	.192	-.593	.553
Fixation Count Paragraph2 Mean	-.776	.438	-1.042	.298	-.128	.932	-.501	.616
Fixation Count Paragraph3 N	-.388	.698	-.411	.681	-.644	.391	-1.363	.173
Fixation Count Paragraph3 Mean	-.157	.876	-.027	.978	-.557	.829	-.617	.537
Fixation Count Paragraph4 N	-1.134	.257	0.000	1.000	-.557	.805	-1.416	.157
Fixation Count Paragraph4 Mean	-1.415	.157	-.532	.595	-.085	.843	-.190	.850
Fixation Count Pie Chart N	-.451	.652	0.000	1.000	-.808	.173	-1.000	.317
Fixation Count Pie Chart Mean	-.652	.514	-.284	.776	-.284	.619	-1.029	.304
Fixation Count Sarah Case N	-.033	.974	-.947	.343	-.227	.753	-1.000	.317
Fixation Count Sarah Case Mean	-.634	.526	-1.118	.263	-.383	.452	-.815	.415
Fixation Count Task Instruction N	-.451	.652	0.000	1.000	-.085	.701	-.333	.739
Fixation Count Task Instruction Mean	-.766	.444	-.744	.457	-.303	.470	-.359	.720
Visit Duration Essay N	-.582	.561	-.907	.364	-.115	.561	-.014	.989
Visit Duration Essay Mean	-1.390	.165	-.460	.646	-.273	.251	-.419	.675
Visit Duration Essay Sum	-.492	.623	-1.475	.140	-.171	.057	-1.515	.130
Visit Duration GraphTitle2 N	-.085	.932	-.380	.704	-1.405	.954	-1.392	.164
Visit Duration GraphTitle2 Mean	-1.291	.197	-.596	.551	-.738	.258	-.697	.486
Visit Duration GraphTitle2 Sum	-.142	.887	-.257	.797	-.057	.586	-1.039	.299

Visit Duration Graph Title N	-1.339	.181	-1.644	.100	-1.829	.565	-.014	.989
Visit Duration Graph Title Mean	-.313	.754	-.176	.860	-.950	.500	-.615	.538
Visit Duration Graph Title Sum	-.910	.363	-1.397	.162	-.393	.883	-.294	.769
Visit Duration Letter-to-Editor N	-.324	.746	-.703	.482	-.460	.693	-.447	.655
Visit Duration Letter-to-Editor Mean	-.777	.437	-.755	.450	-.511	.730	-.565	.572
Visit Duration Letter-to-Editor Sum	-.337	.736	-.895	.371	-1.012	.693	-.518	.605
Visit Duration Bar Graph N	-.668	.504	-1.301	.193	-.606	.754	-.014	.989
Visit Duration Bar Graph Mean	-.440	.660	-.271	.787	-.808	.898	-.203	.839
Visit Duration Bar Graph Sum	-.440	.660	-1.190	.234	0.000	.898	-.190	.850
Visit Duration Newspaper Article N	-.411	.681	-.959	.338	-.581	.520	-.389	.697
Visit Duration Newspaper Article Mean	0.000	1.000	-.984	.325	-.110	.578	-.306	.759
Visit Duration Newspaper Article Sum	-.212	.832	-.969	.333	-.058	.578	-.306	.759
Visit Duration Paragraph1 N	-1.178	.239	-.773	.439	-.401	.932	-.068	.946
Visit Duration Paragraph1 Mean	-.142	.887	-.135	.892	-.575	.419	-.528	.598
Visit Duration Paragraph1 Sum	-1.163	.245	-.325	.745	-.247	.777	-.108	.914
Visit Duration Paragraph2 N	-.327	.743	-.543	.587	-.395	.820	-.759	.448
Visit Duration Paragraph2 Mean	-.512	.609	-.651	.515	-1.008	.702	-.027	.978
Visit Duration Paragraph2 Sum	-.142	.887	-.353	.724	-.313	.932	-.541	.588
Visit Duration Paragraph3 N	-.271	.787	-.082	.935	-.474	.762	-.220	.826
Visit Duration Paragraph3 Mean	-.327	.744	-.108	.914	-.644	.908	-.809	.419

Visit Duration Paragraph3 Sum	-469	.639	-.325	.745	-1.363	.785	-.576	.565
Visit Duration Paragraph4 N	-960	.337	-.137	.891	-.085	.865	-.149	.881
Visit Duration Paragraph4 Mean	-1.529	.126	-.709	.478	-1.305	.160	-.514	.607
Visit Duration Paragraph4 Sum	-1.700	.089	-1.022	.307	-.227	.461	-.135	.892
Visit Duration Pie Chart N	-.808	.419	-.433	.665	-.857	.955	-1.477	.140
Visit Duration Pie Chart Mean	-.264	.792	-1.279	.201	-.303	.067	-.162	.871
Visit Duration Pie Chart Sum	-.615	.539	-.562	.574	-.247	.342	-.717	.473
Visit Duration Sarah Case N	-.479	.632	-.975	.329	-.171	.694	-1.191	.234
Visit Duration Sarah Case Mean	-.599	.549	-.917	.359	-1.363	.646	-.764	.445
Visit Duration Sarah Case Sum	-.788	.431	-.992	.321	-.057	.609	-1.008	.314
Visit Duration Task Instruction N	-.497	.620	-1.069	.285	-.314	.312	-.579	.562
Visit Duration Task Instruction Mean	-1.106	.269	-.744	.457	-.393	.545	-.702	.483
Visit Duration Task Instruction Sum	-1.234	.217	-1.258	.208	-.384	.419	-.564	.573
Visit Count Essay N	-1.181	.238	-.593	.553	-1.012	1.000	0.000	1.000
Visit Count Essay Mean	-.582	.561	-.907	.364	-1.872	.561	-.014	.989
Visit Count Graph Title2 N	0.000	1.000	-.639	.523	-.573	.912	-1.049	.294
Visit Count Graph Title2 Mean	-.085	.932	-.380	.704	-.190	.954	-1.392	.164
Visit Count Graph Title N	-.388	.698	-.411	.681	-.296	.688	0.000	1.000
Visit Count Graph Title Mean	-1.339	.181	-1.644	.100	-.241	.565	-.014	.989
Visit Count Letter-to-Editor N	-.267	.789	-.637	.524	-.557	.805	-.472	.637

Visit Count Letter-to-Editor Mean	-.324	.746	-.703	.482	-.198	.693	-.447	.655
Visit Count Bar Graph N	-.451	.652	0.000	1.000	-.567	.313	-.874	.382
Visit Count Bar Graph Mean	-.668	.504	-1.301	.193	-.331	.754	-.014	.989
Visit Count Newspaper Article N	0.000	1.000	-1.249	.212	-.624	.636	-.411	.681
Visit Count Newspaper Article Mean	-.411	.681	-.959	.338	-1.276	.520	-.389	.697
Visit Count Paragraph1 N	-.062	.950	-.593	.553	-.511	.173	-1.000	.317
Visit Count Paragraph1 Mean	-.780	.435	-.501	.617	-.404	.932	-.068	.946
Visit Count Paragraph2 N	-.388	.698	-.411	.681	-.723	.192	-.593	.553
Visit Count Paragraph2 Mean	-.327	.743	-.543	.587	-.581	.820	-.759	.448
Visit Count Paragraph3 N	-.388	.698	-.411	.681	-1.148	.391	-1.363	.173
Visit Count Paragraph3 Mean	-.271	.787	-.082	.935	-1.900	.762	-.220	.826
Visit Count Paragraph4 N	-1.134	.257	0.000	1.000	-.058	.805	-1.416	.157
Visit Count Paragraph4 Mean	-.960	.337	-.137	.891	-1.132	.865	-.149	.881
Visit Count Pie Chart N	-.451	.652	0.000	1.000	-.544	.173	-1.000	.317
Visit Count Pie Chart Mean	-.808	.419	-.433	.665	-.575	.955	-1.477	.140
Visit Count Sarah Case N	-.033	.974	-.947	.343	-.674	.753	-1.000	.317
Visit Count Sarah Case Mean	-.479	.632	-.975	.329	-.147	.694	-1.191	.234
Visit Count Task Instruction N	-.451	.652	0.000	1.000	-.395	.701	-.333	.739
Visit Count Task Instruction Mean	-.497	.620	-1.069	.285	-.345	.312	-.579	.562
Rate of Total Fixation Duration Essay	-.298	.766	-.379	.705	-.395	.061	-1.001	.317

Rate of Total Fixation Duration GraphTitle2	-0.089	.929	-.199	.843	-.313	.567	-.574	.566
Rate of Total Fixation Duration Graph Title	-1.187	.235	-.741	.459	-.128	.849	-.503	.615
Rate of Total Fixation Duration Letter-to-Editor	-.485	.628	-.889	.374	-.128	.767	-.518	.605
Rate of Total Fixation Duration Bar Graph	-1.534	.125	-.772	.440	-.644	.809	-.840	.401
Rate of Total Fixation Duration Newspaper Article	-1.113	.266	-2.859	.004	-.557	.578	-.286	.775
Rate of Total Fixation Duration Paragraph1	-.852	.394	-.908	.364	-.557	.843	-.243	.808
Rate of Total Fixation Duration Paragraph2	-1.193	.233	-.555	.579	-.085	.571	-1.380	.168
Rate of Total Fixation Duration Paragraph3	-1.100	.271	-1.213	.225	-.808	.741	-.466	.641
Rate of Total Fixation Duration Paragraph4	-.158	.874	-.315	.753	-.284	.532	-.190	.850
Rate of Total Fixation Duration Pie Chart	-.170	.865	0.000	1.000	-.227	.202	-.081	.935
Rate of Total Fixation Duration Sarah Case	-.332	.740	-.918	.359	-.383	.609	-.975	.330
Rate of Total Fixation Duration Task Instruction	-.101	.920	-.411	.681	-.085	.686	-.193	.847

Notes: P-values are exact (2-tailed). P-values below .05 are highlighted in **bold**. N=40 in all comparisons. N

Appendix R Regression results for eye-movement variables in 40 minutes

Dependent variable	Task 1					Task 2				
	Reading predictor		Writing predictor		R-squared	Reading predictor		Writing predictor		R-squared
	Std beta	p-value	Std beta	p-value		Std beta	p-value	Std beta	p-value	
Time to First Fixation Essay Mean	.161	.418	-.131	.509	.019	.126	.526	-.017	.930	.014
Time to First Fixation GraphTitle1_Mean	.003	.988	-.101	.612	.010	-.277	.133	-.200	.275	.179
Time to First Fixation GraphTitle2_Mean	.161	.415	.022	.909	.031	-.331	.067	-.192	.282	.218
Time to First Fixation Letter-to-Editor Mean	-.166	.404	.094	.636	.019	-.077	.645	-.507	.004	.306
Time to First Fixation Line Graph Mean	-.171	.390	.121	.541	.021	-.242	.221	.207	.293	.045
Time to First Fixation Newspaper Article Mean	-.343	.076	.347	.073	.104	.064	.728	-.432	.023	.160
Time to First Fixation Paragraph1 Mean	.214	.281	-.078	.691	.033	.172	.321	-.584	.002	.258
Time to First Fixation Paragraph2 Mean	-.063	.734	-.319	.094	.129	-.034	.859	-.246	.207	.071
Time to First Fixation Paragraph3 Mean	.149	.453	-.049	.803	.016	-.012	.947	-.428	.022	.189
Time to First Fixation Paragraph4 Mean	.265	.180	-.161	.411	.048	-.004	.984	-.092	.644	.009
Time to First Fixation Pie Chart Mean	-.478	.013	.275	.141	.156	-.053	.776	-.310	.105	.117
Time to First Fixation Sarah Case Mean	.136	.488	.091	.642	.041	.230	.245	-.134	.495	.036
Time to First Fixation Task InstructionMean	-.075	.697	-.211	.279	.068	-.105	.598	.144	.470	.015
First Fixation Duration Essay Mean	.332	.085	-.367	.058	.108	-.043	.829	.013	.949	.001
First Fixation Duration GraphTitle1 Mean	.287	.145	-.102	.600	.060	.014	.943	-.060	.763	.003
First Fixation Duration GraphTitle2 Mean	.187	.336	-.307	.118	.065	-.197	.322	.150	.448	.028
First Fixation Duration Letter-to-Editor Mean	.165	.391	.166	.389	.085	-.432	.022	.446	.018	.169

First Fixation Duration Line Graph Mean	.142	.472	-.202	.309	.029	-.159	.414	-.126	.515	.064
First Fixation Duration Newspaper Article Mean	.235	.233	-.210	.287	.044	-.382	.050	.240	.211	.101
First Fixation Duration Paragraph1 Mean	.103	.602	-.207	.297	.030	-.146	.455	.260	.188	.046
First Fixation Duration Paragraph2 Mean	.114	.566	-.021	.914	.011	-.024	.902	-.231	.237	.060
First Fixation Duration Paragraph3 Mean	.133	.503	-.029	.884	.014	-.001	.997	-.223	.257	.050
First Fixation Duration Paragraph4 Mean	.291	.147	-.212	.286	.060	.056	.777	.141	.474	.032
First Fixation Duration Pie Chart Mean	.181	.362	-.059	.765	.024	.181	.362	-.059	.765	.024
First Fixation Duration Sarah Case Mean	.277	.157	-.033	.866	.068	.060	.763	-.050	.803	.003
First Fixation Duration Task Instruction Mean	.173	.366	-.379	.052	.100	.328	.093	-.074	.701	.086
Fixation Duration Essay Mean	.220	.260	-.290	.141	.061	.057	.776	.029	.884	.006
Fixation Duration Graph Title1 Mean	.477	.012	-.410	.029	.176	-.041	.834	-.161	.414	.035
Fixation Duration Graph Title2 Mean	.195	.327	-.080	.684	.027	-.150	.449	.177	.372	.024
Fixation Duration Letter-to-Editor Mean	.153	.428	.149	.440	.072	-.439	.022	.390	.040	.153
Fixation Duration Line Graph Mean	.287	.143	-.046	.810	.069	-.117	.554	-.080	.684	.031
Fixation Duration Newspaper Article Mean	.227	.249	-.223	.258	.044	-.333	.089	.236	.224	.079
Fixation Duration Paragraph1 Mean	.255	.195	-.054	.781	.053	-.245	.215	.203	.302	.045
Fixation Duration Paragraph2 Mean	.266	.172	.006	.974	.073	-.077	.700	.077	.698	.005
Fixation Duration Paragraph3 Mean	.268	.171	-.009	.965	.069	-.079	.692	.083	.677	.006
Fixation Duration Paragraph4 Mean	.351	.072	-.132	.492	.089	-.065	.743	.131	.510	.012
Fixation Duration Pie Chart Mean	.299	.129	-.207	.288	.063	.072	.714	.118	.550	.029

Fixation Duration Sarah Case Mean	.376	.051	-.077	.685	.115	.080	.687	-.006	.974	.006
Fixation Duration Task Instruction Mean	.265	.179	-.215	.274	.052	.153	.439	.012	.951	.026
Total Fixation Duration Essay Mean	-.100	.599	.361	.063	.100	.145	.435	.287	.126	.150
Total Fixation Duration Graph Title1 Mean	.171	.390	-.044	.823	.023	-.127	.515	.288	.143	.058
Total Fixation Duration Graph Title2 Mean	.102	.610	-.059	.769	.007	.119	.548	.077	.695	.030
Total Fixation Duration Letter-to-Editor Mean	.096	.628	.047	.813	.017	-.021	.917	-.060	.765	.005
Total Fixation Duration Line Graph Mean	.096	.629	.083	.676	.025	.028	.887	.189	.336	.043
Total Fixation Duration Newspaper Article Mean	.204	.304	-.168	.396	.031	.074	.710	-.062	.757	.004
Total Fixation Duration Paragraph1 Mean	.313	.107	-.035	.855	.087	-.127	.524	.159	.423	.019
Total Fixation Duration Paragraph2 Mean	.057	.773	.136	.492	.030	-.302	.124	.246	.208	.068
Total Fixation Duration Paragraph3 Mean	.064	.742	.158	.422	.041	.045	.823	.065	.742	.010
Total Fixation Duration Paragraph4 Mean	.119	.546	.103	.599	.039	-.042	.834	.078	.694	.004
Total Fixation Duration Pie Chart Mean	.318	.106	-.172	.376	.069	.038	.841	.329	.085	.123
Total Fixation Duration Sarah Case Mean	.246	.212	-.051	.792	.049	-.076	.704	.140	.483	.013
Total Fixation Duration Task Instruction Mean	.274	.164	-.088	.650	.056	-.041	.837	.017	.934	.001
Fixation Count Essay Mean	-.053	.781	.317	.103	.085	.197	.489	.225	.230	.139
Fixation Count Graph Title1 Mean	.167	.398	.005	.979	.029	.138	.615	-.122	.541	.015
Fixation Count Graph Title2-Mean	-.097	.623	.223	.261	.035	.100	.529	.022	.911	.013
Fixation Count Letter-to-Editor-Mean	.008	.970	-.084	.675	.006	.125	.168	.003	.986	.016
Fixation Count Line Graph Mean	.070	.723	.113	.567	.027	.276	.083	-.253	.205	.061

Fixation Count Newspaper Article Mean	.111	.578	-.071	.723	.009	.337	.172	-.069	.716	.092
Fixation Count Paragraph1 Mean	-.031	.877	.128	.519	.013	.266	.702	.006	.974	.073
Fixation Count Paragraph2 Mean	-.241	.223	.216	.272	.046	.076	.506	.116	.558	.029
Fixation Count Paragraph3 Mean	.071	.722	.060	.761	.013	.131	.092	.095	.628	.040
Fixation Count Paragraph4 Mean	-.002	.991	.037	.851	.001	.330	.216	-.238	.219	.077
Fixation Count Pie Chart Mean	.054	.775	.270	.163	.092	.243	.161	-.005	.980	.058
Fixation Count Sarah Case Mean	.077	.695	.152	.440	.042	.276	.489	-.113	.563	.054
Fixation Count Task Instruction Mean	-.049	.807	.016	.937	.002	.197	.615	.225	.230	.139
Visit Duration Essay Mean	.038	.849	-.103	.605	.008	.035	.841	-.517	.005	.248
Visit Duration Graph Title1 Mean	.269	.171	-.066	.734	.057	-.180	.366	.112	.573	.022
Visit Duration Graph Title2 Mean	.161	.419	-.104	.602	.018	-.344	.073	.383	.047	.117
Visit Duration Letter-to-Editor Mean	.134	.496	.074	.705	.035	-.257	.193	.198	.314	.048
Visit Duration Line Graph Mean	-.058	.747	-.413	.026	.200	.097	.625	.038	.850	.015
Visit Duration Newspaper Article Mean	.208	.317	-.093	.652	.029	-.203	.306	.073	.712	.030
Visit Duration Paragraph1 Mean	.127	.511	-.314	.110	.070	-.253	.200	.212	.282	.049
Visit Duration Paragraph2 Mean	-.084	.675	.046	.819	.005	-.279	.156	.097	.617	.057
Visit Duration Paragraph3 Mean	-.029	.884	.209	.291	.038	-.002	.993	.022	.912	.000
Visit Duration Paragraph4 Mean	.046	.818	-.034	.866	.002	-.195	.322	.235	.234	.042
Visit Duration Pie Chart Mean	.156	.433	-.172	.385	.024	-.162	.395	.383	.049	.103
Visit Duration Sarah Case Mean	.353	.078	-.010	.958	.121	.124	.530	.055	.780	.026

Visit Duration Task Instruction Mean	.063	.720	-.515	.005	.233	.140	.481	-.100	.614	.014
Total Visit Duration Essay Mean	.012	.950	.263	.177	.073	.156	.428	.072	.714	.042
Total Visit Duration GraphTitle1_Mean	.190	.339	-.071	.718	.026	.117	.553	.086	.663	.032
Total Visit Duration GraphTitle2_Mean	.125	.532	-.111	.578	.012	-.164	.403	.259	.189	.046
Total Visit Duration Letter-to-Editor Mean	.072	.718	.052	.795	.012	.019	.925	-.092	.645	.007
Total Visit Duration Line Graph Mean	-.039	.844	-.009	.966	.002	.138	.488	-.005	.980	.018
Total Visit Duration Newspaper Article Mean	.186	.349	-.146	.463	.025	.073	.713	-.066	.740	.004
Total Visit Duration Paragraph1_Mean	.307	.117	-.099	.609	.070	-.175	.376	.180	.363	.028
Total Visit Duration Paragraph2_Mean	.046	.817	.019	.923	.004	-.311	.113	.180	.354	.066
Total Visit Duration Paragraph3 Mean	.072	.715	.086	.666	.020	.021	.916	.063	.753	.006
Total Visit Duration Paragraph4 Mean	.113	.568	.042	.830	.020	-.055	.783	.010	.960	.003
Total Visit Duration Pie Chart Mean	.346	.074	-.344	.075	.105	-.056	.769	.350	.070	.104
Total Visit Duration Sarah Case Mean	.310	.112	-.065	.734	.078	-.013	.948	.132	.507	.016
Total Visit Duration Task Instruction Mean	.298	.119	-.404	.037	.117	.049	.806	-.072	.718	.004
Visit Count Essay Mean	.358	.060	.019	.919	.136	.273	.141	.180	.327	.163
Visit Count GraphTitle1 Mean	.219	.270	-.167	.398	.035	.241	.213	.079	.681	.085
Visit Count GraphTitle2 Mean	.224	.257	-.194	.325	.039	.044	.826	.072	.717	.011
Visit Count Letter-to-Editor Mean	.130	.512	.020	.920	.020	-.062	.757	-.057	.775	.011
Visit Count Line Graph Mean	.179	.367	-.042	.833	.025	.194	.327	-.020	.919	.034
Visit Count Newspaper Article Mean	.283	.144	-.354	.069	.093	.118	.555	-.047	.814	.010

Visit Count Paragraph1 Mean	.351	.072	-.126	.510	.089	.065	.739	.155	.432	.039
Visit Count Paragraph2 Mean	.193	.329	.000	.999	.037	-.161	.417	.182	.359	.026
Visit Count Paragraph3 Mean	.228	.242	.056	.772	.069	.040	.839	.096	.629	.015
Visit Count Paragraph4 Mean	.207	.289	.054	.782	.058	.018	.929	.023	.909	.001
Visit Count Pie Chart Mean	.318	.106	-.203	.296	.070	.033	.865	.236	.226	.066
Visit Count Sarah Case Mean	.255	.193	-.018	.927	.060	.056	.778	.118	.552	.024
Visit Count Task Instruction Mean	.268	.174	-.196	.318	.051	.039	.844	.065	.744	.009
Rate of Total Fixation Duration Essay	-.044	.825	.097	.626	.007	.235	.222	.106	.577	.094
Rate of Total Fixation Duration Graph Title1	.287	.132	-.413	.033	.120	.043	.827	.117	.554	.021
Rate of Total Fixation Duration Graph Title2	.172	.383	-.223	.261	.036	-.102	.608	.130	.514	.012
Rate of Total Fixation Duration Letter-to-Editor	.084	.675	-.040	.841	.005	-.008	.967	-.390	.039	.155
Rate of Total Fixation Duration Line Graph	-.019	.926	-.093	.641	.011	.037	.851	.026	.896	.003
Rate of Total Fixation Duration Newspaper Article	.196	.289	-.473	.014	.158	-.056	.780	-.027	.891	.006
Rate of Total Fixation Duration Paragraph1	.323	.097	-.294	.131	.084	-.265	.179	.084	.668	.052
Rate of Total Fixation Duration Paragraph2	.050	.804	.039	.845	.006	-.354	.064	.004	.982	.124
Rate of Total Fixation Duration Paragraph4	.166	.403	-.149	.452	.022	.006	.975	-.407	.031	.163
Rate of Total Fixation Duration Pie Chart	.234	.216	-.428	.027	.126	-.008	.968	-.221	.261	.051
Rate of Total Fixation Duration Sarah Case	.164	.408	-.212	.286	.033	.046	.815	.184	.350	.045
Rate of Total Fixation Duration Task Instruction	.338	.084	-.135	.482	.081	-.020	.920	.017	.931	.000

Each row contains the results from the two separate regressions, with results first for Task 1 and then for Task 2. P-values less than .05 highlighted in bold. N is 38

Appendix S Regression results for eye-movement variables in the first 10 minutes

Dependent variable	Task 1					Task 2				
	Reading predictor		Writing predictor		R-squared	Reading predictor		Writing predictor		R-squared
	Std beta	p-value	Std beta	p-value		Std beta	p-value	Std beta	p-value	
Time to First Fixation Essay Mean	.363	.057	-.415	.031	.135	-.041	.807	-.515	.004	.291
Time to First Fixation Graph Title1 Mean	.148	.450	.102	.602	.049	.032	.872	.143	.469	.027
Time to First Fixation Graph Title Mean	-.122	.541	-.019	.924	.018	.335	.088	-.223	.249	.078
Time to First Fixation Letter-to-Editor Mean	.028	.889	.057	.775	.006	.095	.623	-.279	.156	.057
Time to First Fixation Line Graph Mean	.037	.854	.012	.951	.002	.384	.049	-.231	.227	.101
Time to First Fixation Newspaper Article Mean	-.024	.895	-.399	.034	.171	.017	.930	.115	.564	.016
Time to First Fixation Paragraph1 Mean	.216	.269	.035	.855	.057	-.199	.310	-.064	.742	.058
Time to First Fixation Paragraph2 Mean	.298	.131	-.157	.419	.061	.022	.912	.003	.986	.001
Time to First Fixation Paragraph3 Mean	.172	.384	-.212	.284	.034	.241	.210	.107	.574	.098
Time to First Fixation Paragraph4 Mean	.089	.652	.072	.715	.020	.077	.697	.088	.656	.021
Time to First Fixation Pie Chart Mean	.171	.384	.059	.764	.044	-.365	.058	.045	.811	.117
Time to First Fixation Sarah Case Mean	.138	.485	-.184	.354	.024	-.041	.839	.068	.732	.003
Time to First Fixation Task Instruction Mean	-.086	.664	.181	.363	.023	-.092	.641	.214	.280	.032
Fixation Duration Essay Mean	.192	.307	-.425	.028	.126	.195	.292	.256	.168	.159

Fixation Duration Graph Title1 Mean	.317	.098	-.402	.038	.119	.268	.135	.265	.139	.222
Fixation Duration Graph Title2 Mean	.293	.137	-.175	.369	.059	-.032	.845	.604	.001	.345
Fixation Duration Letter-to-Editor Mean	-.064	.746	.200	.312	.030	-.153	.439	.177	.374	.024
Fixation Duration Line Graph Mean	.190	.339	-.058	.767	.027	-.113	.567	.189	.341	.025
Fixation Duration Newspaper Article Mean	.198	.318	-.102	.607	.027	-.126	.528	.021	.915	.013
Fixation Duration Paragraph1 Mean	.218	.263	.065	.738	.068	-.282	.152	.213	.277	.058
Fixation Duration Paragraph2 Mean	.279	.155	-.054	.779	.064	.048	.803	.219	.263	.062
Fixation Duration Paragraph3 Mean	.107	.583	.172	.377	.062	-.153	.440	-.034	.861	.030
Fixation Duration Paragraph4 Mean	.023	.910	.015	.941	.001	-.066	.729	.325	.095	.086
Fixation Duration Pie Chart Mean	.388	.050	-.219	.262	.102	.367	.054	.010	.956	.139
Fixation Duration Sarah Case Mean	.153	.427	.181	.346	.087	-.106	.597	.099	.621	.009
Fixation Duration Task Instruction Mean	.239	.226	-.201	.307	.044	.330	.089	-.050	.794	.093
Total Fixation Duration Essay Mean	.254	.182	.131	.488	.119	.371	.046	.091	.616	.184
Total Fixation Duration Graph Title1 Mean	.075	.707	-.018	.930	.004	.033	.865	.257	.187	.076
Total Fixation Duration Graph Title2 Mean	.087	.661	.061	.757	.017	.124	.533	-.046	.818	.011
Total Fixation Duration Letter-to-Editor Mean	-.042	.835	.071	.721	.003	-.071	.723	.096	.632	.007
Total Fixation Duration Line Graph Mean	.083	.668	.199	.307	.065	.276	.139	.167	.367	.156
Total Fixation Duration Newspaper Article Mean	.252	.203	-.122	.535	.044	-.039	.846	-.042	.835	.005
Total Fixation Duration Paragraph1 Mean	.363	.060	-.055	.772	.112	-.114	.567	.129	.517	.013
Total Fixation Duration Paragraph2 Mean	.111	.575	.059	.767	.023	-.200	.301	.336	.086	.077

Total Fixation Duration Paragraph3 Mean	.074	.706	.121	.540	.030	-.184	.349	.244	.217	.043
Total Fixation Duration Paragraph4 Mean	-.104	.602	.135	.498	.013	-.155	.435	.168	.398	.023
Total Fixation Duration Pie Chart Mean	.250	.205	-.189	.336	.045	.152	.438	.109	.574	.054
Total Fixation Duration Sarah Case Mean	.192	.309	.207	.272	.124	-.186	.349	.114	.565	.024
Total Fixation Duration Task Instruction Mean	.232	.239	-.030	.878	.047	-.046	.819	.027	.892	.001
Fixation Count Essay Mean	.226	.235	.156	.408	.115	.378	.040	.109	.543	.201
Fixation Count Graph Title1_Mean	.091	.650	-.050	.801	.006	.155	.437	-.121	.543	.018
Fixation Count Graph Title2_Mean	.098	.620	.070	.725	.022	.090	.643	.160	.415	.050
Fixation Count Letter-to-Editor Mean	.044	.826	.024	.903	.004	-.062	.758	.034	.863	.003
Fixation Count Line Graph Mean	.109	.575	.171	.381	.062	.307	.100	.131	.479	.157
Fixation Count Newspaper Article Mean	.303	.123	-.153	.432	.063	.030	.880	-.066	.740	.003
Fixation Count Paragraph1 Mean	.387	.044	-.053	.776	.130	.019	.926	.077	.701	.008
Fixation Count Paragraph2 Mean	.279	.155	-.054	.779	.064	-.204	.293	.329	.094	.074
Fixation Count Paragraph3 Mean	.073	.713	.117	.554	.028	-.194	.326	.225	.256	.039
Fixation Count Paragraph4 Mean	-.121	.544	.167	.402	.020	-.179	.366	.160	.421	.026
Fixation Count Pie Chart Mean	.251	.203	-.217	.270	.049	.158	.423	.064	.745	.040
Fixation Count Sarah Case Mean	.089	.640	.245	.203	.093	-.130	.513	.155	.435	.018
Fixation Count Task Instruction Mean	.237	.229	-.048	.807	.046	-.055	.783	.034	.866	.002
Visit Duration Essay Mean	-.252	.191	.357	.068	.090	-.046	.819	.029	.883	.001
Visit Duration Graph Title1 Mean	.253	.188	.100	.600	.102	.102	.608	-.018	.927	.009

Visit Duration Graph Title2 Mean	-.151	.395	.531	.005	.215	-.247	.204	.312	.111	.072
Visit Duration Letter-to-Editor Mean	-.111	.578	.036	.856	.009	.046	.796	-.488	.009	.215
Visit Duration Line Graph Mean	.184	.343	.100	.606	.065	.224	.256	-.029	.881	.044
Visit Duration Newspaper Article Mean	-.130	.513	-.012	.952	.019	.018	.928	-.219	.266	.044
Visit Duration Paragraph1 Mean	-.161	.418	.115	.563	.018	.065	.744	.061	.760	.012
Visit Duration Paragraph2 Mean	-.184	.348	.264	.181	.049	-.066	.735	.243	.218	.045
Visit Duration Paragraph3 Mean	-.112	.575	.033	.866	.009	-.322	.100	.231	.236	.074
Visit Duration Paragraph4 Mean	-.398	.039	.329	.086	.120	.296	.132	-.168	.388	.060
Visit Duration Pie Chart Mean	-.090	.653	.114	.568	.010	.165	.383	.208	.275	.109
Visit Duration Sarah Case Mean	-.241	.222	.052	.792	.047	.069	.694	-.520	.005	.235
Visit Duration Task Instruction Mean	.159	.424	-.057	.773	.018	-.046	.819	.029	.883	.001
Total Visit Duration Essay Mean	.152	.416	.261	.166	.136	.152	.194	.042	.827	.077
Total Visit Duration Graph Title1 Mean	.073	.728	-.015	.943	.004	.073	.436	-.086	.664	.016
Total Visit Duration Graph Title2 Mean	.070	.723	.072	.718	.016	.070	.882	.257	.187	.075
Total Visit Duration Letter-to-Editor Mean	-.066	.741	.084	.674	.005	-.066	.735	.052	.794	.003
Total Visit Duration Line Graph Mean	.070	.725	.087	.662	.019	.070	.053	.091	.615	.176
Total Visit Duration Newspaper Article Mean	.243	.220	-.109	.579	.041	.243	.835	-.039	.843	.005
Total Visit Duration Paragraph1 Mean	.378	.052	-.171	.369	.100	.378	.481	.115	.562	.015
Total Visit Duration Paragraph2 Mean	.124	.534	-.041	.836	.011	.124	.245	.311	.113	.069
Total Visit Duration Paragraph3 Mean	.078	.695	.094	.634	.023	.078	.180	.246	.211	.057

Total Visit Duration Paragraph4 Mean	-.286	.145	.238	.224	.062	-.286	.239	.165	.402	.038
Total Visit Duration Pie Chart Mean	.291	.135	-.327	.094	.085	.291	.407	.044	.824	.037
Total Visit Duration Sarah Case Mean	.188	.318	.208	.270	.123	.188	.585	.102	.608	.010
Total Visit Duration Task Instruction Mean	.262	.168	-.413	.033	.118	.262	.647	-.06	.736	.006
Visit Count Essay Mean	.356	.065	-.031	.870	.115	.425	.020	.085	.629	.228
Visit Count Graph Title1 Mean	.148	.457	-.114	.567	.016	.184	.354	-.051	.797	.026
Visit Count Graph Title2 Mean	.165	.401	.067	.733	.044	.266	.173	-.014	.941	.067
Visit Count Letter-to-Editor Mean	.162	.435	-.054	.795	.019	-.289	.143	.193	.324	.058
Visit Count Line Graph Mean	.176	.366	.096	.621	.059	.357	.063	-.016	.933	.121
Visit Count Newspaper Article Mean	.276	.162	-.208	.289	.055	.084	.674	-.095	.633	.007
Visit Count Paragraph1 Mean	.466	.015	-.205	.271	.152	.134	.499	.046	.817	.027
Visit Count Paragraph2 Mean	.180	.365	-.110	.579	.022	.018	.929	.005	.979	.000
Visit Count Paragraph3 Mean	.275	.161	-.032	.870	.067	-.220	.268	.099	.616	.034
Visit Count Paragraph4 Mean	-.014	.944	.027	.894	.000	-.077	.697	-.054	.784	.014
Visit Count Pie Chart Mean	.249	.208	-.215	.275	.048	.249	.203	.002	.991	.063
Visit Count Sarah Case Mean	.053	.785	.229	.240	.069	.099	.616	.054	.787	.019
Visit Count Task Instruction Mean	.264	.181	-.141	.473	.048	.037	.854	.038	.849	.004
Visit Count Essay Mean	.356	.065	-.031	.870	.115	.425	.020	.085	.629	.228
Rate of Total Fixation Duration Essay	.356	.065	-.031	.870	.115	.425	.020	.085	.629	.228
Rate of Total Fixation Duration GraphTitle1	.148	.457	-.114	.567	.016	.184	.354	-.051	.797	.026

Rate of Total Fixation Duration Graph Title2	.165	.401	.067	.733	.044	.266	.173	-.014	.941	.067
Rate of Total Fixation Duration Letter-to-Editor	.162	.435	-.054	.795	.019	-.289	.143	.193	.324	.058
Rate of Total Fixation Duration Line Graph	.176	.366	.096	.621	.059	.357	.063	-.016	.933	.121
Rate of Total Fixation Duration Newspaper Article	.276	.162	-.208	.289	.055	.084	.674	-.095	.633	.007
Rate of Total Fixation Duration Paragraph1	.466	.015	-.205	.271	.152	.134	.499	.046	.817	.027
Rate of Total Fixation Duration Paragraph2	.180	.365	-.110	.579	.022	.018	.929	.005	.979	.000
Rate of Total Fixation Duration Paragraph3	.275	.161	-.032	.870	.067	-.220	.268	.099	.616	.034
Rate of Total Fixation Duration Paragraph4	-.014	.944	.027	.894	.000	-.077	.697	-.054	.784	.014
Rate of Total Fixation Duration Pie Chart	.249	.208	-.215	.275	.048	.249	.203	.002	.991	.063
Rate of Total Fixation Duration Sarah Case	.053	.785	.229	.240	.069	.099	.616	.054	.787	.019
Rate of Total Fixation Duration Task Instruction	.264	.181	-.141	.473	.048	.037	.854	.038	.849	.004

Each row contains the results from the two separate regressions, with results first for Task 1 and then for Task 2. P-values less than .05 highlighted in bold. N is 38

Appendix T Regression results for eye-movement variables in the last 10 minutes

Dependent variable	Task 1					Task 2				
	Reading predictor		Writing predictor		R-squared	Reading predictor		Writing predictor		R-squared
	Std beta	p-value	Std beta	p-value		Std beta	p-value	Std beta	p-value	
Fixation Duration Essay Mean	.095	.635	-.087	.664	.007	.010	.959	-.101	.611	.009
Fixation Duration Graph Title1 Mean	.189	.341	-.025	.901	.031	-.024	.904	-.039	.844	.003
Fixation Duration Graph Title2 Mean	-.199	.317	.133	.500	.028	.254	.193	-.299	.128	.069
Fixation Duration Letter-to-Editor Mean	.100	.611	.112	.571	.035	.137	.484	-.271	.169	.051
Fixation Duration Line Graph Mean	.038	.854	.151	.464	.031	.232	.240	-.218	.269	.045
Fixation Duration Newspaper Article Mean	-.010	.957	-.387	.041	.154	.058	.770	.090	.649	.017
Fixation Duration Paragraph1 Mean	.002	.992	.282	.164	.081	.167	.397	.047	.812	.039
Fixation Duration Paragraph2 Mean	-.041	.837	.295	.149	.074	.014	.946	.140	.482	.022
Fixation Duration Paragraph3 Mean	.137	.487	.043	.829	.027	.114	.563	.085	.666	.031
Fixation Duration Paragraph4 Mean	.043	.828	.115	.562	.021	-.146	.459	.245	.217	.041
Fixation Duration Pie Chart Mean	.179	.358	.096	.619	.061	-.154	.438	.142	.474	.019
Fixation Duration Sarah Case Mean	.076	.698	-.224	.257	.037	-.059	.768	.049	.806	.003
Fixation Duration Task Instruction Mean	-.026	.896	.146	.464	.018	-.106	.583	.309	.116	.070
Total Fixation Duration Essay Mean	.254	.182	.131	.488	.119	.371	.046	.091	.616	.184
Total Fixation Duration Graph Title1 Mean	.075	.707	-.018	.930	.004	.033	.865	.257	.187	.076
Total Fixation Duration Graph Title2 Mean	.087	.661	.061	.757	.017	.124	.533	-.046	.818	.011
Total Fixation Duration Letter-to-Editor Mean	-.042	.835	.071	.721	.003	-.071	.723	.096	.632	.007

Total Fixation Duration Line Graph Mean	.083	.668	.199	.307	.065	.276	.139	.167	.367	.156
Total Fixation Duration Newspaper Article Mean	.252	.203	-.122	.535	.044	-.039	.846	-.042	.835	.005
Total Fixation Duration Paragraph1 Mean	.363	.060	-.055	.772	.112	-.114	.567	.129	.517	.013
Total Fixation Duration Paragraph2 Mean	.111	.575	.059	.767	.023	-.200	.301	.336	.086	.077
Total Fixation Duration Paragraph3 Mean	.074	.706	.121	.540	.030	-.184	.349	.244	.217	.043
Total Fixation Duration Paragraph4 Mean	-.104	.602	.135	.498	.013	-.155	.435	.168	.398	.023
Total Fixation Duration Pie Chart Mean	.250	.205	-.189	.336	.045	.152	.438	.109	.574	.054
Total Fixation Duration Sarah Case Mean	.192	.309	.207	.272	.124	-.186	.349	.114	.565	.024
Total Fixation Duration Task Instruction Mean	.232	.239	-.030	.878	.047	-.046	.819	.027	.892	.001
Fixation Count Essay Mean	.104	.586	.239	.215	.095	.072	.706	.267	.166	.098
Fixation Count Graph Title1 Mean	.021	.916	.008	.967	.001	.028	.887	.200	.308	.047
Fixation Count Graph Title2 Mean	.008	.968	.134	.499	.019	-.017	.933	.119	.552	.012
Fixation Count Letter-to-Editor Mean	.025	.899	.038	.849	.003	.199	.314	-.223	.259	.039
Fixation Count Line Graph Mean	.052	.789	.235	.232	.072	-.043	.828	.131	.510	.013
Fixation Count Newspaper Article Mean	.245	.216	-.125	.525	.041	.192	.332	-.003	.987	.036
Fixation Count Paragraph1 Mean	.187	.338	.080	.681	.058	.000	1.000	.261	.181	.068
Fixation Count Paragraph2 Mean	-.041	.837	.295	.149	.074	.057	.776	-.021	.918	.002
Fixation Count Paragraph3 Mean	-.063	.747	.221	.265	.037	-.041	.838	.002	.994	.002
Fixation Count Paragraph4 Mean	-.303	.120	.299	.125	.080	-.122	.537	-.045	.818	.023
Fixation Count Pie Chart Mean	.185	.351	-.171	.388	.028	-.182	.331	.438	.023	.135

Fixation Count Sarah Case Mean	.089	.640	.245	.203	.093	.085	.670	.008	.968	.008
Fixation Count Task Instruction Mean	.180	.364	.005	.980	.033	-.090	.653	.100	.618	.008
Visit Duration Essay Mean	-.414	.028	.473	.013	.175	-.017	.933	-.172	.383	.033
Visit Duration Graph Title1 Mean	-.135	.498	.109	.583	.014	-.127	.525	.127	.523	.014
Visit Duration Graph Title2 Mean	-.008	.968	.052	.794	.002	.006	.977	-.002	.991	.000
Visit Duration Letter-to-Editor Mean	-.264	.174	.327	.095	.080	.182	.346	-.328	.094	.074
Visit Duration Line Graph Mean	.018	.917	-.471	.011	.213	.082	.682	-.076	.703	.005
Visit Duration Newspaper Article Mean	.160	.419	.001	.996	.026	.124	.531	.036	.854	.022
Visit Duration Paragraph1 Mean	-.074	.705	-.157	.424	.043	-.035	.858	.148	.458	.017
Visit Duration Paragraph2 Mean	.038	.849	.089	.655	.013	.195	.313	-.331	.092	.075
Visit Duration Paragraph3 Mean	-.213	.272	.338	.084	.079	-.010	.961	.144	.470	.019
Visit Duration Paragraph4 Mean	-.396	.041	.272	.155	.110	-.017	.929	-.187	.343	.039
Visit Duration Pie Chart Mean	.228	.258	-.122	.544	.035	-.476	.013	.299	.110	.156
Visit Duration Sarah Case Mean	.165	.383	.208	.275	.109	-.019	.925	.160	.419	.023
Visit Duration Task Instruction Mean	.053	.762	-.507	.006	.230	-.078	.696	.081	.685	.006
Total Visit Duration Essay Mean	.070	.721	.177	.365	.050	.118	.546	.144	.461	.054
Total Visit Duration Graph Title1 Mean	.141	.479	-.054	.785	.014	.010	.960	.237	.227	.059
Total Visit Duration Graph Title2 Mean	-.246	.213	.182	.354	.043	-.006	.977	.085	.672	.007
Total Visit Duration Letter-to-Editor Mean	.112	.571	.041	.835	.020	.228	.248	-.216	.272	.043
Total Visit Duration Line Graph Mean	-.192	.323	-.090	.642	.065	-.048	.809	-.009	.964	.003

Total Visit Duration Newspaper Article Mean	.153	.421	-.379	.051	.102	.214	.280	-.041	.835	.038
Total Visit Duration Paragraph1_Mean	.039	.846	.009	.965	.002	.001	.996	.229	.245	.052
Total Visit Duration Paragraph2 Mean	-.016	.937	.138	.488	.017	.130	.514	-.111	.578	.013
Total Visit Duration Paragraph3 Mean	.066	.739	.091	.646	.019	-.021	.917	-.027	.892	.002
Total Visit Duration Paragraph4 Mean	.159	.422	-.021	.915	.022	-.185	.348	-.025	.898	.040
Total Visit Duration Pie Chart Mean	.059	.779	.085	.689	.017	-.291	.115	.507	.008	.176
Total Visit Duration Sarah Case Mean	.065	.745	-.156	.434	.017	.012	.953	.030	.882	.001
Total Visit Duration Task Instruction Mean	.044	.825	.041	.836	.006	-.069	.729	.091	.647	.006
Visit Count Essay Mean	.230	.238	.053	.785	.069	.117	.547	.141	.472	.052
Visit Count Graph Title1 Mean	.076	.704	-.057	.777	.004	.088	.649	.218	.260	.077
Visit Count Graph Title2 Mean	.057	.771	.158	.424	.038	.004	.984	.181	.361	.033
Visit Count Letter-to-Editor Mean	.156	.432	-.062	.756	.017	.183	.357	-.185	.351	.030
Visit Count Line Graph Mean	.065	.738	.194	.322	.056	.084	.673	.092	.642	.024
Visit Count Newspaper Article Mean	.240	.224	-.191	.333	.043	.166	.402	.024	.904	.032
Visit Count Paragraph1 Mean	.298	.129	-.079	.682	.068	-.001	.994	.215	.276	.046
Visit Count Paragraph2 Mean	.073	.715	.013	.948	.007	-.028	.887	.121	.543	.012
Visit Count Paragraph3 Mean	.184	.350	.046	.815	.045	-.075	.707	.065	.746	.004
Visit Count Paragraph4 Mean	-.144	.468	.156	.431	.020	-.150	.450	.018	.929	.020
Visit Count Pie Chart Mean	.183	.357	-.164	.407	.027	-.117	.535	.402	.037	.123
Visit Count Sarah Case Mean	.053	.785	.229	.240	.069	.027	.893	-.081	.685	.005

Visit Count Task Instruction Mean	.189	.343	-.081	.681	.025	.055	.779	.177	.367	.045
Rate of Total Fixation Duration Essay	.014	.945	.005	.981	.000	.181	.360	-.020	.920	.029
Rate of Total Fixation Duration Graph Title1	.232	.229	-.355	.069	.088	.139	.482	.074	.706	.036
Rate of Total Fixation Duration GraphTitle2	-.267	.172	.004	.984	.070	.186	.348	-.052	.793	.027
Rate of Total Fixation Duration Letter-to-Editor	.082	.678	.047	.813	.013	.202	.292	-.383	.049	.101
Rate of Total Fixation Duration Line Graph	-.210	.280	-.074	.703	.067	-.095	.635	.005	.979	.008
Rate of Total Fixation Duration Newspaper Article	.057	.761	-.373	.053	.119	.215	.276	-.038	.845	.039
Rate of Total Fixation Duration Paragraph1	.176	.367	-.292	.139	.058	-.034	.862	.228	.247	.045
Rate of Total Fixation Duration Paragraph2	.156	.432	-.165	.407	.023	.071	.717	-.265	.178	.054
Rate of Total Fixation Duration Paragraph3	.168	.396	.008	.968	.030	.013	.946	-.323	.095	.100
Rate of Total Fixation Duration Paragraph4	-.267	.176	.172	.381	.049	-.187	.341	-.063	.745	.052
Rate of Total Fixation Duration Pie Chart	.134	.501	-.002	.994	.018	-.107	.580	.302	.125	.066
Rate of Total Fixation Duration Sarah Case	.036	.858	-.141	.480	.015	-.028	.888	.001	.996	.001
Rate of Total Fixation Duration Task Instruction	.013	.948	-.153	.443	.021	-.076	.704	.076	.703	.005

Each row contains the results from the two separate regressions, with results first for Task 1 and then for Task 2. P-values less than .05 highlighted in bold. N is 38.