

# IELTS Research Reports Online Series

## Independent versus Integrated Writing Tasks for IELTS Writing: A Mixed Methods Study



Aek Phakiti

---

## Independent versus Integrated Writing Tasks for IELTS Writing: A Mixed Methods Study

This study investigates: 1) ESL international students' writing performance (independent versus integrated writing tasks), their writing processes, and attitudes and beliefs toward an integrated writing task; and (2) raters' evaluation of integrated writing performance and their reactions, and perceptions toward an integrated writing task.

### Funding

This research was funded by the IELTS Partners: British Council, IDP IELTS, and Cambridge University Press & Assessment. Grant awarded 2017.

### Publishing details

Published by the IELTS Partners: British Council, IDP IELTS, and Cambridge University Press & Assessment © 2024.

This publication is copyright. No commercial re-use. The research and opinions expressed are of individual researchers and do not represent the views of IELTS. The publishers do not accept responsibility for any of the claims made in the research.

### How to cite this report

Phakiti, A. (2024). Independent versus Integrated Writing Tasks for IELTS Writing: A Mixed Methods Study. *IELTS Research Reports Online Series*, No. 1/24. British Council, IDP IELTS, and Cambridge University Press & Assessment. Available at <https://ielts.org/researchers/our-research/research-reports>

### Keywords

IELTS Writing test, integrated writing, independent writing, online test, cognitive validity, scoring validity, raters, reading-to-write

# Introduction

This study by Phakiti was conducted with support from the IELTS Partners (British Council, IDP: IELTS Australia, and Cambridge Assessment English), as part of the IELTS joint-funded research program. Research funded by the British Council and IDP: IELTS Australia under this program complement those conducted or commissioned by Cambridge Assessment English, and together inform the ongoing validation and improvement of IELTS.

A significant body of research has been produced since the joint-funded research program started in 1995, with over 200 empirical studies receiving grant funding. Many of the studies have been published in academic journals, in several IELTS-focused volumes in the *Studies in Language Testing* series (<http://www.cambridgeenglish.org/silt>), and in the *IELTS Research Reports* series. Since 2012, for timely access, the reports have been published on the IELTS website immediately after the peer review and revision process.

Integrated writing tests are tasks that involve other forms of language input (such as reading or listening) in producing written texts, and thus require a test-taker to use more than one language skill. Integrated writing tasks have become more common in high-stakes English language assessment and in classroom-based activities, but further research is needed to understand the nuanced differences between integrated and traditional writing tasks.

The current project delves into understanding the perceptions and performances of test-takers and raters regarding integrated writing tasks in IELTS Writing exams. Adopting a mixed methods research design, the study encompasses various aspects, including the relationships between IELTS scores and writing performances, the processes involved in completing the integrated writing task, and the perceptions towards its usefulness and feasibility. The results show that the test-takers in this study generally performed better on independent writing tasks compared to integrated ones. Distinct features in integrated writing essays across different proficiency levels were identified, particularly concerning source text utilisation. The survey findings showed that while test-takers found the integrated writing task beneficial for university preparation, raters expressed concerns about its practicality and cognitive load.

The study provides valuable insights into the dynamics of integrated writing tasks that may be considered for the IELTS testing context. By examining the interplay between independent and integrated writing performances, identifying key features of integrated writing responses, and exploring perceptions of test-takers and raters, the research provides another piece of evidence in this domain. The findings underscore the importance of considering both quantitative and qualitative aspects when evaluating writing tasks and highlight the potential benefits and challenges associated with incorporating integrated writing tasks in IELTS. Understanding these nuances can inform future test design and assessment practices, ultimately contributing to the enhancement of language proficiency evaluation methods and ensuring their relevance and effectiveness in diverse educational settings.

**Dr Elaine Schmidt**  
**Senior Research Manager**  
**Cambridge University Press & Assessment**

# Independent versus Integrated Writing Tasks for IELTS Writing: A Mixed Methods Study

## Abstract

The current project aims to understand test-takers' and raters' reactions, leading to empirical analysis to investigate the following areas:

1. the relationships between reported official IELTS scores, independent writing scores, and integrated writing scores
2. writing processes to complete the integrated writing task
3. scoring processes of integrated writing responses
4. features of integrated writing responses across different success levels
5. perceptions toward the usefulness of an integrated essay task by test-takers and raters.

A mixed methods research design was adopted to address these areas of investigation. Quantitative data were from 154 ESL international students who provided their previous official IELTS scores (overall, Reading and Writing scores), took an online writing test that included an independent writing task and integrated writing task, and answered two subsequent questionnaires about their test-taking processes and perceptions toward an integrated writing task.

Eight test-takers with different success levels in the integrated writing task were recruited for an individual interview about their test-taking experiences and perceptions of the usefulness of an integrated writing task for their study preparation before university commencement. Two domestic students were also recruited for an interview after they completed the test.

Two IELTS examiners were employed to rate test-takers' responses to both writing tasks. Whenever they could, they were asked to provide comments on test-takers' responses after they rated. They were also interviewed about their rating processes and perceptions toward an integrated writing task.

Key findings in the current study include:

1. there were statistical relationships between official IELTS scores and independent and integrated writing performance (with moderate statistical effect sizes)
2. independent writing scores accounted for about 25% of the integrated writing score variance
3. test-takers performed significantly better on the independent writing task than the integrated writing task
4. there were statistical relationships between reported test-taking and writing processes in the independent and integrated writing tasks
5. prominent features of integrated writing essays between different success levels based on raters' comments were identified, particularly in relation to critical issues of source text use

6. quantitative and qualitative data suggested that test-takers found an integrated writing task useful to help them prepare for university study
7. raters reported concerns about the practicality of an integrated writing task in terms of their cognitive load, preparation prior to rating, and required time for accurate evaluation.

The current study has shed new light on the interrelationships between independent and integrated writing performances, unique features of integrated writing performance and processes, individual differences in integrated writing performance, and perceptions toward the usefulness and feasibility of an integrated writing task that may be adopted as part of the IELTS. Implications and limitations of the current study are discussed.

---

## Author's biodata

### Aek Phakiti

Aek Phakiti is an Associate Professor in TESOL at The University of Sydney. His research focuses on language testing and assessment, second language acquisition, and research methods in language learning. He is the author of *Strategic Competence and EFL Reading Test Performance* (Peter Lang, 2007), *Experimental Research Methods in Language Learning* (Bloomsbury, 2014), and, with Carsten Roever, *Quantitative Methods for Second Language Research: A Problem-Solving Approach* (Routledge, 2018), *Language Assessment: A Practical Approach* (RELC, 2021), *Assessment for Language Teaching* (Cambridge University Press, 2024), and *Calibration and Confidence in Language Learning and Use* (Cambridge University Press, forthcoming). With Brian Paltridge, he has edited *Continuum Companion to Research Methods in Applied Linguistics* (2010, Continuum) and *Research Methods in Applied Linguistics: A Practical Resource* (2015, Bloomsbury). With Peter De Costa, Luke Plonsky and Sue Starfield, he is a co-editor of *The Palgrave Handbook of Applied Linguistics Research Methodology* (Palgrave, 2018).

He was Associate Editor of *Language Assessment Quarterly* (2012–2021), *University of Sydney Papers in TESOL* (2006–2013) and *University of Sydney Journal of TESOL* (2023 to present).

In 2010, Aek was a recipient of the TOEFL Outstanding Young Scholar Award. He received three University of Sydney Faculty Teaching Excellence Awards (2009, 2017, 2019). He was Vice President of ALTAANZ (Association for Language Testing and Assessment of Australia and New Zealand (2014-2016) and is Vice President of ALAA (Applied Linguistics Association of Australia, from 2024).

---

---

# Table of contents

<b>1</b>	<b>Introduction .....</b>	<b>9</b>
<b>2</b>	<b>Review of the literature .....</b>	<b>10</b>
2.1	Previous IELTS Writing research.....	10
2.1.1	Test-takers' writing performance, processes and factors.....	10
2.1.2	Relevance and authenticity of IELTS Writing tasks to academic contexts .....	11
2.1.3	Rating and rater processes.....	11
2.2	Research on assessing integrated writing in academic settings.....	12
2.2.1	The constructs of reading-to-write tasks .....	13
2.2.2	Previous research that compares independent writing with integrated writing .....	13
2.2.3	Research that investigates test-takers' processes in integrated writing .....	14
2.2.4	Research that examines factors affecting integrated writing performance.....	14
2.2.5	Task and delivery characteristics.....	15
2.2.6	Language proficiency level.....	15
2.2.7	Raters and rating .....	15
<b>3</b>	<b>Aims of the current study.....</b>	<b>16</b>
3.1	The validity framework.....	17
<b>4</b>	<b>Research questions.....</b>	<b>18</b>
<b>5</b>	<b>Research method.....</b>	<b>18</b>
5.1	Research design .....	18
5.2	Research context.....	20
5.3	Research participants .....	20
5.3.1	Test-taker participants.....	20
5.3.2	Test-taker interviewees.....	20
5.3.3	Raters.....	21
5.4	Research instruments.....	22
5.4.1	An overview of the IELTS Writing test .....	22
5.4.2	IELTS Independent Writing Task .....	22
5.4.3	IELTS Integrated Writing Task .....	22
5.4.4	Reflective questionnaire 1 .....	23
5.4.5	Reflective questionnaire 2.....	23
5.4.6	Semi-structured interview questions for test-takers.....	24
5.4.7	Semi-structured interviews for raters .....	24
5.4.8	Analytic scoring rubric for the independent writing task .....	25
5.4.9	Analytic scoring rubric for the integrated writing task .....	25
5.5	Data collection.....	25
5.5.1	Ethical considerations.....	25
5.5.2	Key stages in data collection .....	26
5.6	Data preparation and analyses .....	27
5.6.1	Quantitative data analysis .....	27
5.6.2	Qualitative data analysis .....	28
5.7	Reliability of quantitative data .....	30
5.7.1	Score reliability.....	30
5.7.2	Questionnaire reliability.....	33
<b>6</b>	<b>Findings and discussion .....</b>	<b>34</b>
6.1	Research question 1: What is the relationship between integrated and independent writing performance?.....	34
6.1.1	Quantitative findings .....	34
6.1.2	Qualitative findings from raters .....	36
6.2	Research question 2: Do test-takers differ in performances between independent and integrated writing tasks?.....	39
6.2.1	Quantitative findings .....	39
6.2.2	Qualitative findings from test-takers' interviews.....	40
6.2.3	Qualitative findings from raters' interviews .....	45

6.3	Research question 3: What are the relationships between reported IELTS scores and independent and integrated writing performances? .....	49
6.4	Research question 4: Do test-takers with different integrated writing success levels differ across different writing criteria? .....	53
6.4.1	Quantitative findings .....	53
6.4.2	Qualitative findings.....	55
6.5	Research question 5: What is the nature of test-takers' reported reading, writing and test-taking strategies during independent and integrated writing tasks?.....	63
6.5.1	Quantitative findings .....	63
6.5.2	Qualitative findings.....	67
6.6	Research question 6: What are test-takers' perceptions toward the usefulness and relevance of an integrated writing task to academic study? .....	77
6.6.1	Quantitative findings .....	77
6.6.2	Qualitative findings.....	79
6.7	Research question 7: What are raters' perceptions toward an integrated writing task for assessing academic writing? .....	84
6.7.1	Perceptions of the rater training and meeting sessions .....	84
6.7.2	Key considerations for stakeholders in the use of an integrated writing task .....	85
6.7.3	Raters' perceptions toward the feasibility of an integrated writing task for an official IELTS .....	90
<b>7</b>	<b>Implications and conclusions .....</b>	<b>91</b>
7.1	Summary of the key findings.....	91
7.1.1	Relationship between independent and integrated writing performance.....	91
7.1.2	Assessing integrated writing responses .....	91
7.1.3	Differences between independent and integrated writing performance.....	91
7.1.4	IELTS scores and integrated writing performance.....	92
7.1.5	Integrated writing responses by success levels .....	92
7.1.6	Cognitive processes in integrated writing.....	93
7.1.7	Test-takers' perceptions of an integrated writing task.....	93
7.1.8	Raters' perceptions of an integrated writing task.....	93
7.2	Strengths and limitations of the current study .....	94
7.3	Insights into the nature of integrated writing in the current study.....	96
7.4	Implications for IELTS and future research.....	98
7.5	Concluding remarks .....	98
	<b>References .....</b>	<b>99</b>
	<b>Appendix 1: Screenshots of the online writing test.....</b>	<b>105</b>
	<b>Appendix 2: IELTS independent writing task.....</b>	<b>107</b>
	<b>Appendix 3: IELTS integrated writing task.....</b>	<b>108</b>
	<b>Appendix 4: The reflective questionnaire for the independent writing task.....</b>	<b>110</b>
	<b>Appendix 5: The reflective questionnaire for the integrated writing task.....</b>	<b>111</b>
	<b>Appendix 6: Semi-structured interview questions for test-takers .....</b>	<b>112</b>
	<b>Appendix 7: Semi-structured interview questions for raters.....</b>	<b>113</b>
	<b>Appendix 8: Analytical rubric for the independent writing task .....</b>	<b>114</b>
	<b>Appendix 9: Analytical scoring rubric for the integrated writing task .....</b>	<b>116</b>
	<b>Appendix 10: Coding schemes for test-taker interviews .....</b>	<b>118</b>
	<b>Appendix 11: Coding schemes for raters' integrated writing comments .....</b>	<b>120</b>
	<b>Appendix 12: Coding schemes for rater interviews.....</b>	<b>122</b>

---

## List of tables

Table 1: Demographic and writing performance information of the interviewees.....	21
Table 2: Taxonomy of a reflective questionnaire for the independent writing task .....	23
Table 3: Taxonomy of a reflective questionnaire for the integrated writing task .....	24
Table 4: Adjacent agreements between the two raters (accompanied by exact agreements) (N = 154).....	30
Table 5: Cronbach's alpha coefficients (N = 154).....	31
Table 6: Pearson correlation coefficients between the two raters' scores for the independent writing task for the different criteria (N = 154).....	31
Table 7: Pearson correlation coefficients between the two raters' scores for the independent writing task for each scoring criterion (N = 154).....	32
Table 8: Interclass correlation coefficients for the independent writing task (N = 152).....	32
Table 9: Interclass correlation coefficients for the integrated writing task (N = 152).....	32
Table 10: Cronbach's alpha coefficients of the reflective questionnaire on the independent writing task (N = 154) ..	33
Table 11: Cronbach's alpha coefficients of the reflective questionnaire on the integrated writing task (N = 154) ....	33
Table 12: Descriptive statistics of test-takers' independent writing scores based on two raters (N = 154).....	34
Table 13: Descriptive statistics of test-takers' integrated writing task scores based on the two raters' scores.....	34
Table 14: Pearson correlations between independent and integrated writing scores (N = 154) .....	35
Table 15: Pearson correlations between the five criteria and total integrated writing scores (N = 154) .....	36
Table 16: Mean scores of performances in the independent and integrated writing tasks (N = 154) .....	39
Table 17: T-test results between independent and integrated writing performances (N = 154) .....	40
Table 18: Effect sizes of the differences among independent and integrated writing task performance criteria .....	41
Table 19: Descriptive statistics of the overall IELTS, IELTS Reading and IELTS Writing scores (N = 152) .....	49
Table 20: Pearson correlation coefficients between reported IELTS scores and independent and integrated writing task performance (N = 152) .....	50
Table 21: Multiple regression model coefficients.....	51
Table 22: Mean scores of the three groups (N = 154).....	53
Table 23: MANOVA tests of between-subject effects .....	54
Table 24: Tamhane post-hoc tests .....	55
Table 25: Descriptive statistics of the reflective questionnaire 1 .....	63
Table 26: Frequency counts of the reflective questionnaire 1 .....	64
Table 27: Descriptive statistics of the reflective questionnaire on the integrated writing task.....	65
Table 28: Frequency counts of the reflective questionnaire 2 .....	65
Table 29: Pearson correlations between independent and integrated strategies .....	67
Table 30: Descriptive statistics of items 16 to 19.....	77
Table 31: Frequency counts of items 16 to 19.....	77
Table 32: Pearson correlations between items 16 to 19 .....	78
Table 33: Descriptive statistics of items 16 to 19 by the three success groups (N = 154) .....	78
Table 34: The test of between-subjects effects (N = 154).....	79

## List of figures

Figure 1: Research design and procedures .....	19
Figure 2: Cui's integrated essay response.....	42
Figure 3: The hypothesised path model of the relationships between IELTS variables and independent and integrated writing variables (N = 152) .....	52
Figure 4: Group 1's features of integrated writing performance .....	56
Figure 5: Group 2's features of integrated writing performance .....	58
Figure 6: Group 3's features of integrated writing performance .....	61
Figure 7: Key features and requirements for successful integrated writing.....	97

---

# 1 Introduction

The statistics of international students in Australian education in 2023 suggest a consistently rising number of international students in higher education over years (Department of Education, the Australian Government, 2023). For example, the number of international students between January and August 2023 (N = 725,582) is 31% more than that of the same period in 2022. The current statistics show that 48% of the students (the highest proportion) are in the higher education section (N = 411,576), with 22% more students compared to the same period in 2022. The statistics of international students in Australia confirm that international students are invaluable assets for the Australian economy and the Australian education sector. It is imperative to ensure that international students are well-supported and prepared to succeed in their educational and personal goals.

English as a second language (ESL) international students (hereafter international students) in English-medium universities such as in Australia and the UK require communicative and academic language skills to read, write, speak, understand, and learn. They are expected to be able to illustrate their learning achievement by combining what they have learnt from lectures and tutorials with other additional sources, such as academic books, journals, data and multimedia, to show their academic learning, progression and contributions. Across academic disciplines, written essays, assignments and research papers are typical tasks requiring students to integrate information from several sources appropriately. Integrated academic writing tasks are challenging for international students to master (e.g., Grabe & Zhang, 2013, 2016) due to various reasons such as first language writing conventions, limited practical English reading and writing proficiency, and a lack of experience in integrating reading and writing in academic texts (e.g., Hirvela, 2016; Shaw & Pecorari, 2013).

Early research (e.g., Mullins et al., 1995) found that international students had three or four times more difficulties than domestic students when writing academic assignments. Pecorari (2008) found that many ESL students begin their academic writing by copying from existing texts, indicating a lack of adequate and appropriate preparation in L2 English academic writing. Research by Phakiti and Li (2011) found a positive correlation coefficient of 0.68 ( $R^2 = 0.46$ ; attenuated correlation = 0.80) between academic reading and writing difficulties. According to Shaw and Pecorari (2013), one major challenge for ESL students is knowing how to appropriately address intertextuality in academic writing. Many ESL students find dealing with intertextuality challenging as it requires understanding how to move from reading to writing (Hirvela, 2016).

To ensure study success and sustain high-quality education service for international students, it can be argued that ESL international students in higher education need to be well-prepared to cope with reading-to-write activities before they commence their studies in English-medium academic programs. One way to help them develop such skills is by promoting a positive washback of high-stakes English for academic purposes (EAP) tests such as the International English Language Testing System (IELTS). In Australia, IELTS is the most widely used English for academic purposes (EAP) test for ESL international students to illustrate their academic English proficiency for study admission.

Although integrated writing assessment has emerged in other high-stakes EAP tests (e.g., TOEFL iBT) and classroom teaching (Gebriel & Plakans, 2014; Plakans et al., 2019; Weigle & Malone, 2016), in its current form, IELTS Writing does not have an integrated writing task that requires test-takers to extract information from written and audio texts to be subsequently included in their written essays.

---

The current study explores whether, or the extent to which, an integrated writing task (reading-to-write task) is feasible for inclusion in the IELTS test and if so, what benefits it will have for international students and what further considerations the IELTS developers need to do to include this writing task in the Official IELTS.

This project investigates the nature of test-takers' integrated writing performances, their cognitive processes, and perceptions toward an online integrated writing task (hereafter called the integrated IELTS Writing task). It also focuses on raters' comments, rating processes, and reactions to and perceptions of a potential integrated IELTS Writing task. The study employed a mixed methods approach using online writing tasks, questionnaires, individual interviews with test-takers and raters, stimulated recalls, and case studies of selected test-takers.

## 2 Review of the literature

This section begins a brief review of previous IELTS Writing research. It then reviews previous research on integrated writing assessments that informs the directions of the current project (e.g., aims and research questions and design).

### 2.1 Previous IELTS Writing research

Eighteen funded IELTS research projects have examined various areas related to IELTS Writing. Three main areas of research are: (1) test-takers' writing performances, processes, and factors affecting writing performance; (2) relevance and authenticity of IELTS Writing tasks to academic and workplace contexts; and (3) raters and rating issues.

#### 2.1.1 Test-takers' writing performance, processes and factors

Seven studies were found to be related to this theme. Mickan and Slater (2003), who examined the linguistic patterns of test-takers' written texts, suggested evaluation criteria for judging test-takers' writing performance. O'Loughlin and Wigglesworth (2003) investigated task difficulty in IELTS Writing Task 1 among different proficiency levels (high, medium and low), focusing on the effect of the amount and form of presentation of information in a table or diagram on test-takers' writing performance. Yu et al. (2011) examined the nature of test-takers' cognitive and writing processes when taking IELTS Writing Task 1 before and after a short training course. It was found that the writing processes in this task included three stages (comprehending task instruction, followed by understanding graphic information, and lastly, writing the task response). As a follow-up study, Yu et al. (2017) examined test-takers' cognitive processes when writing their responses to IELTS Writing Task 1 using an eye-tracking system, retrospective interviews, and focus groups. The study found the cognitive processes required to complete IELTS Writing Task 1 complex, especially in relation to the interaction between test-takers' characteristics, such as familiarity with a given type of graph, writing ability, and task features, such as information provided in a graph. Weir et al. (2007) compared modes of IELTS Writing delivery (online versus paper) to investigate whether they resulted in differential performance. No significant differences were found between the two modes in terms of writing processes. Computer familiarity and anxiety did not appear to influence performance negatively.

Riazi and Knox (2013) examined the role of test-takers' first language (Arabic L1, Hindi L1, and European-based L1) in writing features of IELTS Writing Task 2. While quantitative analysis did not suggest differences among the three L1 groups, discourse analysis of scripts suggested some differences in using discursive patterns among different L1 groups. Barkaoui (2016) measured changes in test-takers' linguistic responses to IELTS Writing Task 2 (linguistic accuracy, syntactic complexity, lexical density, diversity

---

and sophistication, and cohesion) across three occasions. The study found that IELTS scripts on later occasions were significantly longer and more complex, accurate, and coherent than the earlier ones. Finally, Révész et al. (2017) explored test-takers' cognitive processes during an online IELTS Writing Task 2 via an eye-tracking device, key-stroke logging, and stimulated recalls. The roles of phonological short-term memory and executive control functions were also examined.

This review suggests that IELTS-funded projects transitioned from researching Task 1 to Task 2 Writing. This indicates a void to explore test-takers' performance and processes in an integrated writing task as a potential writing task for IELTS.

### 2.1.2 Relevance and authenticity of IELTS Writing tasks to academic contexts

Four studies were related to academic settings. Moore and Morten (1999) evaluated the authenticity of IELTS Writing Task 2 by comparing performances in this task to performances in university assignments. The study found that written arguments, as required in IELTS Writing Task 2, were similar to some university writing tasks. However, most university tasks required students to use external sources to complete academic writing. The author suggested that IELTS Writing Task 2 should resemble university writing tasks more closely to increase test authenticity and prepare students to cope with university language demands. This study provides a relevant reason for the current study to investigate whether an integrated writing task should be included in a new IELTS test. Ingram and Bayliss (2007) aimed to identify the extent to which students' behaviour in IELTS matched that in tertiary contexts and whether the required level of IELTS performance was sufficiently high for IELTS results to indicate that students could cope with actual academic writing. IELTS Writing was among the four areas focused on in this study. Through four case studies, Paul (2007), in work that complemented that of Ingram and Bayliss (2007), matched IELTS language behaviours with those in real-life academic settings. The study found that IELTS Writing was generally predictive of language behaviours in real-life academic study. However, in all four cases in the study, academic demands played a critical factor in students' writing difficulties and experiences. Uitdenbogerd et al. (2018) examined the appropriate level of academic writing competence required by PhD students in computer science in Australia, how it might change during their courses of study, and how PhD students and supervisors perceived the level of academic writing sufficiency or preparedness. The study used a survey method, asking PhD students and supervisors to complete a questionnaire. It was found that student participants indicated difficulties in writing cohesion, clarity, accuracy of expression, and structure. Similarly, PhD supervisors reported writing cohesion and clarity issues among their students.

Previous IELTS research has shed light on the importance of academic writing in educational settings. In all the studies, there is a clear case for the need to examine whether an integrated writing task in IELTS Writing can be viable and value-added in academic settings in the 21<sup>st</sup> century, which require the integration of various authentic sources.

### 2.1.3 Rating and rater processes

Three studies were found to be directly related to issues of rating and raters in IELTS Writing. Examining features of the rating of academic writing performance of IELTS Writing Tasks 1 and 2 (General Training), Mickan (2003) found the rating of such academic texts complex and proposed the adoption of holistic descriptors, given that language elements did not exist in isolation. Banerjee et al. (2007) focused on developing a rating scale for IELTS Writing Tasks 1 and 2 that can capture the quality of academic written texts through descriptors that reflect differences in language proficiency. Cotton and Wilson (2011) focused on marking coherence and cohesion

---

in IELTS Writing Task 2, asking whether these areas were more difficult to evaluate than others. Through think-aloud protocols and follow-up interviews, coherence and cohesion were found more challenging to rate than grammatical range, accuracy, and lexical resource.

Based on the review of the three studies, it is important to investigate rating processes in independent and integrated writing tasks. The current project focuses on human raters rather than automated scoring mechanisms and aims to gain insights into rater processes and perceptions on the usefulness of an integrated writing task.

## **2.2 Research on assessing integrated writing in academic settings**

Academic writing is a specific writing genre distinctive from other writing genres (e.g., personal and business writing) at various levels. It requires writers to understand the unique discourse structure of micro and macro texts they refer to as they write (see Hyland, 2016; Paltridge et al., 2008; Weigle, 2002; Weigle & Malone, 2016). Academic writing is also interrelated to academic literacies, which are socially, culturally, and educationally specific (see, e.g., Lillis & Tuck, 2016; Street & May, 2017). The notion of academic literacies goes beyond the academic language abilities of reading and writing. The term includes general and unique social, cultural and localised aspects of literacies. Lea and Street (1998) pointed out that academic writing is within a complex set of discourses, identities and values. People need to learn, adjust and shift their academic writing styles and genres as they move from one academic context to another. According to Paltridge and Starfield (2019) and Hyland (2016), for instance, academic writing is shaped by the educational context in which it takes place (e.g., the level of study and the academic discipline of the writer). For example, at the high school level, students are frequently asked to write a description, a summary, or an opinion essay on a topic. In contrast, at the tertiary level, they are asked to write various essay and assignment types (e.g., a critical review of a theory or discipline-specific topic). Typical tertiary academic written assignments require students to conduct laboratory or library research, to illustrate a deep understanding of the content, and to convince the readers or academics (e.g., professors, lecturers, and tutors) of a point using well-informed texts with citations of sources of information (a characteristic of integrated writing).

Theoretically, integrated writing is authentic writing involving other modes of language activities, such as reading and listening, in producing written texts. Based on Plakans (2022), the term 'integrated task' is broadly used to refer to a task that requires an individual to use more than one language skill. For example, in an assessment context, test-takers may be asked to read a text and write an essay in response to a statement and incorporate information extracted from the given text in their essay. It can be said that the more language modes and external sources are involved in an integrated task, the more complicated it is for test-takers to complete the task, as the task requires not only comprehension and writing but also synthesis skills. In the current project, an integrated writing task only involves the receptive skill of reading and the productive skill of writing.

Developing an integrated writing task for writing assessment is also challenging in many ways (see Plakans, 2022; Grabe & Zhang, 2013; Knoch & Sitajalabhorn, 2013). For example, questions such as "how many other language skills should be involved?", "which topics are appropriate?", "how complex should the texts be?", and "how long should the texts be?" need to be considered and addressed carefully. There are also issues of test practicality to consider, including test delivery methods, minimum expected word counts to make a meaningful text, time allowance, allocation of marks, and marking demands (Plakans, 2022).

---

The following sections present five themes of the relevant literature that inform the current research project.

### 2.2.1 The constructs of reading-to-write tasks

To employ an integrated writing task successfully, test developers need to consider the construct definition of integrated writing ability (i.e., what constitutes successful integrated writing; see e.g., Gebriel & Plakans, 2013; Knoch, 2022; Knoch & Sitajalabhorn, 2013; Plakans, 2008, 2009a, 2009b, 2010, 2015, 2022; Sawaki et al., 2013; Weigle, 2002; Weigle & Malone, 2016). Since the current study examines a reading-to-write integrated task (Hirvela, 2016), research that contributes to the roles of reading on integrated writing performance should be considered.

According to Plakans (2015), the reading–writing ability is known as *discourse synthesis* ability. This describes the higher-order processes in which reading interacts with writing. These processes include the connecting, selecting, and organising of text (Plakans 2015). Specifically, discourse synthesis includes: taking notes from texts; summarising texts; paraphrasing textual resources; meaning-making; selecting, combining, and synthesising information from different texts; comparing multiple viewpoints from different texts; responding to assigned reading texts; and writing up a text (Grabe & Zhang, 2013; Plakans, 2022). A TOEFL integrated writing study by Sawaki et al. (2013) described the distinctive nature of reading, listening, and writing skills during an integrated writing task, with *comprehension* as an overarching, transcendent skill that mediates the writing processes. Qualitative research (e.g., Gebriel & Plakans, 2014; Plakans, 2009b) suggests that reading is critical for success in integrated writing. For instance, Plakans (2009b) observed that high-proficient test-takers exhibited more high-quality discourse synthesis behaviours than low-proficient ones, who may struggle with basic reading comprehension and sentence structure construction. A review of the literature suggests that criteria for assessing integrated writing can include accuracy, fluency, and complexity (e.g., Cumming et al., 2005; Gebriel & Plakans, 2014).

Nevertheless, it is important to be explicit in terms of the construct definition of integrated writing ability in a given writing test, the prompts provided, the expected responses, the task selection, the test instructions, the delivery methods, the scoring criteria, and the rating procedures to be used (Knoch, 2022; Plakans, 2022). For the current study, *integrated writing ability* is defined as the ability to write an essay to address a relevant writing task, which includes accurate and appropriate comprehending, extracting, synthesising, utilising and acknowledging information from the given sources to support or make a case in one's essay. This integrated writing ability requires an individual to produce a sufficient text length that yields meaningful texts to acknowledge source texts when used; to use appropriate word choices; to create a range of more or less complex sentence structures; to organise ideas logically and conventionally; and to produce grammatically correct and pragmatically appropriate sentences.

### 2.2.2 Previous research that compares independent writing with integrated writing

According to Plakans (2012, 2022), studies that examine the differences between independent and integrated writing tasks yield intriguing yet mixed results. For example, Cumming et al. (2005) found that test-takers produced significantly longer essays in an independent task compared with an integrated one, whereas Lewkowitz (1994) found no significant difference in the word count of responses to the two tasks and observed that in an integrated writing task, more points were presented than in an independent task. Cumming et al. (2005) found that successful writers summarise more often and more effectively integrate information from source texts than less successful ones. Low-ability writers tend to use fewer source texts and depend largely on paraphrasing and using a source text verbatim (Cumming et al., 2005).

---

In addition, Cumming et al. (2005) did not find differences in the grammatical accuracy of responses to the two task types but found that test-takers had better performance in rhetorical use in their responses to an integrated task compared to those to an independent task. Watanabe (2001) found that writers exhibited more originality in response to an independent task than an integrated one.

A review of previous research suggests that comparative research can focus on the extent of the relationship between integrated and independent writing performance, differences in the quality of writing in aspects such as grammatical accuracy, content development, and the structural and rhetorical characteristics of responses to both essay types. Within-group analysis may address similarities and differences between integrated and independent performances. More research is needed to establish a relationship between independent and integrated writing performances.

### 2.2.3 Research that investigates test-takers' processes in integrated writing

According to Plakans (2022), studies have begun to examine the processes involved in completing integrated writing tasks, which include linguistic and cognitive processes. Researchers need to investigate the theoretical nature of reading-to-write processes (e.g., Hirvela, 2016; Paltridge et al., 2009), such as understanding writing task requirements, reading comprehension and text use, note-taking, planning and outlining, drafting, intertextualising, citing and referencing, and revising, monitoring, checking, and evaluating. For example, Plakans (2009b) found that compared to the behaviour of test-takers when performing an independent writing task, pre-planning in an integrated writing task is used less frequently, while planning is more predominant during integrated writing. A study on the effects of intertextuality on reading-to-write tasks by Cheong et al. (2019) found a significant prediction of discourse synthesis strategies and multiple-text comprehension on integrated writing performances in both L1 Chinese and L2 English. Research by Golparvar and Khafi (2021) found that various self-efficacy variables (e.g., writing self-efficacy) significantly predicted reading-to-write test performance (i.e., a summary writing task). In contrast, linguistic self-efficacy was predictive of discourse synthesis and source use. Payant et al. (2019) found that reading performance and anxiety predicted Canadian Academic English Language (CAEL) integrated writing performance. One predominant factor affecting integrated writing performance is topic unfamiliarity and content complexity. Plakans (2022) pointed out empirical evidence that test-takers reported some preference for integrated writing tasks as the provided sources furnished them with more ideas, words and ideas on how to organise their writing.

A review of this area of research has implications for the current project in that test-takers' processes and perceptions in both integrated and independent writing tasks should be examined as they explain the nature of integrated writing performance and how test-takers perceive an integrated writing task.

### 2.2.4 Research that examines factors affecting integrated writing performance

In language testing and assessment, Bachman and Palmers' (1996) approaches to factors affecting test performance have influenced researchers' understanding of language test performance.

---

### 2.2.5 Task and delivery characteristics

According to Plakans (2022), researchers have investigated the influences of writing genres, topics, reader audiences, and source materials on integrated writing performance. For example, Swain et al. (2009) found that more strategies were used, and their use was more prevalent when more source or skill modes were required in an integrated writing task than an independent task. Cho and Choi (2018) found that integrated writing tasks designed for a specific audience were more complete in terms of necessary background details and more accurately targeted in terms of the actual task than those designed for an unspecified audience. According to Weigle and Malone (2016), computer-mediated reading-to-write tasks can be considered as such tasks are not only convenient and authentic in technology-driven academic environments. Brunfaut et al (2018), for instance, examined the effect of test delivery modes on reading-to-write task performance and found that a statistically significant difference (yet slight) in a mean measure between performances in the two delivery modes on the reading-to-write task for the Trinity College London Integrated Skills in English (ISE) test I level (CEFR B1), but not for ISE II (CEFR B2) and ISE III (CEFR C1). The researchers also found that test-takers perceived online delivery more positively than paper-based mode. Kim et al. (2018) also found that most test-takers preferred an online delivery mode in the process-oriented integrated English Placement Test writing to its paper-based version. The current project does not aim to compare delivery modes but focuses on online delivery of both independent and integrated writing tasks.

### 2.2.6 Language proficiency level

It is undeniable that language proficiency plays a critical role in successful language use, including in integrated writing. Based on Plakans' (2022) review, test-takers who obtain high integrated writing scores exhibit more grammatical accuracy and better use of source integration, suggesting that language proficiency levels influence differential integrated writing performance. For example, although the researchers did not compare integrated and independent writing performance, Plakans et al. (2019) examined the linguistic features of complexity, accuracy, and fluency of integrated writing using multiple regression analysis. They found that fluency (as measured by word counts) was the strongest predictor of integrated writing performance, followed by accuracy. Test-takers with higher integrated writing scores have a higher language proficiency level than those who perform less successfully in this task type (see Gebрил & Plakans, 2013; Plakans, 2022; Weigle & Malone, 2016).

### 2.2.7 Raters and rating

The scoring of integrated writing needs to take into account the presence of source texts in writing. Research on rater processes of scoring addresses multi-layered issues, including: (1) assessment-related issues (e.g., assessment criteria, rater training, rater reliability, scoring accuracy, and rater agreement); and (2) rater processes which inform their scoring (Brown, 2012; Davis, 2022; Plakans, 2022). Ohta et al. (2018) found that integrated writing scores derived from a multi-trait rubric were more reliable than those derived from a holistic rubric. A range of research methods has been adopted to examine rater processes (e.g., think-aloud protocols and individual interviews, e.g., Cumming et al., 2005; Gebрил & Plakans, 2014). The key focus in this area is the question of how raters go about assigning scores. Some studies focus on raters' decision-making processes. For example, Cumming et al. (2005) found that compared to when scoring independent writing tasks, raters paid more attention to rhetoric and content in integrated essays.

Gebрил and Plakans (2014), who examined rater decision-making behaviours in integrated writing scoring, found that raters spent more time on judgment strategies (evaluating the quality of writing) than on interpreting strategies (making sense of content in writing).

---

The researchers also found that raters exhibited behaviours such as checking source information, comparing test-takers' language to the source texts, and acknowledging source texts in writing. Raters were also found to ask questions of adequacy, clarity, quality, and relevance of source texts in writing. Gebril and Plakans (2014) argued that these are examples of raters' processes that would not necessarily be found in independent writing tasks. The researchers also found that test-takers' proficiency levels determined the kind of attention raters paid when rating integrated writing. For example, for a low-ability test-taker (e.g., Level 1), raters tended to pay more attention to 'basic language use' (e.g., linguistic errors, coherence). Still, for a mid-ability test-taker (e.g., Level 3), they focused on the 'organisation' of writing. For a high-ability test-taker (e.g., Level 5), they focused on the 'development' of the topic. Gebril and Plakan's (2014) study has several implications for the current project in that, for example, "what are raters' decision-making processes when rating integrated essays?", "how do they consider source texts in their rating?", and "what are the key challenges they face in rating integrated essays, as compared with rating independent essays?".

In sum, research on raters' behaviours and processes in assessing integrated writing performance can shed light on the scoring validity of the assessment/ rating processes that the current study can use as a base in the IELTS context.

### 3 Aims of the current study

The review of the literature suggests the need for an IELTS study that examines the nature of integrated writing performance and determines the viability of this type of writing assessment for IELTS of the 21<sup>st</sup> century. Integrated writing is considered authentic and reflective of what happens in real-world academic contexts. Previous research on a reading-to-write task and integrated writing has informed the thinking behind the current study in terms of the area of research and the methodology to be used to conduct the research.

*It is essential to know whether there is a potential for an integrated writing task to be included in the official IELTS Writing. The current study, therefore, aims to navigate what it may look like if an online integrated writing task (academic module) is implemented and what both test-takers' performance in an integrated writing task (i.e., simulated IELTS integrated writing task) and raters' processes, behaviours, and perceptions regarding integrated writing performance would be like.*

Introducing an integrated reading-to-write test task can be challenging for test-takers and raters who have rarely encountered it. For test-takers, writing would no longer be mainly about drawing on their own ideas, knowledge, or experience (Brown et al., 2005) but about synthesising and using source texts in their writing (Plakans, 2022). For IELTS raters, scoring integrated writing would be new to them. Their scoring would no longer be based purely on the criteria they use to evaluate an IELTS independent writing task (i.e., IELTS Writing Task 2) but on additional criteria that may include acceptable textual borrowing, intertextuality, and the quality of integrated writing texts (Gebril & Plakans, 2014).

---

### 3.1 The validity framework

In language testing and assessment, validation is the ongoing process of collecting and analysing evidence that supports test validity (e.g., evidence that a test measures what it is supposed to measure; evidence that inferences drawn from scores are meaningful and valuable for their intended purpose; and evidence that a test is fair for all test-takers (see American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (APA) 2014).

There are several approaches to test validation that are distinctive in terms of their respective rationales (e.g., Bachman & Palmer, 2010; Chapelle, 2021; Chapelle & Lee, 2022; Chapelle & Voss, 2021; Kane, 2022; Kunnan, 2018; Weir, 2005). However, they share some common validation themes: they all involve the gathering and evaluation of evidence of test content and its characteristics (e.g., consistency, appropriacy and adequacy of task prompts, questions or items), test-takers' processes, raters' (markers') processes, criterion-related validity, ethics, test consequences, and test fairness. The current project considers Weir's (2005) socio-cognitive framework of test validation but is informed by other validation approaches. Six types of validity evidence are investigated in the current study.

1. *Content validity* is examined within cognitive validity as it relates to how a given test task elicits sampling behaviours associated with the target construct (in this case, integrated writing ability).
2. *Criterion-related validity*: A test will have this form of validity if there is evidence of an association between integrated and independent writing performances. This validity overlaps with convergent validity, which concerns the extent of the relationship between two tests or tasks designed to test the same or a similar construct. The current study examines the relationships between performances in independent and integrated writing tasks and official IELTS scores.
3. *Cognitive validity*: This type of validity concerns how the cognitive processes required to use the language to complete an integrated writing task resemble those associated with the target construct.
4. *Context validity*: This type of validity concerns test task settings, including test rubrics, response format, and the time allowed for task completion; task demands, including representation of language use tasks and task types to be used in the test; and administrative conditions.
5. *Scoring validity*: A test exhibits scoring validity if there is evidence of scoring consistency and accuracy in raters' processes and behaviours in assigning scores to test-takers' work. This type of validity is also concerned with how well observed scores reflect test-takers' language abilities; establishing this involves test item analysis.
6. *Consequential validity*: The current study asks participants how an integrated writing task may benefit their preparedness to study in an English-medium university (e.g., washback impact). This touches on an aspect of consequential validity. The current study does not primarily aim to address consequential validity because, in this study, the test tasks are not intended for decision-making and do not have a high-stakes impact on research participants.

---

## 4 Research questions

This report focuses on answering seven research questions.

1. What is the relationship between independent and integrated writing performances?
2. Do test-takers differ in performances between independent and integrated writing tasks?
3. What are the relationships between IELTS scores and independent and integrated writing performances?
4. Do test-takers with different integrated writing success levels differ across different scoring criteria?
5. What is the nature of test-takers' reported reading, writing and test-taking strategies during independent and integrated writing tasks?
6. What are test-takers' perceptions toward the usefulness and relevance of an integrated writing task to academic study?
7. What are raters' perceptions toward an integrated writing task for assessing academic writing?

## 5 Research method

### 5.1 Research design

The present study adopts a mixed methods approach to collecting and analysing data to address the research aims and questions. Mixed methods research takes the best of both quantitative and qualitative approaches for test validation. The two approaches can provide evidence to support, complement, and expand a validation study to a more comprehensive level (see Cresswell & Plano Clark, 2018; Ivankova & Greer, 2015; Mackey & Bryfonski, 2018). The current study can be described as having a sequential design in which one data set has been collected and analysed (e.g., writing tests and questionnaires) before the collection of another dataset begins (e.g., interviews with test-takers and raters). Figure 1 summarises the design and procedures of the current study.

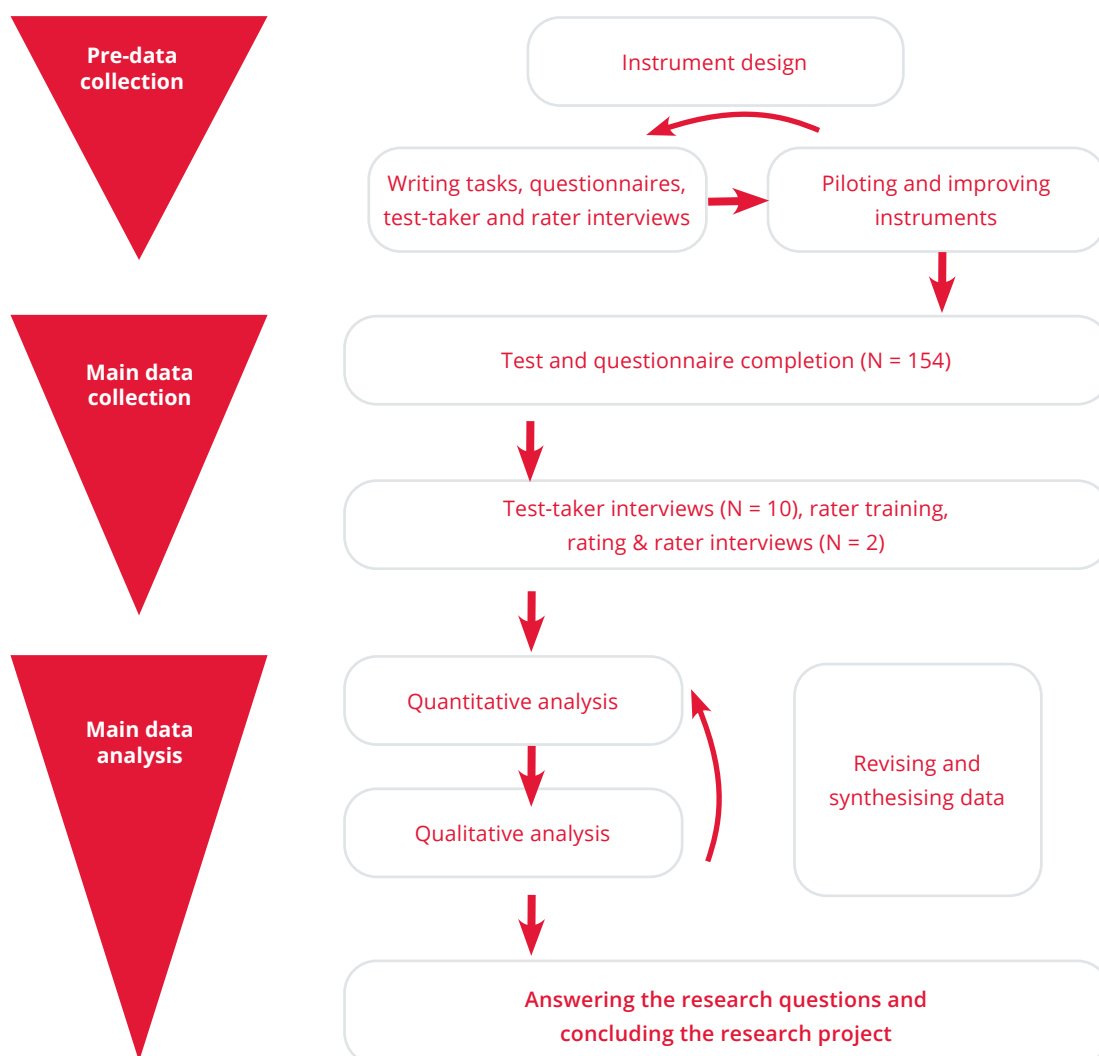
In the pre-data collection phase, the research instruments, including the two writing test tasks, reflective questionnaires for assessing test-takers' cognitive processes (labelled as strategy use), perceptions towards the integrated writing task, and interview questions for test-takers and raters, were developed and piloted. This phase was critical because the tests were to be delivered online, and the test platform needed to collect both test and questionnaire responses sequentially.

Test-takers completed an online writing test in the main data collection phase, including independent and integrated writing tasks (further discussed below). After completing the independent test task, they answered a reflective questionnaire about their writing and test-taking processes (i.e., reported strategy use). They were then prompted to complete an integrated writing task, followed by the second questionnaire about their test-taking processes and perceptions toward an integrated writing task.

Eight test-takers with three different success levels were recruited for an interview. The interview took place within one week of test-taking. The interviews aimed to collect data regarding their processes, experiences and perceptions toward the two writing test tasks. Individual semi-structured interviews utilised a stimulated recall technique by showing participants the test tasks and their writing responses. In addition, two domestic students were recruited to complete the two test tasks and questionnaires; these students also had individual interviews after they had completed the tests. The aim was to understand how they performed in the test tasks, their writing processes, and their perceptions toward integrated writing at university. After all the test data were collected, two accredited IELTS raters were recruited, had three sessions of rater training and meetings, and independently rated the test-takers' writing responses. They rated the independent writing task responses first, followed by the integrated ones.

Before the main data analysis, the reliability of the test scores and questionnaire data were statistically examined and evaluated. Quantitative data analyses, including the use of descriptive statistics, Pearson correlations, paired-samples *t*-tests, multiple regression, analysis of variance (ANOVA), multivariate analysis of variance (MANOVA), and path analysis, were carried out to address the research questions about the associations and differences between the scores for the two test types and questionnaire data. Qualitative data analyses include content analysis of test-takers' interview data, raters' comments on essay responses, and rater interview data.

**Figure 1: Research design and procedures**



---

## 5.2 Research context

The research context could be described as situated in Sydney, NSW, Australia. The participants were international ESL students studying undergraduate and postgraduate degrees and academic English pathway programs to prepare for university admission.

## 5.3 Research participants

The participants in the current study were grouped into three categories. The first category comprised those who completed the writing test tasks and questionnaires. The second included those selected to participate in a stimulated recall and individual interview (including two domestic students). The third comprised the two raters who evaluated the students' essay responses.

### 5.3.1 Test-taker participants

Initially, 215 participants agreed to take part in the study. However, 40 participants did not attempt the test tasks after registering. Twenty participants completed the first writing test task but did not subsequently complete the second task. Therefore, it was decided that only participants who completed all research instruments were included in the current study. Thus, the total number of participants in the current study was 154 (male = 72, female = 80, other gender = 2). Of the participants, 95% were Chinese (N = 147); the others were from Vietnam and Indonesia. Most participants were in the 18–25 age group (N = 134, 87%), with fewer older participants (14 were in the 26–32 age group, 9%; and four were from the 33–49 age group, 3%).

Participants reported previous overall IELTS test scores were between 4 and 8 (mean = 5.55, SD = 0.86). A large number of students had an overall IELTS band of 5 (N = 63, 41%), while fewer students scored 5.5 (N = 19, 12%) or 6 (N = 24, 16%). Thirty (30) participants had overall IELTS scores of 6.5 (N = 10, 6.5%), 7 (N = 12, 7.8%), 7.5 (n = 7, 4.5%) or 8 (N = 1, 0.6%). Two participants did not report their scores. IELTS Reading scores ranged from 4 (N = 1, 0.6%) to 8.5 (N = 3, 2%) (mean = 5.64, SD = 0.94). Fifty-six (56) participants had a score of 5 (37%), 28 participants had a score of 5.50 (18%), and 23 participants had a score of 6 (15%).

IELTS Writing scores ranged from 4 (N = 3, 2%) to 7.5 (N = 2, 1.3%) (mean = 5.46, SD = 0.73; with two students not reporting their data). Sixty-eight (68) participants reported a score of 5 (44%), and 59 participants reported a score of 5.5 (N = 23), 6 (N = 25), or 6.5 (N = 11), making up 38% of participants. Only 12 participants reported scores of 7 (N = 10) or 7.5 (N = 2), which was 7.8% of participants. It should be noted that some statistical analyses were based on 152 test-takers due to missing IELTS score reports.

### 5.3.2 Test-taker interviewees

Ten test-takers were recruited for individual interviews and a stimulated recall session. Eight interviewees were international students, and two (Lizzy and Sam) were L1 English domestic students (Australian). Their participation was voluntary. Table 1 presents demographic information for ten participants. All international interviewees were from China. The names listed are pseudonyms. The names in Table 1 were ordered by success level in their integrated writing scores (integrated success groups).

**Table 1: Demographic and writing performance information of the interviewees**

Participant	Gender	Age	Official IELTS score	Official IELTS Reading	Official IELTS Writing	Overall independent writing score of 24*	Overall integrated writing score of 30*	Time spent on independent writing task <sup>§</sup>	Time spent on integrated writing task <sup>§</sup>	No. of words in the independent writing task	No. of words in the integrated writing task
<b>Cui (Group 1)</b>	Female	18-25	6.50	6.5	6.00	18.00 (4.50)	19.00 (3.80)	40 mins	46 mins	292	261
<b>Yan (Group 1)</b>	Male	18-25	6.00	6.00	5.50	18.00 (4.50)	19.50 (3.90)	39	50	254	251
<b>Liang (Group 2)</b>	Male	26-32	7.50	8.00	6.50	19.75 (4.94)	21.50 (4.30)	34	42	384	321
<b>Zhao (Group 2)</b>	Male	18-25	6.00	5.50	5.50	18.00 (4.50)	21.75 (4.35)	40	50	382	329
<b>Lu (Group 3)</b>	Female	18-25	6.50	6.5	6.00	18.25 (4.56)	22.25 (4.45)	39	50	402	401
<b>Teng (Group 3)</b>	Male	18-25	7.00	7.0	7.00	21.25 (5.31)	22.50 (4.50)	40	50	418	374
<b>Ren (Group 3)</b>	Female	33-49	7.00	7.0	7.00	20.00 (5.00)	23.00 (4.60)	38	49	419	364
<b>Wang (Group 3)</b>	Male	26-32	7.00	7.5	7.50	20.50 (5.13)	25.50 (5.10)	40	50	447	418
<b>Lizzy (Group 3)</b>	Female	26-32	N/A	N/A	N/A	21.75 (5.44)	23.00 (4.60)	37	26	668	445
<b>Sam (Group 3)</b>	Male	33-49	N/A	N/A	N/A	21.75 (5.44)	26.00 (5.20)	32	38	372	356

Notes: \* on a 6-point scale. <sup>§</sup>in minutes.

### 5.3.3 Raters

IDP Australia provided a list of highly recommended IELTS examiners from which one male and one female examiner were selected and employed to rate the essays. For anonymity, these are referred to as Rater 1 and Rater 2. Both raters were native speakers of English and spoke more than two languages. Rater 1 was male and had worked as an IELTS examiner for about 12 years (at the time of data collection). He had a degree in linguistics and a Master of Teaching and several years of English teaching experience. Rater 2 was female, had similar qualifications to Rater 1, and had worked as an IELTS examiner for about 10 years (at the time of data collection). Rater 2 also had extensive experience teaching English in different countries.

---

## 5.4 Research instruments

The current study employs the following five instruments which are described below.

1. Two essay tasks: one independent writing task and one integrated writing task
2. Two Likert-type scale questionnaires for assessing participants' test-taking processes and perceptions toward the essay tasks;
3. Two analytic scoring rubrics: one for the independent writing task and one for the integrated writing task
4. Semi-structured interview questions (with a stimulated recall technique) for selected test-takers
5. Semi-structured interview questions for the two raters.

### 5.4.1 An overview of the IELTS Writing test

An online, simulated IELTS Writing test was developed. The writing test could be accessed via a secure test website. The test website was created and tested for its functionality, robustness, appearance, and data retrieval efficiency (see Appendix 1 for a screenshot of the test outlook). The apparent technical issues were addressed and resolved before using the test in the current project. The test was designed to collect test-takers' writing task responses and questionnaire data in the same session. The data could be viewed on screen and retrieved in the CVS format. To take the test, each participant was required to complete an online registration form, which the researcher or an administrator then checked. On approval of a participant's registration form, an account was created for that participant. That participant was then able to sign in to complete the test tasks. Participants were asked to register at least one day before taking the test.

### 5.4.2 IELTS Independent Writing Task

One official Cambridge IELTS Writing Task 2 set was chosen for the current study (see Appendix 2). The topic was public health and sports facilities. This task requires test-takers to write at least 250 words within 40 minutes. Automatic spelling checkers and autocorrection tools were disabled. Once they had clicked the "Start now" button, the clock provided at the top right of the screen started counting down. Test-takers wrote their essays in the space provided. An ongoing word count was also provided at the bottom of the screen. Test-takers could revise and submit their essays at any time within 40 minutes. Once 40 minutes had been reached, test-takers could no longer write in the space provided and were prompted to submit their essays. Therefore, the time spent on this task varied from test-taker to test-taker but was limited to 40 minutes. Time data were reported in the findings section. After submitting an essay, each test-taker was asked to complete a reflective questionnaire about this writing task (see Reflective Questionnaire 1 below).

### 5.4.3 IELTS Integrated Writing Task

Several considerations were made about the integrated writing task, including the topics, text types, and the lengths and number of academic reading texts to be used. The integrated writing task by Gebril and Plakans (2013) was adopted in the current study. The topic of this writing task was global warming, which was timely and intellectually stimulating. It was also decided that the ESL international students should be familiar with the topic but should not require excessive background knowledge to write well on it. This task has been used and validated by Gebril and Plakans. This task includes two short texts for test-takers to read and use in their writing. The two texts were considered meaningful, not highly technical, yet sufficiently detailed to be used to support writing (see Appendix 3).

It was decided that test-takers should be allowed 50 minutes to complete this task as they needed to read the texts and plan and write their essays. Test-takers were recommended to spend no more than 10 minutes reading and planning their essays. It was anticipated that most participants would not be familiar with this task type, so giving them less than 50 minutes to complete the task could be overly challenging cognitively and would make them rush to complete it. IDP Australia was consulted regarding this decision to allow 50 minutes to complete this task. Moreover, a pilot study suggested that 50 minutes was sufficient for an average-ability writer to complete this task. In this test task, the two texts could be accessed throughout the 50-minute period, and test-takers could use their mouse to copy the text for their writing convenience. Test-takers could move up or down to read the texts. As with the independent writing task, test-takers could see a clock showing the time remaining to complete the task and an ongoing word count. Once test-takers commenced this task, they could submit their essays at any time within 50 minutes. When 50 minutes had passed, they could no longer write in the provided space and were prompted to submit their essays. They were then asked to complete a reflective questionnaire about this task (see Reflective Questionnaire 2), which ended the test administration.

#### 5.4.4 Reflective Questionnaire 1

To gather information about test-takers' writing processes and test-taking strategies in the independent writing task, a 5-point Likert-type scale questionnaire was developed (see Appendix 4). Due to considerations of the time and cognitive effort the test-takers put into writing, the questionnaire needed to be brief and focused on a few key aspects. The questionnaire collected some background information from the participants, including their previous IELTS scores. There was a total of 14 items in this questionnaire. Participants were asked to choose 1 (strongly disagree) to 5 (strongly agree) for each statement. Table 2 shows the taxonomy of this questionnaire.

**Table 2:** Taxonomy of a reflective questionnaire for the independent writing task

Part	Construct	Explanation
A	Writing Strategies Items 1 to 9	One's conscious perception of writing processes. Writing strategies include planning what to include in a text, organising paragraphs, and completing an essay task.
B	General Test-taking Strategies Items 11 to 14	One's conscious perception of the overall processes of completing the given writing task. General test-taking strategies include being aware of time limitations to complete the task, self-monitoring and evaluation.

#### 5.4.5 Reflective Questionnaire 2

There were four sections to this questionnaire for capturing test-takers' writing processes and test-taking strategies in the integrated writing task (Appendix 5). The first concerned test-takers' reading strategies; the second was regarding their integrated writing strategies; the third was about their general test-taking strategies; and the fourth was about their perceptions of an integrated writing task. Table 3 presents the taxonomy of this questionnaire.

**Table 3:** Taxonomy of a reflective questionnaire for the integrated writing task

Part	Construct	Explanation
A	Reading Strategies Items 1 to 5	One's conscious perception of reading processes relevant to the test task. Reading strategies involve identifying key points, selecting information relevant to the assigned tasks, and using prior knowledge to assist reading comprehension.
B	Integrated Writing Strategies Items 6 to 10	One's conscious perception of discourse synthesis writing processes. Writing strategies include planning what to include in a text, paraphrasing and acknowledging source information or borrowed ideas.
C	General Test-taking Strategies Items 11 to 15	One's conscious perception of the overall processes of completing the given writing task. General test-taking strategies include being aware of time limitations to complete the task, self-monitoring, and evaluation.
D	Evaluation of an Integrated Writing Task Items 16-19	One's perceptions about an integrated writing task in terms of authenticity, accuracy, and usefulness.

#### 5.4.6 Semi-structured interview questions for test-takers

Ten participants were recruited for semi-structured, individual interviews. Interview data can provide additional perspectives about participants' test-taking processes during the two writing tasks to those collected via the questionnaire. The interviews were conducted via Zoom and lasted between 25 and 42 minutes. The interviews focused on four main issues (see Appendix 6):

1. test-takers' general background, IELTS test-taking experiences, and general perceptions about integrated writing
2. the processes they used while completing the independent writing task
3. the processes they used in completing the integrated writing task
4. their beliefs, attitudes and perceptions toward an integrated writing task for the IELTS test.

In relation to issues 2 and 3, participants were shown the writing tasks and their responses to help them remember the tasks and their written responses. When this stimulated recall approach is taken, prompts or clues (e.g., written texts, test scores and comments) are provided to facilitate participants' recollection of their thought processes whilst taking the test. The interview questions were trialled and improved before their use in the main study.

#### 5.4.7 Semi-structured interviews for raters

After the raters completed their assessment of the writing responses for both writing tasks, individual interviews with them took place via Zoom. Each lasted about 60 minutes. A stimulated recall strategy, similar to how it was applied with test-takers (e.g., test tasks and response screenshots), was also used to help the raters recall their marking and some writing responses. Each of the interviews had three main parts (see Appendix 7). The first part enquired about participants' general background, rating experiences and perceptions about rater training and meetings. The second part focused on their experiences and issues in marking the independent writing task (e.g., characteristics of good writing responses and their challenges). The third part was about their experiences and issues when marking the integrated writing task. This section also asked them to outline the assessment strategies they used in marking the integrated writing task and their views about including an integrated writing task in the IELTS Writing test.

---

#### 5.4.8 Analytic scoring rubric for the independent writing task

The current study decided that analytic rating scales would be used to evaluate the participants' writing (see Appendix 8). This is a key distinction between the scoring methods used in the current study and the actual standardised IELTS. Therefore, the scores used in the current study may not be directly comparable to those provided in the IELTS Writing, which are derived from holistic scoring. Nonetheless, aspects of the scoring criteria were based on the IELTS scoring rubrics (public version) for the independent writing scales. The rubrics are a 6-point analytic scale that measures four criteria, and raters can assign half-band scores.

1. Task achievement (e.g., addressing the task and providing supporting ideas)
2. Coherence and cohesion (e.g., demonstrating organisation and progression of ideas and using cohesive devices)
3. Lexical resource (e.g., using appropriate, varied vocabulary; showing control of lexical features)
4. Grammatical range and accuracy (e.g., using a range of grammatical structures with flexibility and accuracy).

The scoring rubric was reviewed by a language testing expert, who also provided feedback for improvement. It was also piloted using some writing responses. The two raters also provided further feedback for the rubric refinement during the three rater training sessions. Some original rubrics were adjusted to clarify content, expressions, and meanings.

#### 5.4.9 Analytic scoring rubric for the integrated writing task

The analytic scoring rubric for the integrated writing task was based on the independent writing task. Five criteria were used to score the responses to this writing task (see Appendix 9). Four aspects were identical to those used for scoring the independent writing task. However, the 'source texts' aspect was added to the integrated writing rubric. This criterion was adapted from the TOEFL iBT integrated writing rubric. Still, information about text and discourse synthesis discussed in the literature was further incorporated into the rubrics (e.g., Plakans et al. 2019). The scoring rubric was also piloted before actual use in the current study. Feedback from the two raters was used to refine the rubric for the main study to improve the scales' clarity and to underline the differences between them.

## 5.5 Data collection

The data collection procedure followed in the current study is described in the following sections.

### 5.5.1 Ethical considerations

Ethics approval was sought from and approved by the Human Resource Ethics Committee (The University of Sydney) before data collection (#2020/805). A course coordinator or director informed potential test-taker participants about this research project. They were sent a flyer about the project and a participant information statement explaining the project and what they would have to do if they participated. Participants were informed that they should contact the researcher via email if they were interested in participating in the project. Before data collection, a group meeting was organised to explain the project's purpose and procedures. This session allowed participants to ask any questions they wanted and allowed them to provide their consent to participate in the study. All data were treated in a way that protected the privacy and anonymity of the participants involved in the study.

---

### 5.5.2 Key stages in data collection

The four main stages in the data collection procedure are summarised as follows.

#### Stage 1: Test-taking

Participants met with the researchers before taking the test. They were provided information about the test tasks they would need to complete, and it was verified that they had already registered to take the test. They were made to understand that they could take notes and write an outline during the test, but they would not be allowed to use any online resources (including dictionaries). They were asked not to discuss the test tasks with their peers participating in this project. It was stressed that this test would not be used to make high-stakes decisions on them and that the results would have no consequences for their current study or program. Participants were encouraged to do their best in the writing tasks. All participants acknowledged their understanding of these points. The spell-check feature of the internet browser was disabled to ensure that test-takers were not prompted with spelling correction suggestions. Participants were shown how to disable the spell-check feature in their browser before the test. All participants were asked if they had any questions before starting the test.

In the current study, the scores from the two raters were averaged in each writing task to form a composite of each test-taker's score. Scores derived from both raters' scores provide richer information about their writing performance than those based on a single rater. The scores were also re-scaled to range between 1 and 6, as reflected in the scoring rubrics, rather than a sum of all awarded scores from each assessment criterion.

#### Stage 2: Test-taker interviews

Test-taker participants were informed of the possibility of being interviewed about their test-taking experiences. They were asked to contact the researcher via email if they were interested in participating in this. In this report, a total of eight international students and two domestic students contributed to the interview data. The domestic students also took the two test tasks and completed the questionnaires before the interviews.

#### Stage 3: Rater training and rating

Two qualified IELTS raters were recruited and employed. They met with the researcher three times (1 hour each time in 3 consecutive weeks) to discuss the scoring procedures and scoring rubrics and to grade some sample independent and integrated essays. As the raters were experienced IELTS raters, discussing the rating procedure did not take long. In the current study, two issues need to be noted. First, the scoring rubrics were analytic, so the raters were asked to assign a score to each assessment category and write a brief comment about each essay. This requirement differed from what they had to do when marking an official IELTS Writing test. Second, initially, Turnitin was to be used in the integrated writing task to help raters conveniently identify text that test-takers had taken directly from the reading texts (indicating possible plagiarism or lack of acknowledgement of source texts). However, a pilot study suggested that text similarity identification by Turnitin was distracting and tended to prompt some rater bias or adverse reactions toward a given essay. It was decided that this Turnitin feature would not be used in the current study.

The first meeting focused on explaining the purpose of the study and providing rating orientation to the raters. The raters were then asked to score some independent writing responses. Their ratings were compared, and their opinions were discussed. Any scoring discrepancy was discussed and resolved. The second meeting focused on the rating of integrated writing responses. The raters were asked to complete the integrated writing task before the second meeting so they would have direct experience in completing the task. The third meeting revisited the key issues in the rating of both tasks.

---

Marking schedules and a timeline to complete the marking were discussed and agreed. The raters were asked to complete their rating of the independent writing task responses and only then rate the responses to the integrated writing task. It was recommended that they rate no more than 15 responses in a sitting. In summary, raters provided a score for each aspect of writing and a comment (optional) about each essay. Regarding their comments on the integrated essay task, they were asked to observe source text use while marking the responses.

In the current study, whenever possible, the raters were asked to add a comment for each integrated written response after they had rated it using the scoring rubric. In official IELTS ratings, raters generally do not comment on test-takers' responses, so the current study did not reflect a common IELTS Writing examination rating practice. Raters 1 and 2 provided approximate totals of 3,400 and 3,000 words for all test-takers, respectively. While scoring each test response, Raters 1 and 2 highlighted some texts in test-takers' responses related to writing issues in their comments (e.g., direct copy without acknowledgements). That is, such text highlights were aligned with their comments. Sometimes, the raters used coloured texts to remind themselves when scoring.

#### Stage 4: Rater interviews

Each rater was interviewed separately to discuss their viewpoints and experiences in rating responses to the independent and integrated writing tasks.

## 5.6 Data preparation and analyses

### 5.6.1 Quantitative data analysis

Several standard statistical analyses, such as descriptive statistics, rater agreement and reliability analysis, interclass correlations, and inferential statistics, such as Pearson correlation, paired-samples *t*-tests, multiple regression analysis, and path analysis, were used to address the research questions.

*Descriptive statistics:* The current study used measures of central tendency such as mean, median and mode, and measures of dispersion such as the standard deviation, skewness, and kurtosis statistics with the test and questionnaire data. Frequent counts, percentages, and minimum and maximum observed scores were also used.

*Pearson correlations:* The Pearson product moment correlation is an approach for examining the linear statistical relationship between two continuous variables. When a relationship is detected, the strength and direction of the relationship can be determined. The current study used Pearson correlations to explore the relationship between the independent and integrated writing scores and questionnaire responses.

*Paired-samples t-tests:* The paired-samples *t*-test can be useful to examine differences between attributes and performances within the same individuals. The current study compared test-takers' independent and integrated writing task scores to determine whether the two performances differed.

*Path analysis:* Due to a smaller sample size than anticipated, a path analysis was performed to explore the direct and indirect relationships between variables such as reported official IELTS test scores and independent and integrated writing scores. EQS program was used for testing and modelling hypothesised relationships among observed variables. The stages involved in the current study included initial model specification for the interconnections among latent variables; model identification for assessing the degree of freedom for parameter estimates; model estimation in which the maximum likelihood (ML) was used for estimation; model fit assessment for evaluating whether and the extent to which the dataset could explain the model (e.g.,

---

using statistics such as comparative fit index (CFI); and root-mean-squared error of approximation (RMSEA)) and model respecification and modification in which the original model was revised and retested (see Phakiti, 2007, 2016, 2018; Schumacker & Lomax, 2015).

### 5.6.2 Qualitative data analysis

Miles et al.'s (2014) data analysis framework was adopted mainly as a general qualitative approach to the study. This framework involves three cyclical activities: data reduction, data display, and drawing and verifying conclusions. Content analysis (e.g., Schreier, 2012, 2014) is one of the oldest approaches to analysing qualitative data. It aims to describe patterns and meanings in qualitative data. Methodologists take different approaches to content analysis and propose various stages of content analysis (e.g., Schreier, 2012, 2014). For example, content analysis driven by a theory that informs a qualitative study can lead to particular methods for developing coding schemes and data interpretations. The key stages in content analysis include:

1. the development of a coding scheme or framework
2. the refinement and expansion of codes through pilot coding
3. the development of rules for systematic coding and consistency in coding data
4. coding data to reduce the data and segment them into cohesive categories or groups
5. the analysis and interpretation of patterns, using theory and content as evidence to illustrate common patterns or unique cases or findings that deviate from the norm.

The current study used content analysis to gather evidence of test content, test-taking processes, and test usefulness. Following Gebрил and Plakans' (2014) qualitative analysis procedures, interview data were transcribed and organised based on the different groups of participants. A coding scheme (e.g., as used in Gebрил & Plakans 2014) was considered and applied to the qualitative data. It should be noted that the focus of the qualitative analysis in the current study was on *meaning-making and understanding test-takers' and raters' processes and perceptions* rather than on exhibiting consistency in applying codes and coding. Such meaning-making is naturally subjective; therefore, an attempt to obtain high coder reliability or agreement may be a distraction and inhibit the ability to make meanings from qualitative data. However, intercoder agreements ranged between 90% and 95% in interview data.

#### *Content analysis of test-takers' interview data*

Since the quantitative analyses also examined differences among test-takers in relation to the integrated writing task, the data were organised according to the participants' integrated writing success levels (discussed in the finding and discussion section). The interviewees were grouped into three success levels, with the largest group being Level 3 (the two domestic students were at Level 3). They are summarised as follows:

- Level 1 (scores between 1.4 and 3.99 inclusive): Cui and Yan
- Level 2 (scores between 4.00 and 4.39 inclusive): Liang and Zhao
- Level 3 (scores between 4.40 and 5.50 inclusive): Lu, Teng, Ren, Wang, Lizzy and Sam.

The interviews were transcribed and checked for accuracy. This was a relatively long process as the original automated transcription was not very accurate, largely due to issues in pronunciation and accents. The average word count of the transcripts was 4,262 words (SD = 991), with the shortest script being 2,451 words (Cui) and the longest 6,095 words (Teng). The transcripts were read several times by the researcher, who also interviewed the participants. This process was crucial for gaining insights into participants' writing processes and perceptions toward reading-to-write essays.

---

In the current study, the qualitative analysis followed Miles et al. (2014), which allowed the researcher to go through several coding cycles flexibly. In the first preliminary coding cycle, an 'in Vivo' coding method, in which the languages of the interviewees were used as initial codes, was used. Later, together with the key aspects that shaped the interview questions, a thematic coding scheme was developed for coding and refining the in Vivo codes (see Appendix 10). The current study also adopted a narrative inquiry approach in which the voices of the research participants were used to authentically reflect their own perspectives and stories (see, e.g., Barkhuizen, 2015). It should be noted that test-takers' quotes were subject to a slight language edit for clarity of intended expressions.

It should be noted that iterative and refining coding processes also occurred during the actual coding and writing processes since it was natural that the interviewees did not produce the exact or similar responses to the same interview questions (Miles et al., 2014). Furthermore, follow-up questions were employed to clarify some expressions and to elicit more information, thereby arriving at new unanticipated details. It is important to stress that the researcher only used the coding schemes to guide the analysis of the interview data rather than attempting to replicate coding in a strictly consistent manner to produce evidence of coding reliability. In qualitative research, data analysis is more dynamic than quantitative analysis because of the subjective nature of data interpretation. This report presents and interprets the analysis results using direct quotes from the participants and data matrices. Using direct information from the participants allows us to gain insights into their writing and thought processes and their understanding of the integrated essay task. It also allows readers to relate the researcher's interpretations to participants' data and interpret the information independently; such interpretations may differ from the researcher's.

#### *Content analysis of features of integrated writing responses*

Test-takers' integrated performance success levels were adopted to sort raters' comments. Three success groups were formed:

- scores between 1.4 and 3.99 inclusive = Group 1 (n = 43)
- scores between 4.00 and 4.39 inclusive = Group 2 (n = 57)
- scores between 4.40 and 5.50 inclusive = Group 3 (n = 54).

An individual rater's comments were compiled in an Excel sheet, first in order of test-taker ID; then, a column on test-taker success level was added. Next, this file was sorted in order of success level. Rater 1 provided all 154 comments (100%), whereas Rater 2 provided 124 comments (80.5%). Then, the data were exported to a Word document for coding and data analysis.

The current study adopted the approach taken by Miles et al. (2014) for qualitative data analysis. The analysis processes used included first- and second-cycle codes and coding (e.g., in the first cycle, codes were assigned to chunks of data; in the second coding, a refined coding scheme was reapplied to the data (see Appendix 11 for the coding scheme). Initially, the preliminary coding scheme was based on an 'in Vivo' coding method in which the rater's language in the data record was used as codes (Miles et al., 2014). This first cycle coding scheme was then examined with the rating scale for the specific terminology for categorising the data. This process allowed an application of evaluation coding that led to a refined and more selective coding scheme to understand raters' comments for different success levels.

In the second cycle, this qualitative analysis required an iterative process of referring back to the codes and the actual data, which can lead to some changes in the grouping or the codes. Qualitative data analysis based on Miles et al. (2014) requires researchers to be open and flexible to new, unanticipated information or ideas as they delve into the data rather than restricting themselves to only looking for targeted or predetermined

constructs. In the current study, the qualitative data analysis does not aim to generalise patterns but to organise and make sense of written comments made by the raters. The focus of the qualitative data analysis in the current study is on understanding raters' perspectives in the assessment of integrated writing responses. Nonetheless, the overall high level of intercoder agreements (93%) might be explained by the fact that each rater's comments on integrated writing responses were relatively brief.

### Content analysis of raters' interviews

The interviews were transcribed and double-checked for content accuracy. Each transcript (approximately 7,700 and 7000 words for Raters 1 and 2, respectively) was read several times. Then, they were coded regarding the characteristics of essays, source text use, rater training, difficulty and challenges in rating integrated essays, and feasibility of including an integrated writing task in an official IELTS test (see Appendix 12 for the coding schemes). Quotes from the raters were used comprehensively to authenticate their viewpoints and positions. This allows the reader to review and verify any inferences the researcher made in relation to the issues being addressed. Some quotes from the raters were slightly edited for brevity and clarity of expressions.

## 5.7 Reliability of quantitative data

### 5.7.1 Score reliability

The scores used in the study must be reliable. For each test-taker, an average score from both raters was used for quantitative analysis. Since the study employs human raters to evaluate test-takers' writing performance, scores derived from the raters are examined. Four measures are used to assess inter-rater reliability; the results provide different information about the scores.

1. An *adjacent agreement* determines whether the two raters exhibit significant score discrepancies. The percentage of scores differing by no more than one point is calculated. The current study calculated and evaluated an adjacent agreement for each independent and integrated writing assessment criterion. The difference between the scores by each rater was first computed to compute the adjacent agreements. Then, the percentage of agreements was calculated. The adjacent agreements ranged from 99.3 to 100, suggesting a strong agreement between the two raters across different criteria in both writing tasks (Table 4).

**Table 4:** *Adjacent agreements between the two raters (accompanied by exact agreements) (N = 154)*

Scoring criteria	Adjacent agreement	Exact agreement
Task 1 Task achievement	100.000	63.000
Task 1 Cohesion and coherence	100.000	68.000
Task 1 Lexical resource	100.000	50.000
Task 1 Grammatical range and accuracy	100.000	74.000
Task 2 Task achievement	99.300	56.000
Task 2 Source text	99.300	51.000
Task 2 Cohesion and coherence	99.300	74.000
Task 2 Lexical resource	99.300	57.000
Task 2 Grammatical range and accuracy	99.300	72.000

Note: Task 1 = Independent writing

Task 2 = Integrated writing

2. *Cronbach's alpha coefficients* are used to examine the internal consistency of each rater's scores. This information is useful to understand which aspects were easy or difficult for the two raters to agree on. Table 5 presents the Cronbach's alpha coefficients for each rater. The coefficients range from 0.65 (Task 2 for Rater 2) to 0.85 (Task 1 for Rater 1). The Cronbach's alpha coefficients for the integrated task were lower than Task 1 for both raters. An inspection of 'scale if item deleted' suggests that the lower alpha coefficients were influenced by the raters' scoring of the source text. Scores of Source text reduced the Cronbach's alpha coefficient, implying that the ratings of Source text were not necessarily linearly related to other assessment categories. Test-takers might have received higher scores for other criteria but lower for the Source text. It should be noted that consistent with the adjacent agreement, the correlation coefficient of the source text use between the two raters was 0.904, implying that the raters strongly agreed with each other in terms of the ratings.

**Table 5: Cronbach's alpha coefficients (N = 154)**

	Cronbach's Alpha	N of Items
Rater 1 Independent task	0.853	4 (task achievement, coherence and cohesion, lexical resource, and grammatical range and accuracy)
Rater 2 Independent task	0.860	4 (same as above)
Rater 1 Integrated task	0.739	5 (task achievement, source text, coherence and cohesion, lexical resource, and grammatical range and accuracy)
Rater 2 Integrated task	0.650	5 (same as above)

Note: Task 1 = Independent writing

Task 2 = Integrated writing

3. *Correlation coefficients* examine the scoring consistency between the two raters. Correlations were performed on both individual scoring aspects and total scores. Table 6 presents Pearson correlation coefficients, with 95% confidence intervals, for each scoring criterion between the two raters in the independent writing task. The coefficients ranged from 0.598 (Lexical resource) to 0.740 (Cohesion and coherence). The coefficients were considered large effect sizes in terms of the strength of the association.

**Table 6: Pearson correlation coefficients between the two raters' scores for the independent writing task for the different criteria (N = 154)**

	Correlation	Lower 95% C.I.	Upper 95% C.I.
Task achievement	0.625**	0.519	0.713
Cohesion and coherence	0.740**	0.659	0.804
Lexical resource	0.598**	0.485	0.691
Grammatical range and accuracy	0.650**	0.548	0.733

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

Table 7 presents Pearson correlation coefficients between the two raters' scores for the integrated writing task for each scoring criterion. The correlation coefficients ranged from 0.69 (Task achievement) to 0.90 (Source text) and were higher in the integrated writing task than in the independent writing task.

**Table 7: Pearson correlation coefficients between the two raters' scores for the independent writing task for each scoring criterion (N = 154)**

	Correlation	Lower 95% C.I.	Upper 95% C.I.
Task achievement	0.688**	0.595	0.763
Source text	0.904**	0.870	0.929
Cohesion and coherence	0.773**	0.700	0.830
Lexical resource	0.760**	0.683	0.819
Grammatical range and accuracy	0.759**	0.683	0.819

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

4. *Interclass correlation coefficients* are used to examine the extent to which the two raters agree on which test-takers deserve a high rating and which deserve a low one, etc. Tables 8 and 9 present the interclass correlation coefficients for each writing task. The average absolute agreement between the two raters was 0.838 for Task 1 and 0.856 for Task 2. It should be noted that the interclass correlation coefficients show absolute agreements between the two raters.

**Table 8: Interclass correlation coefficients for the independent writing task (N = 152)**

	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	0.721 <sup>a</sup>	0.635	0.789	6.167	151	151	<0.000
Average Measures	<b>0.838<sup>c</sup></b>	0.777	0.882	6.167	151	151	<0.000

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.

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.

**Table 9: Interclass correlation coefficients for the integrated writing task (N = 152)**

	Intraclass Correlation <sup>b</sup>	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	0.749 <sup>a</sup>	0.670	0.811	6.961	151	151	<0.000
Average Measures	<b>0.856<sup>c</sup></b>	0.802	0.896	6.961	151	151	<0.000

In the two-way mixed effects model, people effects are random, and measures effects are fixed.

a. The estimator is the same whether the interaction effect is present.

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.

In summary, item-level and statistical analyses of score reliability have indicated that the agreements and reliability estimates are strong, suggesting that the writing scores used to address the research questions are suitable for data analysis for answering the research questions.

### 5.7.2 Questionnaire reliability

Table 10 presents the Cronbach's alpha coefficients of the Reflective Questionnaire 1. The overall Cronbach's alpha coefficient was 0.88. The Cronbach's alpha coefficients for Items 1 to 9 (Writing strategies) and 10 to 14 (Test-taking strategies) were 0.799 and 0.806 respectively.

**Table 10:** Cronbach's alpha coefficients of the reflective questionnaire on the independent writing task (N = 154)

Aspects	Items	Cronbach's alpha coefficient
Writing strategies	1 to 9	0.799
Test-taking strategies	10 to 14	0.806
Overall	1 to 14	0.880

Table 11 presents the Cronbach's alpha coefficients of the Reflective Questionnaire 2. The overall Cronbach's alpha of this questionnaire was 0.918. The Cronbach's alpha coefficients for Items 1 to 5 (Reading strategies), 6 to 10 (Integrated writing strategies), 11 to 15 (Test-taking strategies) and 16 to 19 (Evaluation of an integrated writing task) were 0.746, 0.819, 0.816, and 0.664 respectively. It should be noted that items 16 to 19 were designed to measure different areas of participants' beliefs or perceptions about the integrated writing task.

**Table 11:** Cronbach's alpha coefficients of the reflective questionnaire on the integrated writing task (N = 154)

Aspects	Items	Cronbach's alpha coefficient
Reading strategies	1 to 5	0.746
Integrated writing strategies	6 to 10	0.819
Test-taking strategies	11 to 15	0.816
Evaluation of the integrated writing task	16 to 19	0.664
Overall	1 to 19	0.918

In summary, the reliability estimates were acceptable for the data to be used for subsequent analysis to address the research questions.

## 6 Findings and discussion

This section presents and discusses answers to the research questions.

### 6.1 Research question 1: What is the relationship between integrated and independent writing performance?

This research question is first achieved by exploring and interpreting the correlations between integrated and independent writing scores (addressing content, criterion-related, and scoring validity). Second, content analysis of raters' interviews is used to examine the connections between independent and integrated writing performance.

#### 6.1.1 Quantitative findings

Table 12 presents the descriptive statistics of the test-takers' independent writing task scores based on the two raters' scores. The total mean score for the independent writing task was 4.55 (SD = 0.29), and the mean scores for the different scoring criteria ranged from 4.48 (for Cohesion and coherence, SD = 0.39) to 4.63 (for Lexical resource, SD = 0.30). Table 13 presents the descriptive statistics of the test-takers' integrated writing task scores based on the two raters' scores. The total mean score for the integrated writing task was around 4.15 (SD = 0.45). The mean scores of the different assessment criteria ranged from 3.24 (for Source text, SD = 1.19) to 4.48 (for Lexical resource, SD = 0.44).

**Table 12:** Descriptive statistics of test-takers' independent writing scores based on two raters (N = 154)

	Task achievement	Cohesion and coherence	Lexical resource	Grammatical range and accuracy	Total score
Mean	4.502	4.481	4.625	4.584	4.548
Median	4.500	4.500	4.500	4.500	4.500
Mode	4.500	4.500	4.500	4.500	4.500
SD	0.346	0.387	0.295	0.308	0.291
Minimum	3.000	3.500	3.500	3.500	3.438
Maximum	5.500	5.500	5.500	5.500	5.438

**Table 13:** Descriptive statistics of test-takers' integrated writing task scores based on the two raters' scores (N = 154)

	Task achievement	Source text	Cohesion and coherence	Lexical resource	Grammatical range and accuracy	Total score
Mean	4.304	3.239	4.304	4.482	4.422	4.150
Median	4.500	3.750	4.500	4.500	4.500	4.200
Mode	4.500	4.000	4.500	4.500	4.500	4.400
SD	0.479	1.187	.5161	0.443	0.427	0.454
Minimum	1.750	1.000	1.250	1.500	1.500	1.400
Maximum	5.250	5.000	5.500	5.250	5.250	5.200

A Pearson correlation was performed to explore the correlation between the independent and integrated writing scores. The data were normally distributed. Table 14 presents the Pearson correlation coefficients, shared variances ( $R^2$ ) and  $p$ -values. The positive correlation coefficients were found to be moderate across the evaluation criteria. The only non-significant correlation was found between the independent task achievement and the integrated source text use ( $p = 0.259$ ). The effect sizes of the coefficients between all independent writing criteria and source text use were small. This implies that performance in the independent writing task could not predict how well test-takers could use source texts in their integrated writing. It was commonly found that independent lexical resource and grammatical range and accuracy measures had the strongest correlation coefficients with other integrated writing measures (e.g.,  $r = 0.578$  between independent lexical resource and integrated grammatical range and accuracy). Of most interest is the finding that the total independent score was significantly related to the total integrated score ( $r = 0.505$ ,  $R^2 = 0.26$ ,  $p < 0.01$ ). This, however, indicated a moderate effect size, suggesting that approximately 26% of the independent writing score variance accounted for that of the integrated writing performance. It implies that the other 74% of the integrated writing performance needs to be explained. The finding may mean that the two tasks measure two different, though relevant, academic writing abilities. It may also imply that test-takers who are good at independent essay tasks may still struggle to write integrated essay tasks successfully.

**Table 14:** Pearson Correlations between independent and integrated writing scores ( $N = 154$ )

		Integrated task achievement	Integrated source text	Integrated cohesion and coherence	Integrated lexical resource	Integrated grammatical range and accuracy	Total integrated score
Independent task achievement	Pearson correlation	0.423** $R^2 = 0.179$	0.092 $R^2 = 0.008$	0.411** $R^2 = 0.169$	0.389** $R^2 = 0.151$	0.424** $R^2 = 0.180$	0.386** $R^2 = 0.149$
	Sig. (2-tailed)	<.001	.259	<.001	<.001	<.001	<.001
Independent cohesion and coherence	Pearson correlation	0.367** $R^2 = 0.135$	0.175* $R^2 = 0.030$	0.424** $R^2 = 0.180$	0.336** $R^2 = 0.113$	0.359** $R^2 = 0.129$	0.398** $R^2 = 0.158$
	Sig. (2-tailed)	<.001	.030	<.001	<.001	<.001	<.001
Independent lexical resource	Pearson correlation	0.522** $R^2 = 0.272$	0.170* $R^2 = 0.029$	0.431** $R^2 = 0.186$	0.534** $R^2 = 0.285$	<b>0.578**</b> <b><math>R^2 = 0.334</math></b>	0.509** $R^2 = 0.259$
	Sig. (2-tailed)	<.001	.035	<.001	<.001	<.001	<.001
Independent grammatical range and accuracy	Pearson correlation	0.481** $R^2 = 0.231$	0.177* $R^2 = 0.031$	0.413** $R^2 = 0.171$	0.511** $R^2 = 0.261$	0.544** $R^2 = 0.296$	0.489** $R^2 = 0.239$
	Sig. (2-tailed)	<.001	.028	<.001	<.001	<.001	<.001
Total independent score	Pearson correlation	0.507** $R^2 = 0.257$	0.175* $R^2 = 0.031$	0.482** $R^2 = 0.183$	0.498** $R^2 = 0.239$	0.535** $R^2 = 0.286$	<b>0.505**</b> <b><math>R^2 = 0.255</math></b>
	Sig. (2-tailed)	<.001	.030	<.001	<.001	<.001	<.001

\*\* . Correlation is significant at the 0.01 level (2-tailed). \* . Correlation is significant at the 0.05 level (2-tailed).

**Table 15: Pearson Correlations between the five criteria and total integrated writing scores (N = 154)**

		Integrated source text	Integrated cohesion and coherence	Integrated lexical resource	Integrated grammatical range and accuracy	Total integrated score
Integrated task achievement	Pearson Correlation	0.116 R <sup>2</sup> = 0.013	0.768** R <sup>2</sup> = 0.590	0.863** R <sup>2</sup> = 0.745	0.857** R <sup>2</sup> = 0.734	0.775** R <sup>2</sup> = 0.600
	Sig. (2-tailed)	.152	<.001	<.001	<.001	<.001
Integrated source text	Pearson Correlation	1	0.254** R <sup>2</sup> = 0.065	0.119 R <sup>2</sup> = 0.014	0.156 R <sup>2</sup> = 0.024	0.657** R <sup>2</sup> = 0.432
	Sig. (2-tailed)		.001	.141	.053	<.001
Integrated cohesion and coherence	Pearson Correlation			0.760** R <sup>2</sup> = 0.578	0.784** R <sup>2</sup> = 0.615	0.817** R <sup>2</sup> = 0.667
	Sig. (2-tailed)			<.001	<.001	<.001
Integrated lexical resource	Pearson Correlation				0.962** R <sup>2</sup> = 0.925	0.793** R <sup>2</sup> = 0.629
	Sig. (2-tailed)				<.001	<.001
Integrated grammatical range and accuracy	Pearson Correlation				1	0.816** R <sup>2</sup> = 0.666
	Sig. (2-tailed)					<.001

\*\* . Correlation is significant at the 0.01 level (2-tailed). \* . Correlation is significant at the 0.05 level (2-tailed).

Table 15 presents the Pearson coefficients between various measures of the integrated writing performance. Similar to Table 14, it was found that other measures had a non-significant but one weak significant correlation to source text use (ranging from 0.116,  $p = 0.152$  (ns) for Task achievement to 0.254,  $R^2 = 0.065$ ,  $p < 0.01$  for Cohesion and coherence). Compared to other measures, source text use had the lowest correlation coefficient to the total integrated score (0.657,  $R^2 = 0.432$ ,  $p < 0.01$ ). The correlation coefficients between other measures were relatively large (ranging from 0.760,  $R^2 = 0.578$ ,  $p < 0.01$  between Cohesion and coherence and Lexical resource, to 0.962,  $R^2 = 0.925$ ,  $p < 0.01$  between Lexical resource and Grammatical range and accuracy).

### 6.1.2 Qualitative findings from raters

Relationships between independent and integrated writing performance in qualitative analysis can be expressed in terms of common characteristics of good, mediocre and weak responses. An analysis of raters' interviews suggests that good independent and integrated writing responses share some common qualities: the language and the analysis sides. Rater 1 pointed out that the language side has to do with the precision of the *intended meaning*. He defines precision as "the ability to get exactly what the writer is saying".

So, I guess the hallmarks of a good essay would be, first of all, if we start from the language side of it, the candidate must be able to, or the writer must be able to express precisely what they mean. OK, so if I'm looking at something that's excellent, like, I mean, I was reading George Orwell's essay a few weeks ago. And I mean, the man precisely writes what he means, and the sentences are just so superbly written. And that's the language side of it.

---

Second, the analysis side, according to Rater 1, is related to the assessment of the development in an essay that should present relevant cohesive ideas to deliver intended messages.

There needs to be some clear development in the writing, the ideas need to be relevant, and the ideas need to be explored ... so as much as that particular time allows, you need to see an exploration of that topic, the development of that topic and also the cohesiveness of everything. You know, everything needs to flow. Everything needs to be in the right place. So, it's the combination of the language, ideas, and complexity of both, working together to deliver a particular message and in this case, answering the question.

Rater 2 has a similar view to that of Rater 1 in terms of what makes a good academic essay in both writing tasks. Generally, she believes that test-takers need to illustrate that they *have understood the task and addressed it relevantly and sufficiently within a given task condition*. She primarily looks for the level of task achievement and then *coherence and cohesion*.

Initially, I'd be looking to see that they've understood the question they're attempting, at least attempting to answer the question that it's not somehow tangential to what's being asked. Then, you're looking to see how developed that response is. But are they making generalisations, or are they in a better response? You'd expect to see things developed, extended and supported with examples ... So that's obviously more task achievement than at the level of coherence and cohesion. You'd be looking at the overall organisation.

Rater 2 pointed out that coherence and cohesion are subtle in good or skilful essays as they do not attract much attention during reading.

As a reader like you, if I read something by a professional writer, I do not notice *the cohesive devices*. You know, it just feels natural and kind of inevitable in the way something is structured. So, yeah, it's a little bit different in that way to vocabulary, but you can pinpoint, you know.

According to Rater 2, coherence and cohesion are related to the organisation of an essay that addresses a given task. In high-quality essay responses, writers employ a sophisticated referencing substitution, so the reader can follow the message quite easily. Rater 2 pointed out that:

ideally, you want to have some kind of introduction, then obviously a body, and then some kind of concluding section. Then paragraphing is another important thing. Yeah, and to say that the paragraph that's somehow organised logically, you know so that you might have, for instance, advantages in one paragraph, disadvantages in another, and then for the candidate's position in a really good like in an average response that might be quite mechanical. You know, it's just like here are the advantages. Here are the disadvantages. This is what I think in conclusion. But in a higher response, you would get a sense of the author's opinion right from the start, you know, even though the choice of vocabulary, perhaps, and they might consider opposing arguments. Still, it's like a sustained thing through the response.

Rater 1 compared a good essay with a mediocre or average one as he did not see many low candidates or many low responses in the current project. There are some dilemmas in their writing, which Rater 1 has expressed below.

They're [mediocre writers] able to get their point across with something ... So many mediocre writers, they, what they do is first of all, there's a lot of errors in their language and on top of that, they can't express what they mean.

---

Their vocabulary is quite limited, but not quite limited. But there's enough for you to get the idea, but not enough for you to get it precisely. So, they can give you an idea. Also, what they tend to do is they'll talk about a problem, but instead of exploring it, they'll go to another one and then instead of exploring that, they'll go to another one. And so that's how it works.

Rater 2 pointed out that in an average response, the use of cohesive devices tended to be more mechanical.

I guess that the more average level, like the bands five/six in regular IELTS, the difficulty might be that students have memorised a whole bunch of cohesive devices. So, things like "furthermore", "moreover", "additionally", but they're kind of like slotting them in for the sake of it. When you read carefully, you'll see actually that "moreover", like, they don't understand the function of that word. I'll just pop it in there for the sake of it, or they might write in conclusion, and then it's not actually a conclusion at all. So, as a marker, you don't want to be kind of tricked into thinking, oh, yeah, you use the creative device because it's not enough to have them in there like they have to be using them meaningfully.

Both raters pointed out that weak essays were found to be *short and lacked paragraphing*. Rater 1 further reflected on what he found in weak responses, which differed from what he found in mediocre ones, pointing out that the writing made it difficult for the reader to grasp the writer's ideas.

And then the other ones, I mean, the really low ones. Then what you're getting is you're not getting much meaning. You're not getting many ideas. Some were quite short. You know, with, say, one hundred and eighty words, you can't, you can't, especially if you don't have that language to be able to do it, so, yeah, that's the thing.

Rater 2 observed that:

so many of the candidates did not paragraph. So many candidates did not start with an introduction or end with a conclusion. It was just a slab of language; is a slap response ... That was what I found, and yes, I think I mentioned this maybe in our first meeting, it's very, very rare. Like I'll speak on regarding IELTS. It's very rare to get any candidate to get over a band seven that's just written the minimum requirement of two hundred and fifty words.

Rater 1 also noted that in contrast to strong test-takers, weak test-takers did not write an introduction, a body, or a conclusion, suggesting structural issues with their essays.

And if they did, their body was just one paragraph. So, what they do is they have an introduction. They'd write one body paragraph, meaning they had one main idea and the conclusion. And generally, I mean, it's just a rule of thumb to have two main ideas, at least, you know, or if you're going to have two main ideas, you should have two paragraphs. I mean, some did have two main ideas, but just in one paragraph. So, yeah, it was the structural issues.

For Rater 2, poor academic writing responses tend to comprise *illogical* and *incohesive* chunks of text.

It's just a big chunk of text. So, there's no intervention in paragraphs. It might just end abruptly. So, there's no kind of attempt to summarise the arguments in the form of a conclusion. And it might be a bit illogical. So, you don't get the sense that the author knows where they're going, like they say something and then they might contradict it and then they might return to the first point. And then the conclusion kind of comes out of nowhere, you know, like it's a bit of surprise and all. Like this is what the author thinks.

In summary, quantitative analysis suggests a moderate relationship between independent and integrated writing performances. The shared variance between the two writing performances was 26%. In qualitative analysis, performances in both writing tasks overlap in terms of the level of task fulfilment, essay lengths, qualities of cohesion and coherence, vocabulary and grammatical range, and paragraphing. It was found that the ability to utilise discourse synthesis of provided source texts was not well-exhibited compared with other skills. This might be because many test-takers were unfamiliar with this type of writing and were, therefore, unsure how the source texts could be used as part of their writing.

## 6.2 Research question 2: Do test-takers differ in performances between independent and integrated writing tasks?

Although it was found that performance in the independent writing task was moderately related to that of the integrated writing task, it cannot be assumed that test-takers performed similarly or better or worse in one task than the other. The second research question, therefore, aims to address whether the same test-taker performed similarly in both tasks and if they did not, which task was more difficult (addressing content and criterion-related validity). First, paired-samples *t*-tests were performed and interpreted. Paired-samples *t*-tests were performed for the total scores and each assessment criterion (except Source text). Then, a content analysis of test-takers' interviews is presented to discuss test-takers' experiences in addressing the integrated writing task, followed by the raters' expectations when evaluating integrated writing responses.

### 6.2.1 Quantitative findings

Table 16 presents the mean scores in the independent and integrated writing tasks. The mean scores suggest that test-takers performed better in the independent writing task.

**Table 16:** Mean scores of performances in the independent and integrated writing tasks (N = 154)

		Mean	Std. Deviation
Pair 1	Total independent scores	4.548	0.291
	Total integrated scores	4.150	0.454
Pair 2	Total independent scores	4.548	0.291
	Total integrated scores, excluding source text use	4.378	0.435
Pair 3	Independent task achievement	4.502	0.346
	Integrated task achievement	4.304	0.479
Pair 4	Independent cohesion and coherence	4.481	0.387
	Integrated cohesion and coherence	4.304	0.516
Pair 5	Independent grammatical range and accuracy	4.625	0.295
	Integrated lexical resource	4.482	0.443
Pair 6	Independent grammatical range and accuracy	4.584	0.308
	Integrated grammatical range and accuracy	4.422	0.427

Table 17 presents the *t*-test results between independent and integrated writing performances. As can be seen in this table, there were statistically significant differences in performances in the independent and integrated writing tasks ( $p < 0.01$  for both one- and two-sided). Test-takers performed significantly better in the independent writing task than the integrated writing task. In other words, the integrated writing task was more difficult than the independent writing task.

**Table 17:** *T-test results between independent and integrated writing performances (N =154)*

		Pair differences				Significance			
		Mean	Std. Deviation	95% Confidence Interval of the Difference		<i>t</i>	<i>df</i>	One-Sided <i>p</i>	Two-Sided <i>p</i>
				Lower	Upper				
Pair 1	Total independent scores - Total integrated scores	0.398	0.397	0.335	0.461				
Pair 2	Total independent scores - Total integrated scores excluding source text use	0.170	0.370	0.111	0.229	5.698	153	<.001	<.001
Pair 3	Independent task achievement - Integrated task achievement	0.198	0.457	0.125	0.271	5.378	153	<.001	<.001
Pair 4	Independent cohesion and coherence - Integrated cohesion and coherence	0.177	0.497	0.098	0.256	4.421	153	<.001	<.001
Pair 5	Independent grammatical range and accuracy - Integrated lexical resource	0.143	0.379	0.083	0.203	4.677	153	<.001	<.001
Pair 6	Independent grammatical range and accuracy - Integrated grammatical range and accuracy	0.162	0.367	0.104	0.221	5.495	153	<.001	<.001

Table 18 presents the effect sizes of the differences (Cohen's *d* and Hedges' correction). The effect size was large for the total writing performance difference ( $d = 0.997$ ). The score difference was nearly one standard deviation. The effect size between the independent and integrated writing scores without the source text use score was medium ( $d = 0.440$ ). In contrast, the effect sizes for the other assessment criteria were medium ( $d = 0.364$  to  $0.465$ ). These medium-size effects suggest nearly half a standard deviation in the score differences. Based on the statistical results, it can be seen that test-takers were more successful in the independent writing task than in the integrated writing task. They might be familiar with the independent writing task and were unsure of what was expected of the integrated writing task.

### 6.2.2 Qualitative findings from test-takers' interviews

The literature suggests that an integrated writing task is more complex than an independent writing task (e.g., Plakans, 2022) due to the need to synthesise information from source texts to written responses skilfully. Content analysis of the test-takers' interview scripts suggests five common explanations for why the integrated writing was more difficult for them to complete. They are not presented in order of importance.

## Unfamiliarity with integrated writing

A lack of task familiarity might explain why test-takers were less successful in the integrated writing task. A lack of task familiarity also meant that test-takers might not be aware of the need to acknowledge the source texts in their writing. For example, Cui realised she had *not understood the task requirements properly*. She thought that she only needed to summarise the source texts. The raters' comments indicated that Cui had not used the source texts (both raters awarded 1 out of 6 for using the source text). Figure 2 presents Cui's integrated writing response that reflected difficulty integrating source texts in writing. The following are excerpts from interview scripts that exemplify the issue.

Oh, yes, because maybe I misunderstand the requirements. I thought perhaps I'd do a summary or just to. And I'm not cite. ... Before I realised that I didn't cite the samples, I thought I did better than the formal one. (Cui)

Before, um, in China we, I think, we are not taught to how to write it, which just to, to write a little, a little article, a little paper and just about in one hundred and fifty words is very difficult I think for me to write academic essays. In China, my teacher taught me, and she said state your point in the first paragraph; you should tell your point. I really worry. It's not academic. (Yan)

And I'm using some materials and logical problems from reading to support my view. I don't know whether it's suitable or not, but I still write it down. (Zhao)

**Table 18:** Effect sizes of the differences among independent and integrated writing task performance criteria (N = 154)

			Standardiser <sup>a</sup>	Point Estimate	95% Confidence Interval	
					Lower	Upper
Pair 1	Total independent scores - Total integrated scores	Cohen's <i>d</i>	0.399	0.997	0.803	1.191
		Hedges' correction	0.400	0.995	0.801	1.188
Pair 2	Total independent scores - Total integrated scores excluding source text use	Cohen's <i>d</i>	0.386	0.440	0.280	0.601
		Hedges' correction	0.387	0.439	0.279	0.599
Pair 3	Independent task achievement - Integrated task achievement	Cohen's <i>d</i>	0.426	0.465	0.287	0.644
		Hedges' correction	0.427	0.464	0.286	0.643
Pair 4	Independent cohesion and coherence - Integrated cohesion and coherence	Cohen's <i>d</i>	0.463	0.382	0.206	0.558
		Hedges' correction	0.464	0.381	0.206	0.557
Pair 5	Independent grammatical range and accuracy - Integrated lexical resource	Cohen's <i>d</i>	0.393	0.364	0.205	0.523
		Hedges' correction	0.393	0.363	0.204	0.522
Pair 6	Independent grammatical range and accuracy - Integrated grammatical range and accuracy	Cohen's <i>d</i>	0.384	0.423	0.264	0.583
		Hedges' correction	0.385	0.422	0.263	0.581

a. The denominator used in estimating the effect sizes.

Cohen's *d* uses the sample standard deviation of the mean difference adjusted by the correlation between measures.

Hedges' correction uses the sample standard deviation of the mean difference adjusted by the correlation between measures, plus a correction factor.

**Figure 2:** Cui's integrated essay response

The problem of the urgent climate change and sea level is under heated discussion. Some scientists argue that human beings are blamed for overusing of the nature resources and polluting the environment, it is no doubt that people need to take actions to reduce the pollution and rescue our planet as soon as possible. However, others hold the view that there is still time for people to save the earth.

From my perspective, it is necessary for scientists to analyze data and find out effective methods to protect the environment, but we still can arise the awareness of the public to decrease environmental unfriendly actions. Although someone claims that there is no strong evidence to prove the global climate is influenced by humans. It is no denying that the temperature of the earth is increasing continuously which is closely related to the carbon emissions of private cars, air conditioners and even the production of our clothes. Sea level is increasing, ice cap is melting and more and more animals are losing their habitat, the impact of our ungreen actions is more than these. If humans still pass the buck and take no notice of their destruction, someday they will lose their home.

To put in a nutshell, it is instant that people need to stop from destroying the earth even start with the little things in our daily. The earth may be repaired with the high technique in the future, but we still need to protect our planet immediately in case of it is destroyed before the remedial solution come out.

**Rating:**

	Rater	Task Achievement	Source Text	Coherence and Cohesion	Lexical Resources	Grammatical Range and Accuracy
	1	4.5	1	4.5	4.5	4.5
	2	4.5	1	4.5	4.5	4.5

**Rater 1's comments:** There is a position but more concrete evidence is necessary. No use of sources.

**Rater 2's comments:** No clear evidence source material has been used.

**Reading demands**

Unlike an independent writing task, an integrated writing task requires *reading comprehension* and *extraction* of relevant information for use in writing. It also requires a skill to retrieve or recall and manage extracted information in written responses. Yan (Level 1 success) reported that while skimming the two texts to find what information they contained that may be useful to include in his response, he said that he had only scanned the texts and had not tried to understand them thoroughly. The following excerpts are from other test-takers that illustrate this reading demand.

But I think reading to write gives me more opportunity to use a wide range of vocabulary because, you see, the reading has all of the resources. I mean, the texts have to have the vocabulary that I need, and it's a reminder because when I say I'm not really good at vocabulary, I mean that I can't easily remind myself of certain vocabularies. I mean, I know these words, but I can't easily use them ... So, the first text is about its undeniable. It's so that would be the proponent, and the second text would be the opponent, which says it's not true. I started reading it carefully for both of the texts. (Wang)

---

It also requires test-takers to strategically read and use information from source texts.

In my head, I kind of like picked out some of the key points that I thought were important, and then I wrote them down. So then, after skimming it and, then kind of wrote down some key points for each of them. And in that writing down those key points gave me more of a chance to kind of take it like read it properly, like read it, not just skimming it, like reading it in depth. And then so I wrote down my key points and then I started writing. (Lizzy)

### Knowledge of the conventions of integrated writing

Several test-takers expressed a lack of academic writing conventions of integrated writing. They felt that good academic writing should have both depth and breadth of ideas, and certain words were unsuitable for their writing.

I always will use the same sentence structure to write ... In China, some teachers will ask students to summarise some topics, such as key sentence[s] while you see that question and that it have fixed the structure. And we will have to memorise that. So, I think that's just to suitable for exam but not suitable for my own writing skill because of that. (Yan)

Some test-takers reported difficulties using the source texts in their integrated writing. For instance, Liang, who was also found to be over-reliant on the source texts by the raters, pointed out that the source texts seem to have made it impossible for him to come up with his own ideas.

However, in terms of the writing process, I think there is a little negative influence (of the source texts). I have a feeling like there's some an emotional impact of having to read two texts and then to write the essay about the topic. After reading the passage, I was like, I was like, didn't have any other ideas because suddenly I couldn't think of any ideas of my own. Yeah, you relied on the ideas from the passage. (Liang)

Unlike an independent writing task, an integrated writing task requires paraphrasing and not knowing how to paraphrase can reduce their integrated writing quality. The following extracts illustrate demands or difficulties to paraphrase.

And I remember that because I didn't want to use the same sentence structure as the first text ... because I memorised the sentence about the issue and its structure. So I want to use that and paraphrasing the question again ... Because I use the reference in the first paragraph, so I find what sentence I can't use. So, I spent more time, right? And as I said, I got so nervous. And sometimes, I would think a long time to how to paraphrasing the key ideas and what's my words? I also spent much time to think my own points to write it down. (Yan)

It had to be selected because there was more information I could have used, but I just chose to use stuff about, like melting ice and also droughts in the US. So, I think the challenge is being selective about what support to use, like switching between reading and writing. And then also incorporating the skill of paraphrasing. Yeah. So that's something that I didn't have to do in the first task. (Sam)

Additionally, integrated writing requires synthesising source texts in their writing. Lizzy reflected on this aspect.

And I noticed after as well, like in the questionnaire when I did it. I did quote it in like each one in the different paragraphs. But then I was thinking about after and I was like, oh, I could have put the two sides in one next to each other to synthesise and compare and contrast the ideas more rather than just having them totally separate.

---

Um, I think it would have been better to have, you know, like the part with the where they say, like, you can wait twenty-five years. I think to compare that to what the climate scientists were saying would have probably been better to put it next to it, to strengthen it, to say, twenty-five years. But like what they're actually saying is that it's not, it's not twenty-five years you can wait. It's like right now. (Lizzy)

### Topical knowledge

Several interviewees suggested that their integrated writing responses were related to topical knowledge.

I think maybe because, ah, I haven't read enough books, so maybe I see some point in the literature, I can't see, I can't write and read critically ... I need to go further reading to store some examples, so that I can find some support, some background knowledge. (Cui)

Because sometimes it's not about language only; it's about knowledge. It is about your worldview and other different things. So, I think reading can be really helpful, you know, that and all the things that I do. (Liang)

If we don't have the background knowledge, I think I, I really don't know how to answer that or write a long essay about that question. (Yan)

You know, but I think it is a kind of writing. We must think from different perspectives. You can't just think this in one way. Right. So, use other examples to support it. But I don't have this kind of example. (Lu)

The first and foremost reason would be my knowledge of the topic. So if I have enough knowledge on this topic, I can write quite easily. So, I think that's the greatest reason that affects my writing ... I guess when I was young, I didn't have much knowledge on many topics. (Wang)

### Lexical knowledge

The interviewees also pointed out that in relation to topical knowledge, they struggled with specific vocabulary in their integrated writing.

About the vocabulary, because, you know, there are a lot of topics about academic writing, and a lot of them are so professional words I'm not familiar with. I just have to, you know, use some other words to replace it. But the words are not as accurate as the original ones. (Zhao)

I think the vocabulary is always one of the biggest issues. You know, if, for example, if we write like we're not doing a limited test, we can rely on, for example, search engines. Yes, search the dictionary for such responses. (Ren)

Well, a vocabulary might include colloquialisms, like colloquial expressions. So, these are the things that I am not really good at ... But maybe I'm not using a good range of it. (Wang)

I actually liked it when I was writing yesterday for the test; I felt like sometimes the words weren't coming to me. So, I'm looking for the right word to use and then it's not coming. And then I had to go back and change several sentences so that it fits. (Sam)

---

### 6.2.3 Qualitative findings from raters' interviews

Context analysis suggests several issues the raters observed when evaluating the integrated writing task.

#### Understanding the task requirements

Both raters reported that some test-takers, especially weak ones, might not properly understand the task instructions and were unable to comprehend the source texts accurately. Both raters pointed out that their primary focus was assessing whether test-takers *answered the question*. Rater 1 explained two key factors that affected his rating of the integrated writing responses (i.e., length and relevant ideas to the task).

Many were under length. And when you're on the length, and you're using somebody else's work, and you're quoting that, that takes away your own work. I took that into consideration while I was doing it (trying out the task).

Success in an integrated writing task depends on comprehending the provided source texts, the task instructions, and the specific prompt.

The second thing was how good with the ideas; I mean, a lot of people talk about global warming, OK, and something has to be done about it, you know. This is bad, and we should do this, and we should do that, OK? That's not what the questions are asking. The question asks, "Is it happening or not?" Some people say it's happening; other people think it isn't, or that there's one group of scientists who think it is, and there's another group of scientists who argue that it isn't, OK? What do you think? It's not asking you what should be done about the problem. It's saying, does the problem exist? Pick a side and argue for that side. And so that's another thing I look at. (Rater 1)

#### Paraphrasing

Many test-takers tended not to paraphrase source texts. They tended to use *only what they could understand in their essays*.

The way they first started was no paraphrasing. Yeah, and, you know, you had the sense that they didn't quite understand what they were reading, or maybe that was the only bit of the thing that they understood. They might have seen the words 'polar bears' and 'seals' and not really understood anything else. 'But there are polar bears and seals losing their homes. So I can understand that. I'll put that in.' (Rater 2)

There were slight test-taker groups that misinterpreted what was happening and what was being said, but not many, you know. [They needed to have an] understanding of what the text means and stuff. Some took liberty with that, like, oh well, this is what this means. And it's like, well, actually, it doesn't mean that ... You need to be able to understand what's being said. (Rater 1)

Rater 1 recalled that most test-takers used direct quotations in their essays.

There were a few who paraphrased. Yeah. That was about it. Yeah. It wasn't. I mean but most I would say were just taking quotes. Mm hmm. Yeah. I mean, to give them credit, many of them did reference the source material.

Rater 2 pointed out that paraphrasing should be more about the global message from the source texts, rather than sentence-level concepts.

There was very little paraphrasing of the global message or, you know, it would be nice to have *paraphrasing of the two and to bring them together somehow or to contrast them*. But they didn't seem to do that. It was more like *granular things* like they might swap two words around.

---

### Engaging with source texts

Both raters expected to see some level of engagement with the input test-takers used from the source texts, for example,

You would want there to be a quote from the text and then some kind of some kind of input from the candidate. To say like this is similar to this example "in my country" or "I don't like". I find this example interesting because often there was none of that. It was just the quotes and nothing around it. (Rater 2)

### Integration of source texts and own information

Both raters, however, observed some excellent responses to this integrated writing task. Good integrated writing task responses appear sophisticated in terms of well-structured organisation, presenting a clear position, utilising relevant source texts, and combining test-takers' knowledge.

There were some good responses, and they were able to do a great job of weaving together their own knowledge and their own examples, plus using the source material to support what they were saying. (Rater 1)

I guess the good responses had understood the global message [from the source texts]. Then they'd gone through and maybe selected a few specific examples quoting Curry or whoever the person was. Smith, I forget. Yeah. You know, not just kind of, yeah, like cut things a little bit randomly and stuck them in. (Rater 2)

Test-takers who performed well in this task tended to discuss both sides of the argument.

Because the better response would say, well, this side says this. However, there are flaws in what it's saying, and this side says that. And here is why this side is. I mean, better responses would be able to look at both sides, and their argument would be so strong because they would sort of undermine what the other side said. Whatever the side they picked, doesn't matter whether it's happening or not. Whatever side they picked, they would undermine the opposite side. They would acknowledge it. And then they, you know, using their knowledge holds in. (Rater 1)

Rater 1 also noted the *difference between responses in band 5 and those in band 6* in the current project rubric scale as follows:

I think if you showed one side and if you just quoted from one source, you could get up to about five. But if you were looking at a band six, well, it says there's nothing that could possibly be added to the response. Well, yeah, you could sort of add that the other side exists, right?

### Vocabulary choices

Rater 2 observed more flexible use of vocabulary choices in Task 1 (independent) than in Task 2 (integrated).

I guess one thing. Yeah, one thing I noticed is the dimension of COVID and the pandemic. Yeah. Yes, so that's kind of interesting because like, yeah, a topic about sports facilities doesn't really naturally lead to talking about diseases. But, yeah, it's kind of interesting that because of the context that we're in, then so many of them did make that connection, which, you know, you can make. But, yeah, that's kind of interesting at a vocabulary level.

---

### Integrated writing skills

Both raters pointed out that test-takers had to realise that *there were some specific expectations in relation to essay content and that a new set of writing skills was required to write an integrated essay successfully*. This task expected test-takers to use source texts and their knowledge and examples.

You need to have your own knowledge. You can't just rely solely on the source material. But at the same time, for this particular task, you can't wholly ignore it. It needs to be a combination of both. And again, I understand that this task, they only had 50 minutes, 10 minutes to read the material and 40 minutes to write. I understand that. So, in that particular time frame, whatever they can do to use their own knowledge and use the supporting, use the source material to support their main ideas and their own knowledge, then, especially on a topic like global warming, I mean, it's everywhere. (Rater 1)

It does require another set of skills. Because it's no longer just being faced with a blank page. It involves reading skills, close reading skills and more like analytical skills and paraphrasing skills. Yeah. The regular task [independent essay task] doesn't really require them. (Rater 2)

Rater 2 explained a scenario of what she would do when she was faced with an integrated writing task. Her approach is useful to understand what successful test-takers would do in response to this task.

What I would do if I were faced with this task would be to like closely read and understand the global. You know, the global message of each of the two source texts and then to make those two opposing messages clear in my response, so even if I didn't agree with those two, I would still want to be referring to Bennett and saying, you know, some kind of rebuttal against him.

### Underuse of source texts

Rater 1 mentioned that approximately 10% of the test-takers did not quote or use the source material in their essays, whereas Rater 2 felt that the source texts were poorly used, even by candidates with high-level English skills. Rater 2 further pointed out that test-takers tended to include a line or two from the source texts but did not engage with them when the source texts were underused. Rater 2 further noted that many test-takers used concrete examples from the source texts (e.g., droughts in the US, Greenland's ice cap, the polar bears, the seals, and statistics) rather than *abstract concepts* from the source texts.

Consequently, not using the source texts and their own knowledge in essay responses would affect the scores awarded for various scoring criteria. Rater 1 exemplified this point.

But that would affect everything around it because if all they have used is that source material, they haven't answered the question. [If] they haven't used their own language. [If] they haven't organised the information the way it should be, and [if] they haven't used their own grammar.

---

### Overuse of the source texts

Both raters reported that those who performed the poorest were the ones who copied and pasted information from the source texts into their essays.

They either relied a little bit too much on the source material, or they relied wholly on the source material. So, in other words, there are no original ideas. There's nothing from the candidate. They're not showing their own knowledge of the topic. (Rater 1)

There were a lot of responses that copied things for Boston without any quotation marks. And I remember in the training, like I personally, I think that should be penalised. (Rater 2)

So, there was plenty of overuse. They can't have their own idea. I mean, yes, some overused it, but not many; some overused it by just copying the text and pasting; others overused it by just rewording it and summarising it. So again, it was just a summary of the texts. It really wasn't an argument. (Rater 1)

Both raters pointed out that several weak test-takers tended to *summarise* the source texts, which suggested some misconceptions regarding this writing task requirement.

Many of them summarised the source material, right, so they just took it and said, OK, in one paragraph, this is what Smith says in another paragraph. This is what Barnett says, Barnett. And that's the conclusion. (Rater 1)

Rater 2 also commented on how referencing the source texts in weak essay responses often involved test-takers only *artificially putting in references without utilising the source texts*.

Like it reminds me a little bit like at university. I remember if I'm in a typical university essay, I might say like an undergraduate level like reference 12 sources, you know and like I was always very conscientious, but I would maybe read dozens of books and find the 12 most suitable and then go through and find like a basic selective about it and careful about it. Whereas some other students, like they write the essay and then afterwards they go and find 12 books and like stick them in. You know, it felt like that to me.

Rater 2 noted an issue of the irrelevance of source text used by test-takers.

I guess that sometimes it was a bit illogical or incoherent in the way the sources were used. Uh. Yeah, I can't think of an example off the top of my head. Um ... Sometimes the sources were confined to the introduction or just one paragraph. They might just have one paragraph that was ninety-nine per cent source material and then three paragraphs of their own stuff, whereas ideally, you would want it to be more evenly spread through the response.

In summary, in this research question, it was found that test-takers performed significantly more poorly in the integrated writing task than in the independent writing task (overall and in each assessment criterion). The qualitative data from both test-takers and raters suggest several issues explaining why the integrated writing task was more difficult. Research question 4 will report a content analysis of raters' comments that further explain features of integrated writing responses in the current study.

### 6.3 Research question 3: What are the relationships between reported IELTS scores and independent and integrated writing performances?

Research questions 1 and 2 did not consider test-takers' previous IELTS scores in data analysis. This research question focuses on quantitative analysis of the relationships between IELTS scores and independent and integrated writing performance through correlations, multiple regressions and path analysis (addressing criterion-related validity).

#### Pearson correlations between reported official IELTS and independent and integrated writing task performance

Table 19 presents the descriptive statistics of the reported IELTS scores. While the maximum overall IELTS score was 8.5, Table 19 suggests that the participants' English proficiency, on average, was not very high, equivalent to Level B2 on the Common European Framework of Reference for Languages (CEFR).

**Table 19:** Descriptive statistics of the overall IELTS, IELTS Reading and IELTS Writing scores ( $N = 152$ )

	Overall IELTS	IELTS Reading	IELTS Writing
Mean	5.553	5.636	5.461
Median	5.000	5.500	5.000
Mode	5.000	5.000	5.000
Std. Deviation	0.858	0.937	0.734
Minimum	4.000	4.000	4.000
Maximum	8.000	8.500	7.500

A Pearson correlational analysis was performed to investigate the linear relationships between reported IELTS scores and independent and integrated writing task performance. Table 20 presents the Pearson correlation coefficients. First, the correlations of the IELTS Reading and Writing scores to the overall IELTS scores were large ( $r = 0.926$ ,  $R^2 = 0.86$ ,  $p < 0.01$  and  $r = 0.906$ ,  $R^2 = 0.82$ ,  $p < 0.01$ , respectively). The relationship between the IELTS Reading and Writing scores was also strong ( $r = 0.827$ ,  $R^2 = 0.68$ ,  $p < 0.01$ ). Second, it was found that the relationships of the independent writing task performance to the overall IELTS, IELTS Reading and IELTS Writing variables (i.e.,  $r = 0.57$ ,  $R^2 = 0.32$ ,  $p < 0.01$ ;  $r = 0.56$ ,  $R^2 = 0.31$ ,  $p < 0.01$ ; and  $r = 0.60$ ,  $R^2 = 0.36$ ,  $p < 0.01$ , respectively) were stronger than those of the integrated writing task performance (i.e.,  $r = 0.35$ ,  $R^2 = 0.12$ ,  $p < 0.01$ ;  $r = 0.34$ ,  $R^2 = 0.12$ ,  $p < 0.01$ ; and  $r = 0.36$ ,  $R^2 = 0.13$ ,  $p < 0.01$ , respectively). It is important to caution that the two writing tasks in the current study were not completed under high-stakes conditions as in the official IELTS test. Under the same high-stakes condition, test-takers' writing task performances may be higher or lower. Third, the correlation between the independent and integrated writing task performances was statistically significant ( $r = 0.51$ ,  $R^2 = 0.26$ ,  $p < 0.01$ ). As expected, since the integrated writing tasks required test-takers to read texts and to synthesise information from them for their writing, the two sets of writing task performance were found not to be strongly correlated.

## Multiple regression analysis

Given the statistical correlations between IELTS scores and independent and integrated writing performance, a multiple regression analysis was performed to explore the relative contribution of IELTS and independent writing scores (i.e., predictors) on integrated writing performance. All predictors were analysed simultaneously in a multiple regression analysis to understand their relative importance in predicting integrated writing performance. The correlation between all predictors (R-value) was 0.511,  $R^2 = 0.261$  (adjusted  $R^2 = 0.241$ ). The value suggests the predictors explained 26% of the integrated writing performance variance. The  $R^2$  change, which is the difference between the statistical regressions without these predictors and the current model, was found to be significant ( $F(4, 147) = 12.99, p < 0.01$ ). Table 21 presents the multiple regression model coefficients. It was found that independent writing had the largest beta coefficient ( $\beta = 0.450, p < 0.001$ ), while all the IELTS scores were statistically non-significant. In this multiple regression analysis, the independent writing variable was the only predictor related to how much the integrated writing performance increased when the independent writing variable increased by one unit. If test-taker A performed a level higher than test-taker B in the independent writing task, test-taker A's integrated writing score would be 0.703 higher than that of test-taker B.

This was also confirmed by the collinearity statistics (the point at which it became difficult to distinguish the contribution of the predictive variables) with a value larger than 0.2 and VIF below 10 (confirming the absence of excessive collinearity) for the independent writing variable. Given the current findings, a hierarchical multiple regression analysis was not necessary.

**Table 20:** Pearson correlation coefficients between reported IELTS scores and independent and integrated writing task performance (N = 152)

		Overall IELTS	IELTS Reading	IELTS Writing	Independent Writing	Integrated Writing
Overall IELTS	Pearson Correlation	1	0.926** $R^2 = 0.86$	0.906** $R^2 = 0.82$	0.567** $R^2 = 0.32$	0.348** $R^2 = 0.12$
	Sig. (2-tailed)		<.001	<.001	<.001	<.001
IELTS Reading	Pearson Correlation		1	0.827** $R^2 = 0.68$	0.560** $R^2 = 0.31$	0.336** $R^2 = 0.11$
	Sig. (2-tailed)			<.001	<.001	<.001
IELTS Writing	Pearson Correlation			1	0.600** $R^2 = 0.36$	0.361** $R^2 = 0.130$
	Sig. (2-tailed)				<.001	<.001
Independent Writing	Pearson Correlation				1	0.505** $R^2 = 0.260$
	Sig. (2-tailed)					<.001
Integrated Writing	Pearson Correlation					1
	Sig. (2-tailed)					

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

**Table 21: Multiple regression model coefficients**

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	0.657	0.515		1.274	.205					
	Overall IELTS	0.040	0.133	0.076	0.302	.763	0.348	0.025	0.021	0.080	12.554
	IELTS Reading	-0.008	0.092	-0.016	-0.082	.935	0.336	-0.007	-0.006	0.138	7.236
	IELTS Writing	0.021	0.108	0.035	0.199	.842	0.361	0.016	0.014	0.166	6.014
	Independent Writing	<b>0.703</b>	0.140	<b>0.450</b>	5.020	<.001	0.505	0.383	0.356	0.625	1.601

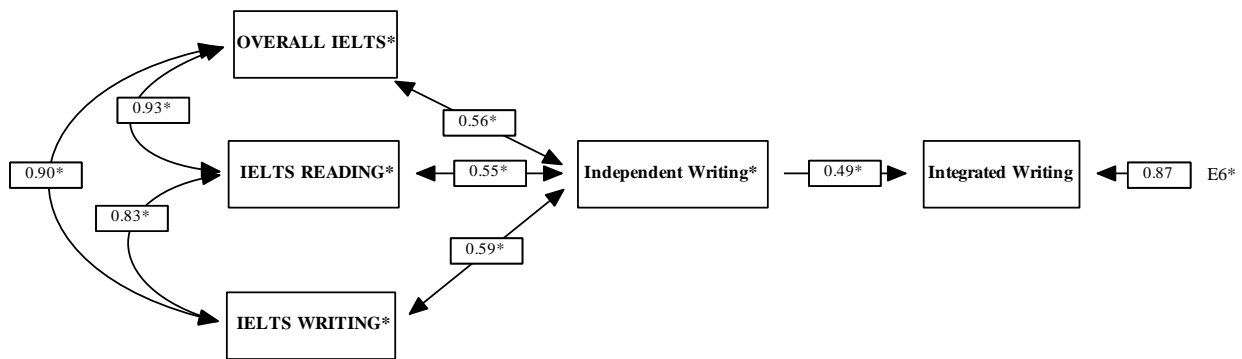
a. Dependent Variable: Integrated Writing

### Path analysis

Although the multiple regression analysis found that independent writing performance was the only significant predictor of integrated writing performance, it was theoretically rational to consider that prior IELTS proficiency had an indirect relationship to integrated writing performance via independent writing performance. It may be said that when other significant language proficiency variables such as academic English proficiency as measured by official IELTS scores are considered simultaneously in a single statistical analysis (e.g., path and structural equation modelling), a more accurate picture of the complex relationship between the independent and integrated writing performances was better explained. Because L2 writing is known to be theoretically complex in that many variables are simultaneously interacting, a correlation between variables taken two at a time (as in standard statistics) can be limited. However, this can still shed some light on their complex relationships.

A path analysis was performed via the EQS 6.3 Program to explore this possibility. It should be noted that a structural equation modelling analysis was initially intended to be used. However, due to a smaller sample size than anticipated ( $N < 200$ ), it was decided that it should not be used as it would be prone to statistical inference error. Several path models were tested. For example, the three IELTS variables correlated to the independent and integrated writing variables. A model with and without correlational coefficients between IELTS variables was also tested. In the current study, a path model in which all three IELTS variables were directly correlated with the independent writing variable, which in turn had a regression path to the integrated writing variable, was found to be best explained by the dataset. In the analysis, the estimation method used was the iteratively reweighted least square solution (Heterogeneous Kurtosis Distribution Theory). Figure 3 presents this hypothesised path model.

**Figure 3:** The hypothesised path model of the relationships between IELTS variables and independent and integrated writing variables (N = 152)



Chi-square ( $\chi^2(3)$ ) = 0.884,  $p$  = 0.829, CFI = 1.00, RMSEA = 0.00 (90% CI = 0.00–0.08)

In this path model, the average absolute residual was 0.0091, with the average off-diagonal absolute residual of 0.0111, which was very small, indicating a good model fit. All residual values (100%) were within 0.0, suggesting the distribution to be symmetric and centred near zero. The path model in Figure 3 had a very good model fit. The Chi-square ( $\chi^2$ ) statistic was non-significant, suggesting a large difference between the independent  $\chi^2$  and the tested model. The fit indices indicate very good model fit (e.g., Bentler-Bonett NFI = 0.998; CFI = 1.00, RMSEA = 0.00 (90% confidence interval = 0.00, 0.08)). The fit indices imply that the regression path from the latent independent writing task performance variable to the integrated writing task performance was correctly specified. All paths and parameter estimates were significant at 0.05.

According to this path model, the overall IELTS scores had a strong correlation to IELTS Reading and Writing scores ( $r = 0.93$ ,  $R^2 = 0.86$ , and  $r = 0.90$ ,  $R^2 = 0.81$ , respectively) and the correlation coefficient between IELTS Reading and Writing scores was 0.83 ( $R^2 = 0.69$ ), which was also very strong. In terms of their correlations to the independent writing variable, it was found that their relationships were moderately strong:  $r = 0.56$  ( $R^2 = 0.31$ ) for Overall IELTS),  $r = 0.55$  ( $R^2 = 0.30$ ) for IELTS Reading, and  $r = 0.59$  ( $R^2 = 0.35$ ) for IELTS Writing. The strengths of the path correlation coefficients were similar to those found in the Pearson analysis (see Table 20). In Figure 3, the correlation coefficient between the independent and integrated writing variables was found to be 0.49 ( $R^2 = 0.24$ ), which was also slightly lower than that found in the Pearson correlation (Table 20). In this path model, the indirect correlations of the overall IELTS, IELTS Reading and IELTS Writing to the integrated writing performance were 0.27 (i.e.,  $0.56 \times 0.49$ ), 0.27 (i.e.,  $0.55 \times 0.49$ ), and 0.29 (i.e.,  $0.59 \times 0.49$ ), respectively.

In summary, since the IELTS variables were found to influence independent writing performance simultaneously and integrated writing task performance (albeit more indirectly), since the independent writing performance was significantly related to the integrated writing task performance, it directly affects integrated writing task performance. It is thereby quite predictive of the integrated writing task performance (i.e., 25% shared variance). Therefore, the integrated writing performance can be regressed to the independent writing task performance. It may be inferred that to perform an integrated writing task successfully, test-takers must first be competent in an independent writing task, which is simultaneously explained by general academic English proficiency, including writing and reading (up to a quarter of the integrated writing task performance variance).

## 6.4 Research question 4: Do test-takers with different integrated writing success levels differ across different writing criteria?

This question investigates variations in performances based on success levels (addressing context and scoring validity).

### 6.4.1 Quantitative findings

This section examines differences in the integrated writing task performance for each criterion. To identify the levels of success in the integrated writing task, the frequency counts of observed scores were examined (ranging between 1.40 and 5.20). Using the ranges of scores and sample sizes, three success groups were formed:

- Scores 1.4 – 3.99 inclusive = Group 1 (n = 43)
- Scores 4.00 – 4.39 inclusive = Group 2 (n = 57)
- Scores 4.40 – 5.50 inclusive = Group 3 (n = 54).

The average word counts for each group were as follows:

- Group 1: 235 (SD = 68)
- Group 2: 281 (SD = 72)
- Group 3: 317 (SD = 74).

It should be noted that the decisions on cut-off scores were subjective as the decision was arbitrary. Table 22 presents the mean scores of the three groups. To investigate differences between the three groups in terms of each writing criterion, MANOVA was performed because a number of dependent variables were compared. The Box's test of equality of covariance matrices was found to be significant ( $F(30, 64176.556) = 9.232, p < 0.001$ ), and Levene's test of equality of variances ( $p < 0.001$ ) suggests similar results. These statistics indicate that the dependent variables (i.e., the aspects of integrated writing task performance) were non-equal across groups. In the current analysis, Tamhane post hoc tests, which assume non-equal variances, were used to examine group differences. Table 23 presents the tests of between-subject effects.

**Table 22:** Mean scores of the three groups (N = 154)

	Success	Mean	Std. deviation	N
Task achievement	1.00	3.948	0.626	43
	2.00	4.316	0.342	57
	3.00	4.574	0.226	54
	Total	4.304	0.479	154
Source text	1.00	2.093	1.038	43
	2.00	3.145	0.945	57
	3.00	4.250	0.389	54
	Total	3.239	1.187	154
Cohesion and coherence	1.00	3.855	0.688	43
	2.00	4.382	0.321	57
	3.00	4.579	0.193	54
	Total	4.304	0.516	154

Lexical resource	1.00	4.198	0.676	43
	2.00	4.504	0.224	57
	3.00	4.685	0.207	54
	Total	4.482	0.443	154
Grammatical range and accuracy	1.00	4.128	0.646	43
	2.00	4.461	0.215	57
	3.00	4.616	0.193	54
	Total	4.422	0.427	154

**Table 23:** MANOVA tests of between-subject effects ( $N = 154$ )

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
<b>Corrected Model</b>	Task achievement	9.406 <sup>a</sup>	2	4.703	27.618	<.001	0.268
	Source text	112.171 <sup>b</sup>	2	56.086	82.026	<.001	0.521
	Cohesion and coherence	13.100 <sup>c</sup>	2	6.550	35.777	<.001	0.322
	Lexical resource	5.734 <sup>d</sup>	2	2.867	17.831	<.001	0.191
	Grammatical range and accuracy	5.831 <sup>e</sup>	2	2.915	19.911	<.001	0.209

a.  $R^2 = 0.268$  (Adjusted  $R^2 = 0.258$ )

b.  $R^2 = 0.521$  (Adjusted  $R^2 = 0.514$ )

c.  $R^2 = 0.322$  (Adjusted  $R^2 = 0.313$ )

d.  $R^2 = 0.191$  (Adjusted  $R^2 = 0.180$ )

e.  $R^2 = 0.209$  (Adjusted  $R^2 = 0.198$ )

According to Table 23, there were statistically significant differences between the performances of the three success groups in the integrated writing task ( $p < 0.001$ ), and the effect sizes were relatively large (partial eta squared ( $\eta^2$ ), ranging from 0.191 (lexical resource) to 0.52 (source text). Table 24 presents the Tamhane post-hoc test results. It was found that all but two (lexical resource and grammatical range and accuracy between Groups 2 and 3) were significantly different. The more successful groups outperformed the less successful groups in three assessment criteria. Observing the large differences among groups in the source text criterion is particularly intriguing. It can be implied that the ability to do discourse synthesis of provided reading texts in this task differentiates success in integrated academic writing in this area (Plakans, 2022). However, test-takers in Groups 2 and 3 did not differ significantly in their ability to illustrate lexical resources and grammatical range and accuracy.

**Table 24:** Tamhane post-hoc tests (N = 154)

Dependent Variable	(I) Success Integrated	(J) Success Integrated	Mean Difference (I-J)	Std. Error	Sig. <sup>b</sup>	95% Confidence Interval for Difference <sup>b</sup>	
						Lower Bound	Upper Bound
Task achievement	1.00	2.00	<b>-0.368*</b>	0.083	<0.001	-0.570	-0.166
		3.00	<b>-0.626*</b>	0.084	<0.001	-0.831	-0.422
	2.00	3.00	<b>-0.258*</b>	0.078	0.004	-0.448	-0.069
Source text	1.00	2.00	<b>-1.052*</b>	0.167	<0.001	-1.456	-0.647
		3.00	<b>-2.157*</b>	0.169	<0.001	-2.566	-1.748
	2.00	3.00	<b>-1.105*</b>	0.157	<0.001	-1.485	-0.725
Cohesion and coherence	1.00	2.00	<b>-0.527*</b>	0.086	<0.001	-0.736	-0.318
		3.00	<b>-0.724*</b>	0.087	<0.001	-0.936	-0.512
	2.00	3.00	<b>-0.197*</b>	0.081	0.049	-0.394	0.000
Lexical resource	1.00	2.00	<b>-0.307*</b>	0.081	<0.001	-0.503	-0.111
		3.00	<b>-0.488*</b>	0.082	<0.001	-0.686	-0.289
	2.00	3.00	-0.181	0.076	0.057	-0.365	0.004
Grammatical range and accuracy	1.00	2.00	<b>-0.333*</b>	0.077	<0.001	-0.520	-0.146
		3.00	<b>-0.488*</b>	0.078	<0.001	-0.677	-0.298
	2.00	3.00	-0.155	0.073	0.103	-0.331	0.021

Based on estimated marginal means

\*. The mean difference is significant at the 0.05 level. b. Adjustment for multiple comparisons: Bonferroni.

#### 6.4.2 Qualitative findings

MANOVA results suggest differences in the qualities and features of integrated writing responses across the three success groups. This has led to a qualitative analysis of the raters' comments that might explain the characteristics of integrated writing performance by different success groups. To present the findings, a communality data matrix based on the two raters was constructed and used to describe and exemplify the nature of integrated writing for each success group (Levels 1 to 3). It is important to note that while the number of counts used for organising the features of raters' comments in the data matrix may represent some common features of writing responses by this group of test-takers, the count number cannot indicate the relative importance or value of these features in the essay evaluation. That is, raters may mention a feature less frequently, but it may be used to determine the level of writing quality more significantly than those mentioned more frequently. Each matrix includes a summary of the raters' comments and evidence using direct quotes from the raters' written comments. In all the matrices, (+) and (-) are used to indicate positive (+) or negative (-) features.

## Group 1's integrated writing features

Figure 4 presents a display matrix that explains the nature of Group 1's integrated writing features, according to Raters 1 and 2. In Figure 4, five key commentary features are identified and organised based on the number of code counts. Raters' comments pointed towards Level 1 test-takers lacking good writing qualities, missing important writing features and having difficulties addressing the task. All negative features (-) are present in this matrix.

**Figure 4:** Group 1's features of integrated writing performance

Success level	Summary of rater comments	Examples
<b>Group 1</b> (n = 43 for Rater 1; n = 40 for Rater 2)		
<p><b>1. Under/unclear development or position (-)</b> (= 28 counts for Rater 1; 21 counts for Rater 2)</p>	<p>Test-takers tended to present insufficient development of essay directions, including no conclusions. Essays were not clearly articulated or extended to meaningful discussion. Underdevelopment tended to be affected by their unclear positions toward the task prompt, thereby under-developed. There may be a lack of or insufficient supporting evidence from the source texts and their own. Weak argument and evidence (either based on their own or source texts) or ineffective source text use (using less important or weak information from source texts) might appear in an essay.</p>	<ul style="list-style-type: none"> <li>• The development is <u>not clear</u>. The candidate simply takes the ideas and pads them with a position. (Rater 1/Test-taker 27)</li> <li>• Clearly this candidate's arguments <u>need a lot more development</u> (Rater 2/Test-taker 51)</li> <li>• A short response with relevant ideas but <u>not much development</u>. (Rater 1/Test-taker 54)</li> <li>• <u>Quite a confused response</u>; difficult as a reader to follow the candidate's logic (Rater 2/Test-taker 90)</li> <li>• A clear position, but obviously this response <u>needs more development</u> (Rater 2/Test-taker 113)</li> <li>• While there is a position, the <u>evidence used by the candidate is quite weak</u> such as the link between tsunamis and global warming (Rater 1/Test-taker 117)</li> </ul>
<p><b>2. No, little or minimal use of source texts (-)</b> (= 23 counts for Rater 1; 16 counts for Rater 2)</p>	<p>Test-takers tended to use no source texts in their writing, or if they did, they tended to use them very little or minimally (i.e., lack of clear evidence). Such a lack of source text use could not fulfil the integrated writing criterion. This feature was found to overlap with the under/unclear development/position category.</p>	<ul style="list-style-type: none"> <li>• Source texts used only <u>tangentially</u> (Rater 2/Test-taker 54)</li> <li>• <u>Minimally uses</u> the sources to develop their point of view. (Rater 1/Test-taker 82)</li> <li>• The source texts <u>are not used</u> (Rater 1/Test-taker 86)</li> <li>• <u>None of the source material is used</u> (Rater 1/Test-taker 94)</li> <li>• <u>No evidence source material has been used</u> (though possibly it has been read re polar bears??). (Rater 2/Test-taker 144)</li> </ul>
<p><b>3. Insufficient responses (-)</b> (= 18 counts for Rater 1; 13 counts for Rater 2)</p>	<p>Test-takers tended to produce short or under-length texts, making them insufficient to assess or understand the positions and arguments of the test-takers. This feature was connected to under essay development and an absence of source text use. Test-takers tended to write a single short paragraph.</p>	<ul style="list-style-type: none"> <li>• The candidate responds <u>minimally</u> (Rater 1/Test-taker 20)</li> <li>• <u>159 words only. No paragraphing</u> (Rater 2/Test-taker 45)</li> <li>• <u>A short response</u> with relevant ideas but not much development (Rater 1/Test-taker 54)</li> <li>• A response <u>65 words short</u> (Rater 1/Test-taker 82)</li> <li>• The response is <u>140 words short</u> (Rater 1/Test-taker 89)</li> </ul>

<p><b>4. Ineffective/inaccurate/improper source text use (-)</b> (= 11 counts for Rater 1; 13 counts for Rater 2)</p>	<p>Test-takers might have attempted to use source texts but did not acknowledge them in their essays. They might have indirectly used source texts or broadly mentioned source texts in their essays but did not acknowledge them successfully. They tended to copy the text verbatim into their essays. They did not tend to paraphrase source texts. Test-takers lacked effective skills in using source texts, such as inaccurate source text use. A lack of proper reading comprehension or engagement in reading might cause this.</p>	<ul style="list-style-type: none"> <li>• This candidate has <u>cut and pasted</u> great chunks of material from the source texts, only inserting two total words of their own. This makes it very difficult to know their actual ability! (Rater 2/Test-taker 23)</li> <li>• Inclusion of <u>direct quotes</u> is done in quite a <u>sloppy way</u>. Within quote marks, the candidate has included their own paraphrasing, and in the final paragraph two separate sentences are joined together. (Rater 2/Test-taker 27)</li> <li>• A good attempt to integrate the source material, however <u>more commentary/paraphrasing</u> of this material required. (Rater 2/Test-taker 71)</li> </ul>
<p><b>5. Over-reliance on source texts (-)</b> (= 13 counts for Rater 1; 6 counts for Rater 2)</p>	<p>Unlike those who did not use source texts, test-takers who used source texts tended to rely heavily on them in their essays. This could take the form of direct copying of the source texts or a descriptive summary of one or two source texts. This often led to a perceived lack of originality. Test-takers may use the same information from the same source text repetitively.</p>	<ul style="list-style-type: none"> <li>• The other paragraphs were <u>copied</u>. (Rater 1/Test-taker 23)</li> <li>• <u>Almost all the essays use the source material</u>. The introduction is the only part that has original writing. (Rater 1/Test-taker 29)</li> <li>• <u>A third of it is the source</u> (Rater 1/Test-taker 51)</li> <li>• However, overall, there is an <u>over-reliance</u> on source material. (Rater 2/Test-taker 58)</li> <li>• <u>The first two paragraphs are spent quoting the source material</u> rather than presenting an argument. (Rater 1/Test-taker 101)</li> </ul>

The qualitative results for Level 1 test-takers suggest that these test-takers exhibited some ineffective skills in addressing the integrated essay task. According to the qualitative analysis of the raters' comments on the Level 1 group, the raters tended to look for a clear position and argument, some originality of ideas, meaningful integration of source texts with accurate and explicit acknowledgements, explicit evidence of source text use, paraphrasing, and sufficient content coverage in written responses. Their comments suggest that many writing features were considered simultaneously.

### Group 2's integrated writing features

Figure 5 presents a display matrix that explains the nature of Group 2's integrated writing performance based on the raters' comments. Eight key features describe Group 2's integrated writing responses. While many of the negative features found in Group 1 remained for Group 2 (e.g., over-reliance on source texts, under/unclear development/position, no acknowledgements of source text use), some positive features have been observed (e.g., position/argument/development present, clear position). In Figure 5, unlike the matrix for Group 1, the matrix is organised from negative features to positive ones based on the number of code counts for ease of illustration and clarity of meaning-making.

According to the qualitative comments for Group 2, qualitative differences emerged between Groups 1 and 2's integrated writing responses. Nevertheless, more instances of negative features were observed in Group 2 compared to those observed in Group 1 essays. A plausible reason for this is that the raters were able to make more sense of their responses, given the increased word counts in Group 2. Furthermore, raters' comments suggested that some Group 2 test-takers did not explicitly provide the specific reference when using the source texts. They tended to copy original texts directly without appropriate acknowledgement. They also misunderstood the meaning of the source text that they used.

Figure 5: Group 2's features of integrated writing performance

Success level	Summary of rater comments	Examples
<b>Group 2</b> (n = 57 for Rater 1; n = 50 for Rater 2)		
<p><b>1. Under/unclear development/position (-)</b> (= 37 counts for Rater 1; 28 counts for Rater 2)</p>	<p>Test-takers at this level still have an issue with essay development and position. Their development or argument (e.g., cohesion) may still be unclear, confusing, partially supported or weak (e.g., specific examples needed). They might attempt to use some source texts to support their position but did this quite ineffectively. Furthermore, their poor development may be reflected in their conclusion (e.g., brief, ineffective, incohesive, confusing).</p>	<ul style="list-style-type: none"> <li>• There is an argument for both sides and the candidate makes references to the source material, but this is <u>done poorly</u>, especially in the last paragraph as they refer to the source as 'some researcher'. (Rater 1/Test-taker 19)</li> <li>• A somewhat <u>confused response</u>, referring to the 'advantages and disadvantages' and 'benefit effect' of global warming... (Rater 2/Test-taker 28)</li> <li>• There is an argument but there needs to be <u>more cohesion</u> between the main ideas presented and supporting ideas. (Rater 1/Test-taker 132)</li> </ul>
<p><b>2. No, little or minimal use of source texts (-)</b> (= 19 counts for Rater 1; 20 counts for Rater 2)</p>	<p>Similar to the Level 1 success group, Level 2 test-takers still tended to either use no source texts in their writing at all, or, if they did, they tended to use them little or minimally, insufficient to make a good argument. The essays appear to rely on test-takers' own knowledge and examples.</p>	<ul style="list-style-type: none"> <li>• There is <u>only one reference</u> to the source material, which is irrelevant to the point being made. (Rater 1/Test-taker 21)</li> <li>• Candidate <u>relies mostly on own knowledge of global warming</u>, rather than engaging in any meaningful way with the source texts. (Rater 2/Test-taker 22)</li> <li>• However, the writer <u>relies heavily on their own knowledge and does not cite the source material at all</u>. (Rater 1/Test-taker 39)</li> <li>• The highlighted reference suggests the candidate has at least familiarised themselves with Source 1, however there is <u>no explicit reference made to the source text/ author</u>. (Rater 2/Test-taker 57)</li> </ul>
<p><b>3. Ineffective/improper source text use (-)</b> (= 16 counts for Rater 1; 17 counts for Rater 2)</p>	<p>This is the same feature for Level 1 group. Test-takers may partially use source texts correctly or appropriately in one or two paragraphs, but may use them inaccurately, ineffectively and improperly in other paragraphs. This includes not acknowledging source texts in their essays or verbatim copying of original texts. This feature is related to 'no, little or minimal use of source texts', 'over-reliance on source texts' and 'no acknowledgements of source texts'.</p>	<ul style="list-style-type: none"> <li>• Last paragraph is inspired by Source (2) however <u>does not explicitly reference the source</u>. (Rater 2/Test-taker 19)</li> <li>• <u>Some relevant information</u> is provided from the source materials but this <u>seems forced</u>. (Rater 1/Test-taker 25)</li> <li>• Good attempts to engage with both of the source texts, however <u>acknowledgement of these needs to be clearer</u> (e.g., titles/ authors rather than just 'first article'). (Rater 2/Test-taker 67)</li> <li>• The sources are used but they <u>cause some confusion</u> rather than add to the argument. (Rater 1/Test-taker 177)</li> </ul>

<p><b>4. Over-reliance on source texts (-)</b> (= 18 counts for Rater 1; 4 counts for Rater 2)</p>	<p>Test-takers who used source texts tended to overuse or rely heavily on the source texts in their essays. This could involve directly copying the source texts, parroting the source use, presenting redundant information, or summarising the source texts. Rater 2 appeared to have fewer comments on this feature.</p>	<ul style="list-style-type: none"> <li>• The second main idea has some development but the <u>quote simply repeats the candidate's point</u>. (Rater 1/Test-taker 28).</li> <li>• Much of the essay is explaining the source material, which is <u>slightly overused</u>. (Rater 1/Test-taker 95)</li> <li>• <u>Over-dependence</u> on source material. (Rater 2/Test-taker 145)</li> <li>• <u>Clear reliance on the source texts</u>, however the material has been skilfully paraphrased. (Rater 2/Test-taker 175)</li> </ul>
<p><b>5. Insufficient responses (-)</b> (= 8 counts for Rater 1; 6 counts for Rater 2)</p>	<p>Although there were fewer comments for this feature for the Level 2 test-takers, some test-takers produced a short or short texts, making it insufficient to assess or understand their positions and arguments. These test-takers might have performed well in other assessment criteria that placed them into Level 2 overall.</p>	<ul style="list-style-type: none"> <li>• <u>193 words only</u>. (Rater 2/Test-taker 66)</li> <li>• There is a clear position, and the candidate uses resource material and their own viewpoint to support their argument, but <u>the response is under-length</u>. (Rater 1/Test-taker 78)</li> <li>• The response is short with limited development and there is minimal use of the sources. (Rater 1/Test-taker 107)</li> <li>• <u>190 words only</u>. More development needed, as well as more substantial use of source material (Rater 2/Test-taker 154)</li> </ul>
<p><b>6. Position/argument/development present (+)</b> (= 20 counts for Rater 1; 0 count for Rater 2)</p>	<p>Only Rater 1 commented on this positive feature. Test-takers' positions, arguments, and development of their essays were commented on as being present. Rater 1's comments tended to be somewhat optimistic about this performance. Rater 1, however, tended to 'but' after his statements. This suggests that essays were not fully developed in terms of positions and arguments. It might have been a reason why Rater 2 did not comment on this feature.</p>	<ul style="list-style-type: none"> <li>• There is a <u>position, but the response is too short</u>, so it does not develop. (Rater 1/ Test-taker 47)</li> <li>• <u>Expresses a position but with unclear conclusion</u>. (Rater 1/Test-taker 67)</li> <li>• There is a <u>position that is supported but the development is not clear</u> – some supporting evidence makes little sense. (Rater 1/Test-taker 77)</li> <li>• There is a <u>position</u> that provides support using some of the ideas from the text, <u>but nothing is acknowledged</u>. (Rater 1/Test-taker 80)</li> <li>• There is a <u>position, but the evidence needs to be stronger</u>. (Rater 1/Test-taker 159)</li> </ul>
<p><b>7. Clear position (+)</b> (= 9 counts for Rater 1; 7 counts for Rater 2)</p>	<p>Unlike the use of 'position present' in Category 6, both raters explicitly included 'clear position' in their comments. Some test-takers exhibited not just a position (as in Feature 6) but also did it quite clearly or explicitly in their essays. Their positions can be more easily identified and have some coherent organisation. The raters explicitly used 'clear position' in their comments.</p>	<ul style="list-style-type: none"> <li>• A very solid argument with a <u>clear position</u> presented. (Rater 1/Test-taker 39)</li> <li>• A <u>clear position</u> presented, including some good original examples. (Rater 2/Test-taker 41)</li> <li>• A <u>clear position with some good ideas</u>. (Rater 1/Test-taker 62)</li> <li>• A <u>clear position</u> taken by this candidate, who aligns themselves with Text 2. (Rater 2/Test-taker 78)</li> <li>• There is a <u>clear position with sound evidence</u>. (Rater 1/Test-taker 104)</li> </ul>

<p><b>8. Some relevant source text use</b> (= 4 counts for Rater 1; 6 counts for Rater 2)</p>	<p>Unlike test-takers at Level 1, test-takers at this level showed some evidence of selection and engagement of source text use relevant to the essay task and content. This includes an attempt to paraphrase source texts in their own words, acknowledging the source used. Test-takers appeared to be able to integrate source texts into their essays. However, their source text use was far from perfect (e.g., issues with imprecision).</p>	<ul style="list-style-type: none"> <li>• Good attempts throughout to <u>paraphrase source material</u> (though this does result in basic errors and imprecision). (Rater 2/Test-taker 19)</li> <li>• Some <u>relevant information</u> is provided <u>from the source materials</u> (Rater 1/Test-taker 25)</li> <li>• A good attempt to <u>include material from both source texts</u>, as well as the candidate's own examples (Rater 2/Test-taker 31)</li> <li>• The sources used are <u>somewhat relevant but there does not seem to be much connection</u> to the point the writer is making. (Rater 1/Test-taker 32)</li> </ul>
---	--	--

### Group 3's integrated writing features

Figure 6 presents a display matrix that explains eight features of the two raters' comments on Group 3's integrated writing responses. The finding in comments related to several negative features for this group might appear surprising for some readers, but the researcher expected to find this. Recall that many test-takers at this level did not have a high English proficiency as measured by the official IELTS test. That is, none of them were at an advanced proficiency level. Some test-takers might be placed in a border area between Groups 2 and 3 (based on the cut-off points being used). Test-takers who could well be either at the high end of Level 2 or at the low end of Level 3 would still face a challenge in fully stating a clear position and developing an essay. In addition to the eight features, it was observed that neither rater commented on the essays being short or insufficient. It seems that Level 3 test-takers did not have an issue with the length of their essay writing.

In summary, the qualitative analysis of Raters 1 and 2's comments on integrated essays based on success levels has allowed us to understand the similarities and differences between various features of essays produced by different groups of test-takers. Some negative features were found across the three levels (e.g., under/unclear development/position, over-reliance on source texts, little or no use of source texts). However, some positive features were found only in higher-level groups (e.g., clear position, integrated source texts with own knowledge).

Figure 6: Group 3's features of integrated writing performance

Success level	Summary of rater comments	Examples
<b>Group 3</b> (n = 54 for Rater 1; n = 33 for Rater 2)		
<p>1. <i>Under/unclear development/position (-)</i> (= 26 counts for Rater 1; 13 counts for Rater 2)</p>	<p>There were similar comments about issues related to under/unclear development/position to the above levels for Level 3 test-takers. Rater 1 provided more comments associated with this feature than Rater 2. However, the nature of their comments appeared to be somewhat less critical than for other success levels. Their comments regarding this feature were not stand-alone but were made after some strong points were provided. In particular, comments were related to the need to use the source texts in the essay development.</p>	<ul style="list-style-type: none"> <li>• but the <u>main ideas need more development</u> (Rater 1/Test-taker 40)</li> <li>• <u>Greater engagement with the source texts would have improved this response.</u> (Rater 2/Test-taker 44)</li> <li>• however, the body of this essay <u>requires further extension and development</u> ( Rater 2/Test-taker 55)</li> <li>• ... the candidate <u>just lists</u> the evidence rather than linking it to a main point (Rater 1/Test-taker 59)</li> <li>• <u>Some of the conclusions drawn (e.g., the very final sentence) are unwarranted.</u> (Rater 2/Test-taker 63)</li> </ul>
<p>2. <i>Ineffective/improper source text use (-)</i> (= 15 counts for Rater 1; 14 counts for Rater 2)</p>	<p>Some essays had a feature of ineffective/improper source text use. Test-takers might use only one source text, rather than both. Some might use both, but the extracted information was not effectively or appropriately used. The source texts might not be properly paraphrased. The feature was related to a lack of source text acknowledgement or insufficient or improper acknowledgements of the source texts in the essays.</p>	<ul style="list-style-type: none"> <li>• Acknowledges source texts via use of quotation marks, however <u>no references to titles/authors are provided.</u> (Rater 2/Test-taker 40)</li> <li>• Candidate <u>could have used Source 2 more comprehensively/</u> included <u>some rebuttal of arguments provided in Source 1</u> (rather than relying so much on 'personal perception'). (Rater 2/Test-taker 129)</li> <li>• <u>Source is just used separately in its own paragraph.</u> (Rater 1/Test-taker 143)</li> </ul>
<p>3. <i>Over-reliance on source text use (-)</i> (= 15 counts for Rater 1; 2 counts for Rater 2)</p>	<p>Both raters, especially Rater 1, commented on some over-reliance on source text use among some Level 3 test-takers. In particular, some test-takers tended to summarise source texts in their essays or lift from source texts without a deep engagement.</p>	<ul style="list-style-type: none"> <li>• <u>Lack of inverted commas despite chunks copied verbatim;</u> lifted from Smith: changes in the water cycle affect; melting ice was taking with; food supply for many animals; lifted from Burnett: no scientific agreement that global warming is a problem or that humans are its cause. (Rater 2/Test-taker 59)</li> <li>• The argument here is <u>more of a summary</u> of the sources texts which are used (Rater 1/Test-taker 108).</li> <li>• The <u>source text seems to dominate</u> the body paragraph (Rater 1/Test-taker 111).</li> </ul>
<p>4. <i>No/little/minimal source text use (-)</i> (= 4 counts for Rater 1; 0 count for Rater 2)</p>	<p>Only Rater 1 noted that few test-takers used source texts minimally in the Level 3 group. This feature, however, appears to overlap with comments in the 'Ineffective/improper source text use' category. This comment occurred when test-takers relied on their own knowledge in their essays.</p>	<ul style="list-style-type: none"> <li>• The source texts are <u>used and acknowledged but very briefly.</u> (Rater 1/Test-taker 59)</li> <li>• A clear plea for taking action against climate change, but <u>not much of the source material is used.</u> (Rater 1/Test-taker 60)</li> <li>• An argument is there with the candidate using mainly their knowledge and a <u>small part of the source material.</u> (Rater 1/Test-taker 88)</li> </ul>

<p>5. <b>Sound/clear argument (+)</b> (= 27 counts for Rater 1; 17 counts for Rater 2)</p>	<p>Unlike in other success levels, both raters used 'sound' and 'clear' arguments or positions in their comments for Level 3 test-takers. The comments were related to reasonable, convincing arguments and substantiated with test-takers' source text use and own sources. This feature suggests a positive quality of integrated essay writing for this success group.</p>	<ul style="list-style-type: none"> <li>• A <u>sound argument</u> is given for global warming (Rater 1/Test-taker 28)</li> <li>• This candidate demonstrates an impressive range of vocabulary appropriate to this context and provides some <u>excellent original arguments</u>. Greater engagement with the source texts would have improved this response. (Rater 2/Test-taker 44)</li> <li>• A <u>clear, convincing response</u>, with reasonable attempts to include examples from the source texts. (Rater 2/Test-taker 48)</li> <li>• The candidate makes a <u>solid argument</u> (Rater 1/Test-taker 129)</li> </ul>
<p>6. <b>Some relevant use of source texts (+)</b> (= 15 counts for Rater 1; 7 counts for Rater 2)</p>	<p>Some test-takers used relevant source texts to a level that enhanced essays, including good paraphrasing in their essays. This feature suggests that test-takers were reading at least one or both texts, extracting key information to support their essay points.</p>	<ul style="list-style-type: none"> <li>• <u>Great paraphrasing</u> of key ideas, some good, less common lexis. (Rater 2/Test-taker 38)</li> <li>• A sound attempt is made making an argument <u>using the resources</u>. (Rater 1/Test-taker 48)</li> <li>• The candidate makes an <u>excellent attempt</u> not only to include the source material but also to comment on its validity. (Rater 2/Test-taker 76)</li> <li>• There is <u>some good use of the source material</u>. (Rater 1/Test-taker 127)</li> <li>• <u>Some good paraphrasing of the source material</u>; however, this is confined to the one paragraph. Fuller use of the sources, including Source 2, could have improved this response. (Rater 2/Test-taker 143)</li> </ul>
<p>7. <b>Position present (+)</b> (= 12 counts for Rater 1; 0 count for Rater 2)</p>	<p>Only Rater 1 indicated that Level 3 test-takers presented a position or argument in their essays. This feature is partially positive since it demonstrates that test-takers presented a position, but it may lack some support or good arguments. Rater 2 did not mention this feature. The comments suggested that while it is desirable for this task, test-takers needed to extend their positions and arguments.</p>	<ul style="list-style-type: none"> <li>• There is a <u>position, but</u> much of the supporting detail is the source, leaving little from the candidate (Rater 1/Test-taker 55).</li> <li>• A <u>clear plea for taking action</u> against climate change, but not much of the source material is used (Rater 1/Test-taker 60)</li> <li>• <u>An argument is there</u> with the candidate using mainly their knowledge and <u>a small part of the source material</u>. (Rater 1/ Test-taker 88)</li> <li>• <u>There is an argument there using the source texts, but more input is required</u> from the candidate. (Rater 1/Test-taker 112)</li> </ul>
<p>8. <b>Integrated source texts with own knowledge (+)</b> (= 8 counts for Rater 1; 7 counts for Rater 2)</p>	<p>Both raters commented on an effective use of source text integration with test-takers' own knowledge or source in the essay task. This comment did not appear in comments for Level 1 and 2 test-takers. Essays with this commentary feature show a good balance between test-takers' knowledge and information extracted</p>	<ul style="list-style-type: none"> <li>• A good attempt at including <u>both source material and original language</u>. (Rater 2/Test-taker 40)</li> <li>• A sound argument with <u>both sources mentioned, with the writer also adding their own knowledge</u>. (Rater 1/Test-taker 91)</li> <li>• The writer uses a <u>combination of their own knowledge/opinions and the source material</u> to develop an argument. (Rater 1/Test-taker 147)</li> <li>• Good paraphrasing of source material and <u>integration into the candidate's own arguments</u> (Rater 2/Test-taker 158)</li> </ul>

In summary, the current research question has found quantitative and qualitative differences between the three groups of test-takers' performances in the integrated writing task. The statistical analyses suggest that more successful groups outperformed the less successful groups in nearly all assessment criteria. The findings allow us to understand a unique characteristic of an integrated task, which results in variation in writing performances.

## 6.5 Research question 5: What is the nature of test-takers' reported reading, writing and test-taking strategies during independent and integrated writing tasks?

This research question aims to explore the content, cognitive, convergent validity of the integrated writing task. It investigates the extent to which test-takers engaged in the writing task. It also explores the extent to which reported writing and test-taking strategies between the two writing tasks were related to one another. Recall that test-takers were asked to complete a reflective questionnaire after each writing task (see Research Method). In addition, the interview data from 10 participants were used to explore the qualitative nature of their test-taking processes and strategy use.

### 6.5.1 Quantitative findings

Table 25 presents the descriptive statistics of items in the Reflective Questionnaire 1 (Independent). Table 26 presents the percentages of responses for each item (the highest percentage for each item is shown in bold). The mean ranges from 3.380 (Item 1) to 4.310 (Item 11). In validation research, it is essential to investigate whether processes that underlie the test constructs are perceived to be used by test-takers.

**Table 25:** Descriptive statistics of the reflective questionnaire 1 (N = 154)

No.	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
1.	1	5	3.380	1.061	-0.343	-0.374
2.	1	5	4.070	0.994	-1.157	1.073
3.	1	5	3.450	1.015	-0.469	-0.098
4.	2	5	3.620	0.906	-0.034	-0.799
5.	1	5	3.550	0.995	-0.433	-0.277
6.	1	5	3.560	1.126	-0.474	-0.462
7.	1	5	3.540	1.041	-0.302	-0.559
8.	1	5	3.950	0.951	-1.029	1.102
9.	1	5	3.550	1.034	-0.725	0.297
10.	1	5	4.180	0.931	-1.475	2.665
11.	1	5	4.310	0.978	-1.646	2.620
12.	1	5	3.640	1.020	-0.666	0.114
13.	1	5	3.950	0.897	-1.010	1.543
14.	1	5	3.430	1.221	-0.294	-0.925

Table 26 shows few participants reported disagreements with the items, while many reported an average-to-strong agreement. Test-takers tended to identify themselves as engaging with the cognitive activities described in the questionnaire. Some responses were worth noting. Six items with at least 75% agreement on each statement were Items 1, 2, 8, 10, 11 and 13.

**Table 26:** Frequency counts of the reflective questionnaire 1 (N = 154)

No.	Item	1	2	3	4	5
1.	I made an outline of what to include in my writing.	8 (5.2%)	21 (13.6%)	50 (32.5%)	<b>52</b> <b>(33.8%)</b>	23 (14.9%)
2.	I wrote using my own words.	4 (2.6%)	9 (5.8%)	19 (12.3%)	<b>62</b> <b>(40.3%)</b>	60 (39%)
3.	I organised my writing to meet the expected academic writing style.	7 (4.5%)	17 (11%)	49 (31.8%)	<b>60</b> <b>(39%)</b>	21 (13.6%)
4.	I made sure I supported my ideas with clear points, principles and examples.	0 (0%)	16 (10.4%)	54 (35.1%)	<b>55</b> <b>(35.7%)</b>	29 (18.8%)
5.	I checked to make sure I used appropriate words.	4 (2.6%)	19 (12.3%)	43 (27.9%)	<b>63</b> <b>(40.9%)</b>	25 (16.2%)
6.	I checked to make sure I didn't have any spelling mistakes.	8 (5.2%)	18 (11.7%)	43 (27.9%)	<b>50</b> <b>(32.5%)</b>	35 (22.7%)
7.	I checked to make sure I had produced sufficient words as required by the task.	4 (2.6%)	21 (13.6%)	47 (30.5%)	<b>52</b> <b>(33.8%)</b>	30 (19.5%)
8.	I used my knowledge and experience to support my points of view.	4 (2.6%)	8 (5.2%)	24 (15.6%)	<b>73</b> <b>(47.4%)</b>	45 (29.2%)
9.	I acknowledged sources of information in my writing.	9 (5.8%)	11 (7.1%)	43 (27.9%)	<b>66</b> <b>(42.9%)</b>	25 (16.2%)
10.	I made sure I understood the test instructions.	5 (3.2%)	2 (1.3%)	18 (11.7%)	63 (40.9%)	<b>66</b> <b>(42.9%)</b>
11.	I was aware of the time limitation of the test.	5 (3.2%)	3 (1.95%)	17 (11%)	43 (27.9%)	<b>86</b> <b>(55.6%)</b>
12.	I evaluated my performance while doing the test task.	6 (3.9%)	14 (9.1%)	38 (24.7%)	<b>67</b> <b>(43.5%)</b>	29 (18.8%)
13.	I considered whether my writing met the task requirements.	4 (2.6%)	4 (2.6%)	29 (18.8%)	<b>76</b> <b>(49.4%)</b>	41 (26.6%)
14.	I checked my writing before submitting.	10 (6.5%)	28 (18.2%)	37 (24%)	<b>43</b> <b>(27.9%)</b>	36 (23.4%)

The items in the questionnaire were then used to create two composites (based on the questionnaire taxonomy):

- 'Independent writing strategies' based on Items 1 to 9
- 'Independent test-taking strategies' based on Items 10 to 14.

The Cronbach's alpha coefficient of these two composites was 0.85. The Pearson correlation coefficient between the two composites was 0.746 ( $R^2 = 0.56, p < 0.01$ ).

The finding suggests a strong linear relationship between the two groups of strategies in this independent writing task.

Table 27 presents the descriptive statistics of items in the Reflective Questionnaire 2 (integrated). Items 16 to 19 will be reported in Research Question 6. The mean ranged from 3.070 (Item 4) to 4.280 (Item 12). Table 28 presents the frequency counts and percentages of participants' responses for each item. The patterns were similar to those for the independent writing reflective questionnaire in that few respondents disagreed with the statements, and several agreed or strongly agreed.

**Table 27:** Descriptive statistics of the reflective questionnaire on the integrated writing task (N = 154)

Item	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
1.	1	5	3.850	1.008	-0.516	-0.633
2.	1	5	4.090	0.876	-0.887	0.570
3.	1	5	4.080	.939	-0.792	-0.044
4.	1	5	3.070	1.137	0.061	-0.658
5.	1	5	3.990	0.880	-0.578	-0.083
6.	1	5	3.360	1.032	-0.319	-0.229
7.	1	5	3.660	1.029	-0.507	-0.156
8.	1	5	3.680	0.966	-0.714	0.434
9.	1	5	3.660	0.906	-0.366	-0.102
10.	1	5	3.890	0.878	-0.507	-0.079
11.	1	5	4.150	0.820	-0.945	1.421
12.	1	5	4.280	0.872	-1.254	1.599
13.	1	5	3.780	0.929	-0.793	0.725
14.	1	5	3.860	0.899	-0.831	0.873
15.	1	5	3.520	1.234	-0.646	-0.424

**Table 28:** Frequency counts of the reflective questionnaire 2 (N = 154)

No.	Item	1	2	3	4	5
1.	I briefly skimmed the texts before I read more carefully.	1 (0.6%)	17 (11%)	34 (22.1%)	<b>55</b> <b>(35.7%)</b>	47 (30.5%)
2.	I considered what the author(s) meant in the texts.	1 (0.6%)	8 (5.2%)	23 (14.9%)	<b>68</b> <b>(44.2%)</b>	54 (35.1%)
3.	I used my prior knowledge and experience to help me understand texts.	1 (0.6%)	9 (5.8%)	29 (18.8%)	54 (35.1%)	<b>61</b> <b>(39.6%)</b>
4.	I took notes from the texts.	13 (8.4%)	34 (22.1%)	<b>56</b> <b>(36.4%)</b>	31 (20.1%)	20 (13%)
5.	I identified the main ideas in the given text.	1 (0.6%)	6 (3.9%)	36 (23.4%)	<b>62</b> <b>(40.3%)</b>	49 (31.8%)
6.	I made an outline of what to include in my writing.	8 (5.2%)	19 (12.3%)	<b>57</b> <b>(37%)</b>	50 (32.5%)	20 (13%)
7.	I paraphrased the texts.	5 (3.2%)	13 (8.4%)	46 (29.9%)	<b>55</b> <b>(35.7%)</b>	35 (22.7%)

8.	I compared and synthesised multiple points of view from the written texts to help me write.	5 (3.2%)	11 (7.1%)	40 (26%)	<b>70</b> <b>(45.5%)</b>	28 (18.2%)
9.	I made sure I supported my ideas with clear points, principles and examples.	2 (1.3%)	12 (7.8%)	49 (31.8%)	<b>64</b> <b>(41.6%)</b>	27 (17.5%)
10.	I acknowledged source information from the given texts in my writing.	1 (0.6%)	8 (5.2%)	38 (24.7%)	<b>67</b> <b>(43.5%)</b>	40 (26%)
11.	I made sure I understood the test instructions.	2 (1.3%)	1 (0.6%)	27 (17.5%)	<b>67</b> <b>(43.5%)</b>	57 (37%)
12.	I knew the time limitation of the test.	2 (1.3%)	3 (1.9%)	22 (14.3%)	51 (33.1%)	<b>76</b> <b>(49.4%)</b>
13.	I evaluated my performance whilst doing the test task.	4 (2.6%)	9 (5.8%)	36 (23.4%)	<b>74</b> <b>(48.1%)</b>	31 (20.1%)
14.	I considered whether my writing met the task requirements.	3 (1.9%)	8 (5.2%)	32 (20.8%)	<b>76</b> <b>(49.4%)</b>	35 (22.7%)
15.	I checked my writing before submitting.	16 (10.4%)	12 (7.8%)	38 (24.7%)	<b>52</b> <b>(33.8%)</b>	36 (23.4%)

In Table 28, four items had at least 75% agreement among the participants: Items 2, 3, 11 and 12. The items in the integrated questionnaire were then used to create three composites:

- 'Integrated reading strategies' based on Items 1 to 5
- 'Integrated writing strategies' based on Items 6 to 10
- 'Integrated test-taking strategies' based on Items 11 to 15.

Pearson correlations suggest that integrated reading strategies were correlated with integrated writing strategies ( $r = 0.753$ ,  $R^2 = 0.57$ ,  $p < 0.01$ ) and integrated test-taking strategies ( $r = 0.727$ ,  $R^2 = 0.53$ ,  $p < 0.01$ ). Integrated writing strategies were correlated with integrated test-taking strategies ( $r = 0.693$ ,  $R^2 = 0.48$ ,  $p < 0.01$ ). The strength of the correlation was smaller but similar to that between independent writing strategies and test-taking strategies ( $0.746$ ,  $R^2 = 0.56$ ).

### Correlations between independent and integrated strategies

Since it is important to know the extent of the relationship among reported strategy use in the independent and integrated writing tasks to know test-takers engaged similarly in core strategic processes during the two different tasks, a Pearson correlation was also performed. The composites from both reflective questionnaires were used in the analysis. Table 29 presents the Pearson correlations among writing processes in the independent and integrated writing tasks. It was found that all variables were statistically correlated to one another ( $p < 0.01$ ). In particular, independent writing strategies were strongly related to integrated reading, writing and test-taking strategies ( $r = 0.523$ ,  $R^2 = 0.27$ ,  $p < 0.001$ );  $r = 0.561$ ,  $R^2 = 0.31$ ,  $p < 0.001$ ); and  $r = 0.611$ ,  $R^2 = 0.37$ ,  $p < 0.001$ ), respectively. Independent and integrated test-taking strategies were also strongly correlated ( $r = 0.686$ ,  $R^2 = 0.47$ ,  $p < 0.001$ ). The findings suggest that test-takers' perceived strategic processes were interrelated, implying the cognitive validity of the integrated writing task.

**Table 29:** Pearson correlations between independent and integrated strategies (N = 154)

		Independent test-taking strategies	Integrated reading strategies	Integrated writing strategies	Integrated test-taking strategies
Independent writing strategies	Pearson Correlation	0.746**	<b>0.523**</b>	<b>0.561**</b>	<b>0.611**</b>
	Sig. (2-tailed)	<.001	<.001	<.001	<.001
Independent test-taking strategies	Pearson Correlation	1	<b>0.555**</b>	<b>0.468**</b>	<b>0.686**</b>
	Sig. (2-tailed)		<.001	<.001	<.001
Integrated reading strategies	Pearson Correlation		1	<b>0.753**</b>	<b>0.727**</b>
	Sig. (2-tailed)			<.001	<.001
Integrated writing strategies	Pearson Correlation			1	<b>0.693**</b>
	Sig. (2-tailed)				<.001

### 6.5.2 Qualitative findings

The content analysis of test-takers' interviews suggests the nature of eight common strategic processes involved in the integrated writing task: assessing the situations, reading comprehension, outlining, integrating source texts, paraphrasing, acknowledging source texts used, monitoring, and appraisal.

#### Assessing the situation

This refers to a strategic process to learn and comprehend writing task conditions and requirements so that performance goals can be set. This results in planning how to proceed with the task and evaluation of performance. The qualitative analysis suggests that test-takers spent more time understanding the integrated writing task than the independent writing task, as they were familiar with the independent writing task. Less successful test-takers tend to spend less time understanding the task requirements, especially for the integrated writing task. This resulted in an essay that did not meet the expected requirements. Cui, for example, did not engage in understanding the integrated writing requirements and jumped to read the two passages before starting to write. She realised that she did not understand the integrated task requirement.

I misunderstand the requirement. I thought maybe I'd do a summary or just two. And I'm not cite. (Cui)

Yan reported spending time trying to understand the integrated writing task. However, he thought he could choose the text he agreed with in his response. He said he had realised that he needed to spend more time reading Text 1.

I just read the two readings quickly. So, I to know what I'm reading and write about and what's the point. How it's so. So, then I just write the first paragraph and then when I write the second paragraph. I agree to the reading one again and to find some evidence.

Liang reported going through the task, planning and conceptualising what to include in his essay, but more so by conceptualising ideas than by reading the source texts. It turned out that his essay read like a summary of the two texts.

After reading the questions, I just started to think about what I'm going to say, to choose, to think of the ideas I want to say, because I yeah I started to think about my own opinion ... So, after reading the questions, I started to read the passage.

---

To get the main idea of those to it. And then I just tried to use the ideas from that passage to support the things that I want to write. (Liang)

Zhao also expressed a lack of certainty regarding how to use the source texts in his writing.

So, I still read the instruction at first, like, what should I do or should I do that in the next 15 minutes? And I should. And I then I start to look at my time, like how many how many times [how much time] should I spend on reading the text? And what should I be doing next? (Zhao)

More successful test-takers showed an awareness of the integrated writing task requirements. Zhao reported that he had been concerned about some of the task conditions, such as the time and word limits (i.e., trying to understand the task and question and conceptualising what to do to address the task). He was conscious of the time allowance and how that may affect his ability to monitor his writing.

The first thing I was thinking is the time and the word minimum requirement and I have to think about whether I can finish in time. And then I was checking the work that I am gonna write, like the argument, their response or what then about a topic I was reading about the topic and about the opinion.

Lu indicated that the processes of addressing the integrated writing task were similar to those she had employed while completing the independent writing task.

I just took a look through the writing requirements and then read to the ... all the reading materials. Yes. And then outline the structure. But when I outlined a structure for this task, it just puts me a little bit more time. So, during the following writing, I must, you know, time my time. So had a last minute. Yes. I still wrote the last sentence. (Lu)

Teng reported trying to understand the writing task requirements and brainstorming ideas initially because he had not completed this task before.

I read the question first, like, I'll do the whole thing. And then I thought about it and then I started to brainstorm. So, there were a lot of different ideas popping up, like why I supported this or why I didn't support that... It took me some time to actually get through the instructions. (Teng)

Similarly, Wang spent time trying to understand the integrated writing task requirements. However, it did not take him long to determine his position on the topic.

Yeah, so I started reading. "Some people believe that global warming is damaging our planet. Others believe that it's not serious. Do you agree or disagree?" Which point do you agree? So, it's an argument I need to support myself and give reasons to support. (Wang)

Lizzy initially thought about the question and what position she would take. Lizzy also reported spending time trying to understand the statement and the task requirements and that doing so helped her formulate her essay's direction.

So, I kind of just assessed the question to start with. So, I was like, do I agree with that? Do I not agree with it? What kind of stance am I going to take? So then once I'd made my mind up about that, I decided I was like, I do want to say that it's important. So I'm going to talk about how important it is.

---

So I read it and like same type of thing. Like, I read the statement, and then I said I was like, OK, which way am I going to like? I like the moment I read it, I already knew kind of which kind of argument I would kind of agree on. (Lizzy)

Sam decided on the structure of his essay and the positions he would adopt soon after he understood the task requirements. He engaged in planning and mental outlining processes. In the integrated writing task, Sam reported that he had only quickly tried to understand the task requirements and the question being asked.

I only skimmed through reading this part here, and then I got down to this part here and made sure that I thought I had a pretty good understanding of the question ... And I thought about, like, but my stance on global warming, but more importantly, like which would be easier to write about, whether I agree that it's an issue or whether I disagree, and in the end, I decide that I agree that it's an issue partly because personally I do, but also partly because I was able to find examples in the source texts which could support my idea. (Sam)

The content analysis, therefore, indicates that assessing the situation, in particular by attempting to understand the integrated writing task requirements and conditions, is critical to success. There were some differences between the more or less successful test-takers.

### Reading comprehension

This process is related to comprehending the two source texts. All interviewees reported reading the source texts in the integrated writing task, but less successful test-takers tended to skim and scan the texts (expeditious reading).

First, I go through the passage and find some examples in the passage which I will use as a supporting idea in my writing and use to my viewpoint. (Cui)

Yan reported that he had skimmed the two texts but decided to focus only on using the first text in his response.

So, I read the two readings and found what points I could use in writing. Mm hmm ... I read that global warming is damaging, it's damaging our planet and so, I find some references and also find the reason. After reading, I found reading one supports that global warming is damaging our planet, and reading two says it's not a serious problem. So, I read, the more I read and I think it's spend more time on reading one to find some evidence. (Yan)

More successful test-takers tended to read more carefully and extract relevant information more accurately than less successful ones. Zhao, for example, reported spending about 8 minutes reading the source texts initially. Zhao believed that his reading was stronger than his writing, so he was confident in his reading ability and engaged actively in reading the source texts. Zhao questioned the credibility of the claims made in the second source text.

And I start to read paragraphs of reading material two. I think about how should I debate this kind of view, what is the logical problem in this paragraph, and why does this author offer the evidence? ... And then I noticed the time of the paragraph was published in 2001 and which is earlier than the first reading paragraph. I think it might be some further study on this problem and there are some more brand new data instead of 2001. (Zhao)

---

Zhao reported that reading helped him include details of the source materials used (author and date) in his writing.

So, at that, at the very beginning, I was making sure the author and the time in order to make reference more clearly. And so, I started really just kind of reading and I confirmed the authors' view is that climate change is caused by human activity. Oh, and also I saw that the title is 'undeniable' and I can make sure that this author is supporting the first view ... So, I start why and what kind of evidence he has. And I started looking for the evidence like he was talking about blah, blah, blah, blah, blah. And then I saw this guy, and he said, he's famous, blah, blah, blah, and indicates according to his new computer models and instead of this one is well on the way. So, this is the evidence one I think I could use for supporting my view. And I also saw this and I noticed that this guy is that is the second author about the reading material. And he also explained why this author's view is wrong. And I think I can use them later too. (Zhao)

Lu reported spending about 7 to 8 minutes understanding the task requirements and reading the two source texts. She reflected that she needed to improve her reading speed after completing this task.

So, you just read as maybe at least, maybe up to five minutes, I think. So, for me, I think my reading speed is not very quick, not very quickly. Yes, I indeed spent a lot of time reading. So that is one of the problems for this kind of writing for me. (Lu)

Teng reported spending around only 4 minutes reading the two texts initially, as he knew he would be able to return to them later in the test. He admitted that having to read two conflicting articles was challenging for him.

Key challenges, I, yeah, I feel like I was a little bit confused because, you know, those two articles are saying the opposite thing, right? So, I have to look for evidence to prove one of them, one of them is wrong. I think that part was a little bit challenging ... I read all the two articles, some as I mentioned, and I actually took some notes. I believe that's what I meant. Because you cannot really circle circle ... For the second text, we have a lot of different numbers, right, so I started to look at the details here ... because I think what like the general idea of the second text actually wrong. So, I start to look at the possible mistakes that that they have within that text. (Teng)

Ren explained how she had approached the two reading texts.

I first read the topic and then I scanned the two articles very quickly and I actually summarised the like the different opinions of the two articles. And then, according to your point, mentioning the two articles, I asked myself, based on the facts mentioned, that here in the articles, which then incorporate, they're saying, so what? What do I think of the topic? I think that's how I deal with this kind of essay. (Ren)

Wang explained that his focus while reading had been to comprehend general ideas rather than to memorise information.

Yeah, well, I didn't scan the text or skim the text. I just started reading carefully from bit to bit, sentence to sentence. So I didn't try to remember anything, memorise anything. I only tried to understand the text, like understand and comprehend what it says. (Wang)

Lizzy reported that she had skimmed over the two readings soon after to get a sense of the information presented. It allowed her to note key points to include in her essay and decide which parts of the texts to read carefully. She determined how she would use the source texts to support her arguments.

---

In my head, I kind of like picked out some of the key points that I think were important, and then I wrote down. So then after skimming it and then kind of wrote down some key points for each of them. And in that writing down those key points gave me more of a chance to kind of take it like read it properly, like read it, not just skimming it, like reading it in depth. And then so I wrote down my key points and then I started writing. (Lizzy)

Sam reported that he initially did not read the two source texts in detail but selectively focused on some parts of the texts as he wrote the essay.

No, I skimmed read them and then I started writing and then I went back and read intensively some parts to find like one I could use to support my idea. (Sam)

It was found that an integrated writing task requires both *expeditious and careful reading*, critical reading and the ability to discriminate between more or less relevant information.

### Outlining

Outlining is related to planning and brainstorming. Many test-takers reported constructing an essay outline before commencing their writing, except Cui, who did not explain how she outlined her essay. Cui responded to the integrated task *without an essay outline*. Yan reported he partially outlined by taking notes after reading the texts. He did not spend time developing an outline or listing ideas to write about explicitly.

And yeah, on outline in some sentence, especially some strong sentence after the name, some professor's name, for example, is someone states that blah blah... and I will point my view. (Yan)

It appears that Liang recognised the importance of planning before writing. He, however, reported that he had not written an essay outline on paper; his outline was in his mind. He explained this as follows.

Yeah, I just think the idea is in my mind. Yeah, I have to write something because you don't have much time in terms of taking the IELTS test. You know, you need to really have really good timing for planning. So many students cannot finish the tasks. (Liang)

Zhao planned his essay, but he had not written an explicit outline. He explained his plan.

So, I just trying to make a very general statement. So, the first challenge is to decide whether to write a very general paragraph like academic writing. We should introduce about the background of the problem ... So, after I finish that instruction, I start to look at the topic. And it was about the global warming. So, I started about the point. Like the first point about global warming is damage to our planet, where other opinion argues that it's not a serious problem. So, I have to first I have to answer it, at which point of view I agree with and why and I should give a reason and support them. So this is typical. I agree. (Zhao)

Wang pointed out that he had started to write straight after he thought he had understood what was required of the task, and had not made an explicit plan, such as a written outline. He had located which text supported or did not support the climate change argument. Instead, he reported using mental planning. He also did not have enough time to review his essay before submitting it.

Yeah, I'm quite confident in my writing, so I think I have to plan in my mind. So, I don't really need to write it down because that takes time and I'm quite tight on time. (Wang)

---

Sam mentioned that he had not written an outline on paper, but had a mental outline. This choice might have been strategic due to the test-taking conditions.

I didn't have a physical outline except for the comments I wrote at the top. I definitely had a **mental outline** in my head. (Sam)

It was found that outlining – whether it is written on a paper or constructed mentally – is essential for addressing an integrated writing task. It is also the quality of an outline that might differentiate between more or less successful test-takers.

### Integrating source texts

Integrating source texts in writing is related to discourse synthesis processes. It is about combining relevant information from the provided source texts with one's ideas or positions. Less successful test-takers reported a challenge to use the source texts in their writing. Liang, for example, found it hard to integrate the source texts into his writing and he ended up summarising them in his essay. He found it difficult to produce his own ideas, possibly due to the plethora of ideas in the source texts. This weakened his essay performance. He explained that the time limit also put pressure on his integration, making it hard to develop his own ideas. He also reported revisiting the source texts several times during his writing.

However, when I was writing the essays, I came back to read passage again to find out the details. So maybe there's something interesting that you wanted to use for your writing. (Liang)

More successful test-takers attempted to synthesise information in their writing. Teng mentioned how he had attempted to synthesise information from the source texts to mix with his own ideas.

I think to me, I think, like, global warming is definitely an issue and it's definitely happening. So, I look closely at what they said here in the first text to maybe see for evidence for which can actually help me. So, I actually mixed out some of the useful information and which can serve my writing. (Teng)

Teng pointed out that he did not rely on the first text in his writing, but used the second text to strengthen his arguments.

And then I also, I also think about yeah, there's actually another article. So it has to I have to do something with it. So, I was thinking like, yeah, maybe I can say like something like yes, I think "global warming is real" and then here's this and not those evidence. But, however, some people think differently if think global warming is like a Hoax or something like that. And then I can come back again to to argue against them. (Teng)

What Teng described above is related to discourse synthesis, which is required when completing an integrated writing task. Ren also reflected on the importance of synthesis in her responses to this writing task. Still, she was concerned that during the task completion, this would be time-consuming, causing her not to perform as well as she otherwise might.

Because when I came to the second integrated task, because I know I have to synthesise and, you know, syntheses are always, you know, one of the things that I'm not sure if it's only me, I guess students will feel a little bit worried to do that because it takes a lot of time. And you have to, you know, weigh between, for example, others' opinions, your opinion and how you balance the two or the three and four. So, I have this kind of feeling that I might not do well. (Teng)

---

It was observed that integrating information successfully requires good synthesis skills that consider both the source texts and one's own ideas and positions.

### Paraphrasing

Paraphrasing requires expressing someone else's information in one's own and acknowledging the source. Zhao had tried paraphrasing the ideas in the provided texts and using his own words rather than simply quoting.

And in the first paragraph, I think I was trying to write down here is the impact of a global warming on some animals that suffer from them. And I gave the evidence based on the text one and also saw additional evidence from the Tim Barnett's research. (Zhao)

Lu also reported that she had paraphrased the information from the source texts rather than quoting from them. This was linked to her previous learning experience.

Oh, yeah, I paraphrased them. Yes. I think this is, also, the, helps from my writing assignments as well. Yeah. Because a lot of professors told us, yeah, you can paraphrase or you can say yes in the brackets, but you can't use others' points directly in your writing that is plagiarism. (Lu)

Teng reported going into various aspects of each reading to understand and extract relevant information. Teng reflected that he had synthesised and paraphrased the source texts.

For me, basically, I would first look at the numbers and also like how or what they did for their research or case studies to actually judge whether their evidence is reasonable or is actually ... Yeah, I think there're some of the words that popped up to me and which actually made me to choose, to choose those. Yeah, those sentences, to paraphrase my own evidence ... And then I paraphrase their ideas and then I tried to argue against them to say, like, actually yeah, they are actually wrong to my view. I used my own words to explain that to my audience. (Teng)

Ren preferred to paraphrase rather than use direct quotes from the texts. She also explained that paraphrasing showed that she understood a point she wanted to make.

My first thought of kind of task is to cite. I prefer to paraphrase because I feel paraphrase is more like, you know, after I understand what it's about. And it's kind of the way I see it, the way I read it. So, I want to incorporate this kind of phrase into my own essay. (Ren)

Lizzy reported using quotes in her essay rather than paraphrasing extracts from the source texts.

I mainly used direct quotes. The general idea for it, you know, I think actually mainly just quote I was going to use, I was going to paraphrase when it was. I don't know. I was going to paraphrase something, but then I realised that I could just like quote half of what they've said. (Lizzy)

Interestingly, in terms of source text use, Sam reported that he had copied and pasted selected text from the source texts into his working space. He then paraphrased them instead of using them directly.

I actually, like, selected the line. I copied and pasted it into my essay, and I put it like that below my paragraph so I could read it and then I tried to paraphrase it. I think it's kind of like soft paraphrasing. I didn't spend a lot of time changing it, but I did change it. (Sam)

---

It was observed that paraphrasing skills are significant to successful integrated writing, and more successful test-takers exhibited these skills.

### Acknowledging source texts used

In academic writing, it is important to acknowledge the sources of information being used in an argument. In both direct quotes and paraphrasing, acknowledgement is needed. This awareness was expressed more often among more successful test-takers. Liang pointed out that he had been aware of the need to acknowledge the source texts he used in his essay.

Oh, I have to acknowledge that. Like a special training at the university. So every time when I wrote something academically, I always tried to find the sources and the references. Yeah. So I need to make sure that I follow the norms of writing some academic essays. (Liang)

Lu reported that she revisited the texts when she forgot the source of the text, such as the person who made a particular point. She showed an awareness of the need to acknowledge the source of the information used in her essay, but at the time of her writing, she did not have time to think deeply about the topic and source texts.

So, I know I must cite now if I use something that we need to write. Yes. Yes. It brings me the first impression is, is this are tasks about, you know, geography or human activities just like that. And I didn't think deeply or more about the source texts. (Lu)

Although Ren tried to vary her source text use with some direct quotes, during the stimulated recall, she realised that she had forgotten to acknowledge the source texts in part of her essay.

Of course, as you can see in a second or in the second paragraph, I, I directly cited. I mean, I put it into quotation marks, but I guess I forgot to, you know, put a name ... I directly quoted the guy who mentioned changes in the water cycle "Though some scientists argue that the current rate of increase is slower than the average rate over the 18,000-year period" I quoted, you know, I directly quoted the author, but I forgot to put a name after that sentence. (Ren)

Lizzy reported experiencing uncertainty in how to cite the quoted information appropriately, e.g., she was unsure whether to provide details of the author of the text quoted or the person the author mentioned in the text.

It's hard to know when to quote the person or when to quote the person who they're quoting type of thing, you know what I mean? (Lizzy)

The quality of the source texts (ambiguity of who to cite) can be an issue in the current integrated writing task. Sam was uncertain about whether he should mention the author of the source text, or the author cited by the source text. Sam decided to cite the main author.

And I didn't pay too much attention to that because I wasn't sure if, like, the writer, so to speak, would care about the correct citations. I just cited Burnett 2001. But actually, I think I'm citing other people who Burnett wrote about it. So I, of course, I could have said like as cited Curry 2001, but I didn't go to that trouble maybe because I thought it's like else, are they be testing APA style here. But of course, in an assignment, yeah, I would do like a cite to be like as a secondary source. (Sam)

It was observed that a good understanding of citations and acknowledgements of source texts in writing was a distinctive characteristic of more successful test-takers.

---

## Monitoring

Monitoring is related to checking ongoing writing processes in relation to the task conditions including time limits and word counts. The analysis suggests that less successful test-takers did not engage in self-monitoring, compared to more successful test-takers. For example, Cui's self-assessment appears not well aligned with her writing processes. Yan reported that he did not check his writing due to the time limit but said he would in an official test. Liang's ability to self-monitor was affected by anxiety in online writing. He commented that he preferred to handwrite his essays rather than type them on a computer. He mentioned that while he was familiar with computers and that providing answers on a computer could be suitable for both test-takers and examiners since some test-takers might have messy handwriting, he felt anxious about taking computer tests.

However, when I was reading while I was writing, I think the computer-delivered test is not really user-friendly; I have to say because it's more complex and troublesome, I think it can be scary for those who are not used to using computers. (Liang)

Lu reported she engaged in time monitoring. She felt anxious as the end of the test approached and made some strategic decisions to cut back on what she wanted to write.

When I realised, oh, there are several minutes left or so. There are ten minutes left, so I must type as fast as I could. I feel very, very, a little bit uncomfortable. And it's that I compile the examples or reduce my writing number of words as well. (Lu)

Teng felt he did not have enough time to monitor his writing.

I felt like when I was typing, there must be some kind of typos because I typed like quickly. Yeah. In the end, I didn't have like, like a long enough time to, to actually check all the word spellings of each single word. So, yeah, I guess there are spelling mistakes and because, you know, in real life computer use, there's like a spelling correction, either Microsoft Word or Grammarly, those kinds of applications that will help you to find the spellings by now. (Teng)

Teng's reflection on computer applications for checking spelling errors reminds us that in a real-life writing situation, it is expected to have technology to aid our writing. It was observed that monitoring is an ongoing process that may be conscious or subconscious. More successful test-takers exhibited this awareness.

## Appraisal

Appraisal refers to evaluating the qualities of written responses, including content, organisation, grammatical and lexical accuracy, and range. Appraisal is related to monitoring, but it is apparent when test-takers check their responses before submission. Less successful test-takers engaged less in self-appraisals (e.g., Cui and Yan). Cui and Yan did not have enough time to evaluate their writing before submitting.

Because the time is very limited, I didn't have time to check, so I see [now] because there many spelling mistakes in it. (Cui)

Oh, no. I mean, in fact, I don't have time to check. I think because before you were writing, I, I wish I had enough time to write. I think it's enough to write originally. And when I write it, I think it's difficult. And then I spent too much time to write it and then then find I don't have enough time. So, I think I'm just a little bit disappointed. (Yan)

Liang said he had not done much double-checking due to the time limit. He, however, thought he checked his writing as he wrote. His appraisal tended to take place during self-monitoring.

---

Um, I would like to. So, I don't remember when I was writing this passage if I had some time to check. Yeah, so the IELTS test, I don't think I have much time. And I just like why I was writing the passage. I know I checked when I have written before. Yeah. So maybe if I had some time I would start to double-check. (Liang)

Zhao did not have enough time to complete the task in the allotted time. This also meant that he was not able to double-check his writing before submitting it.

Wait, I didn't finish. In the end, it's my problem. I didn't have time to check writing. (Zhao)

Ren reported that she had not had time to double-check her work. Nonetheless, she pointed out that in a real test, she would try to evaluate her essay before submission.

Yeah. It happened to me. Yeah. I think I could make it if I was given more time. I could spend time for example, polish language and also I tried to make the content, I mean the evidence part to provide more content to support my point, but because of the time limit. So, I just finished one draft and I didn't have enough time to check it and just submitted. (Ren)

Wang reported self-assessment and evaluation during his writing, which was focused on content, cohesion and coherence, rather than on grammatical structures. However, he might have implicitly checked his grammatical accuracy as he wrote, rather than checking it after he had completed his essay.

But not to check my grammar, you can see that a second sentence has a grammar mistake. I usually come back to read the sentences and to make sure that I am still on topic and also to make sure that two sentences have like kind of coherence and cohesion. (Wang)

Lizzy had about three minutes left at the end, so she reviewed her essay and found some typos, which she corrected. In her self-evaluation, she focused on the readability of her essay; she also made sure she had not used any particular words too frequently.

Mainly just like readability. Like if there were any kinds of grammatical things that didn't match typos and also yet like other words I could use, I know just to check the words around or if you have, I noticed also you saw like repeated words, something you're trying to figure out. Yeah, yeah, I like to make it like a bit more varied. (Lizzy)

Sam reported that since he used the first-person pronoun in his essay, he felt his essay might read like a speech.

I think I found it hard not to use personal pronouns, so I think I use quite a bit of like, the global community should act or like we should act, and I think it kind of felt a bit more like I was. By persuading people to act instead of like writing about the issues of global warming, but I don't know if that's a problem for this kind of task, but. It almost felt like I was giving a speech. (Sam)

It was observed that integrated writing requires both ongoing self-assessment during writing and self-evaluation of the whole essays once completed. Time limit appears to be a key factor affecting whether the writing performance is evaluated.

In summary, both quantitative and qualitative results of writing, reading and test-taking strategy use suggest that test-takers engaged in these processes when addressing both writing tasks. A strong statistical relationship existed between reported writing and test-taking processes in both writing tasks. That is, there was convergent evidence in processes involved in the two different tasks. In the content analysis, a range of strategic processes were found to be used in varying degrees by different test-takers. Strategic processes in an independent task appear to be a foundation of those required in an

integrated writing task, which requires additional processes such as paraphrasing, discourse synthesis and acknowledgement of source texts.

## 6.6 Research question 6: What are test-takers' perceptions toward the usefulness and relevance of an integrated writing task to academic study?

This research question examines test-takers' perceptions of the usefulness of an integrated writing task if it was to be introduced and included in an official IELTS writing task, particularly regarding their academic preparation to study at university (addressing consequential validity). It also seeks to identify the nature of its disadvantages or issues that the task might introduce. This question is answered first by analysing the questionnaire responses to questions 16 to 19, followed by content analysis of the interviews.

### 6.6.1 Quantitative findings

#### Descriptive statistics

Table 30 presents the descriptive statistics of items 16 to 17. The mean scores ranged from 3.340 (Item 19) to 3.930 (Item 16). Table 31 shows the frequency counts of responses to the given descriptors from strongly disagree (1) to strongly agree (5). It was found that 70% (N = 109), 52.6% (N = 81), and 69.5% (N = 107) agreed and strongly agreed with items 16, 17, and 18, respectively. Item 19 had a mixed result in that 45% agreed and strongly agreed with this item (N = 69), 19.5% disagreed and strongly disagreed with this item (N = 30). 36% of the responses (N = 55) were neutral or undecided.

**Table 30:** Descriptive statistics of items 16 to 19 (N = 154)

Item	Minimum	Maximum	Mean	SD	Skewness	Kurtosis
16.	1	5	3.930	0.866	-0.615	0.454
17.	1	5	3.560	0.961	-0.329	-0.089
18.	1	5	3.860	0.854	-0.506	0.081
19.	1	5	3.340	1.062	-0.251	-0.416

**Table 31:** Frequency counts of items 16 to 19 (N = 154)

No.	Item	1	2	3	4	5
16.	I believe a reading-to-write test task is more authentic to assess academic writing than an opinion essay writing task.	2 (1.3%)	4 (2.6%)	39 (25.3%)	<b>67</b> <b>(43.5%)</b>	42 (27.3%)
17.	I believe my ability to write well academically is more appropriately assessed by a reading-to-write test task than by an opinion essay writing task.	4 (2.6%)	13 (8.4%)	<b>56</b> <b>(36.4%)</b>	55 (35.7%)	26 (16.9%)
18.	I believe that a reading-to-write test task would better prepare me to cope with academic writing in a real-life academic situations than an opinion essay writing task.	1 (0.6%)	8 (5.2%)	38 (24.7%)	<b>72</b> <b>(46.8%)</b>	35 (22.7%)
19.	I prefer an opinion essay writing task to a reading-to-write test task.	8 (5.2%)	22 (14.3%)	<b>55</b> <b>(35.7%)</b>	47 (30.5%)	22 (14.3%)

## Correlations

Table 32 presents Pearson correlations between items 16 to 19. Pearson correlations indicate a significant positive relationship between items 16, 17 and 18, implying that test-takers who agreed on one item tended to agree on the others and vice versa. Item 19 was not significantly correlated with items 16 and 18 (negative) but with item 19 ( $r = -0.184$ ,  $p < .05$ ). This finding implies that test-takers believed that a reading-to-write task was an authentic writing task that would better prepare them to study at university. On the contrary, they did not have a strong position on an opinion essay task.

**Table 32:** Pearson correlations between items 16 to 19

		17	18	19
16.	Pearson Correlation	<b>0.571**</b> R <sup>2</sup> = 0.33	<b>0.617**</b> R <sup>2</sup> = 0.38	-0.084
	Sig. (2-tailed)	<.001	<.001	.298
17.	Pearson Correlation	1	<b>0.556**</b> R <sup>2</sup> = 0.31	-0.184* R <sup>2</sup> = 0.03
	Sig. (2-tailed)		<.001	.023
18.	Pearson Correlation		1	-0.120
	Sig. (2-tailed)			0.137

## Differences by success groups

It is useful to know whether test-takers with different success levels differ in their responses to these items. This analysis was achieved through MANOVA. Levene's test of equality of error variances shows non-statistical significance for all four items, indicating that the data could be statistically comparable. Table 33 presents the descriptive statistics of the three groups. Groups 1 and 2 had similar mean scores across items. Group 3 reported slightly higher mean scores than the other two groups. Table 34 presents the test of the between-subjects effect. It was found that there were two significant differences in items 17 ( $p < .05$ ) and 18 ( $p < .01$ ). A post-hoc analysis indicates that only Groups 2 and 3 significantly differed in their mean scores in the two items ( $p < .05$  and  $p < .01$ , respectively), with a small effect size. This analysis implies that all three groups rated similarly in items 16 to 19.

**Table 33:** Descriptive statistics of items 16 to 19 by the three success groups (N = 154)

Item	Success Group	Mean	Std. Deviation	N
16	1.00	3.790	0.833	43
	2.00	3.840	0.862	57
	3.00	4.130	0.870	54
17	1.00	3.420	0.982	43
	2.00	3.370	0.858	57
	3.00	3.870	0.972	54

18	1.00	3.810	0.824	43
	2.00	3.610	0.818	57
	3.00	4.150	0.833	54
19	1.00	2.560	0.983	43
	2.00	2.560	1.035	57
	3.00	2.830	1.129	54

**Table 34:** The test of between-subjects effects (N = 154)

Source	Dependent Variable	F	Sig.	Partial Eta Squared
Success	16	2.335	.100	.030
	17	4.670	.011	.058
	18	5.890	.003	.072
	19	1.176	.311	.015

### 6.6.2 Qualitative findings

Content analysis suggests that all test-taker interviewees consider an integrated writing task relevant and useful to university study.

According to Cui, an integrated writing task would encourage international students to develop critical thinking, summarising and paraphrasing skills that would also benefit their future studies.

And it's, I think if there's this kind of test in the IELTS test, some students will pay more attention on critical thinking and summarising, paraphrasing, which will make their future life in the college more easier. But they might have difficulty in meets the criteria of the entrance score. (Cui)

Despite his perceived lack of success in the integrated writing task, Yan was optimistic that this task type would be useful for his future study.

I think, is, uh, in fact, writing task number two is more difficult, but I think is more useful for my future study ... It's, ah, it's worth to learn and to study how to write and how to read and reference. (Yan)

Yan also reflected that having to prepare to take an integrated writing task would help him prepare for future study.

And because I think, your teachers will ask you to read papers and articles about the project and then you write in it, if you would, just to know writing but not know how to reference, how to take notes. And I think it's not useful, I think, just to state your point and not to have scientific and the academic evidence to support your point. (Yan)

Although Liang found the integrated writing task difficult, he perceived that preparing for an integrated writing task would benefit students as it would help them prepare for academic study. He mentioned that he did not know how to use information in his assignments when he started his degree. It was challenging for him, and he had to learn how to do this during his study.

---

Well, I think, through taking this task, they will have a concept of what academic writing could be when you studied abroad. And also, you know, take Chinese students, for example, Chinese students are really good at taking exams in China. So, in terms of the IELTS training, I think if the IELTS writing tasks could be practised into writing, students make great efforts to prepare for these tests and then have a better or better understanding of what academic writing would be. So, I think it can generate some positive, positive effects. (Liang)

Zhao positively perceived the usefulness and relevance of the integrated writing task.

I think read-to-write tasks are more like the kind of university writing style. The professor will give you a bunch of the reading material; you need to read them and write an essay about how to respond to questions. And so, I think it's pretty good preparation for university life, actually. (Zhao)

He also articulated further that many international students did not have enough experience with this kind of writing, and he thought that limited their study capacity. Zhao indicated that the ways he was taught essay writing in high school and university (undergraduate) were similar. He also mentioned that an integrated writing task was novel for him, and he thought that such a task would “encourage students to improve their reading accuracy” as they needed to use information from the given texts to support their essay writing.

Lu was convinced about the benefits of introducing an integrated writing task for international students. She reflected that, even for this project, she realised she needed background knowledge to help her complete an integrated writing task successfully. She also reported learning new things from this task.

Oh, yes, I definitely think it is helpful. The reading parts offer me the background information about global warming so I can realise, oh, so these people have different attitudes towards global warming ... So, it's to open my mind, I think, the reading. (Lu)

She reported that this writing task would encourage students to realise the importance of reading and improving their reading skills, including their reading speed.

Oh, yes, maybe not only for me, but also for some other students, they think, oh, really, really spends a lot of time to finish the integrated task, but on the contrary, so you can kind of help you to realise, oh, my reading speed is not as satisfied as I expected myself before. So, you can realise. It can help me at least to realise for this point I must improve. (Lu)

She believed that reading had the potential to allow students to accumulate broad background knowledge, and that an integrated writing task would be more interesting than an independent writing task.

And number two is reading, you know, not only offers us the background information and I think this kind of writing task is also interesting because, well, a lot of independent writing, they are only a short sentence, a very concise problem for this kind of writing. I think it is a little bit more interesting as well. (Lu)

When asked about other benefits for students, Lu confirmed that it would influence how they prepared themselves for study, especially when they realise what they need to improve further after taking the test.

I think if the students realise the importance of writing or they will think about it. They take the writing as an important thing rather than just sit through it.

---

They can think, oh, there is something I can improve like that. So, I think this kind of writing task is really helpful to help them to realise that they have there is space for them to improve. (Lu)

Teng believed that an integrated writing task in IELTS would have benefited him. In retrospect, he felt he could have been better prepared for his postgraduate study had he prepared for this type of task before his study.

That's really important. Or maybe you make proper citations or references, and then use all that evidence to prove your ideas or claims. I think that's similar to what you do when writing the research paper. So, I think that would definitely help. I was not trained to write something like this when back in undergrad. So, I feel like this kind of task should be given to people who have never experienced something like this before. So, I think that's really beneficial for those people. (Teng)

Furthermore, as with Lu, Teng reflected that completing this integrated writing task had been a learning experience for him.

I'm thinking that like maybe next time when I encounter something similar to those, I will be able to do better because I'm more familiar with the past, like familiar with what they asked me to do and required me to do so. So, I was thinking about what I read on news, what I heard, those kind of stuff. (Teng)

Teng also mentioned that he could improve his writing by reading.

I can actually learn from them, like how they arrange all those different ideas, how they also like references, quotations, and how they use evidence to prove their ideas or claims, I think I can definitely learn from them. (Teng)

Ren believed it was very important for international students to know how to complete an integrated writing task, as they would be better prepared when starting their university studies.

Before I enter like the university, I did not have any kind of this kind of practice. So especially when I got to do my first assignment, I felt torn because I had no experience in doing it. For young students, if they do not have this kind of practice before, it will be really difficult for them to get used to the academic writing style. Yeah. So, it's not actually about ability, it's about if they know this before. (Ren)

During the interview, Wang shared accounts of some of his experiences during his master's degree study. He frequently needed to engage in reading-to-write activities.

At the university we, of course, do the reading first and then write. So, I think that's essential. Without reading, I couldn't write very well, especially on a topic that I am not familiar with. (Wang)

He pointed out that an integrated writing task was relevant to academic study as it was an authentic activity. He also said it would help students prepare for their study before commencing it.

I think definitely it's better than just an independent essay because, well, at university we write essay. I mean, we it's more authentic, you know, like at university we do the reading first and then write. It's more authentic. So, if IELTS introduced reading to write then it's more close to the real life. I think that will prepare students better for the university. (Wang)

Wang also pointed out the potential benefit of improving reading ability if an integrated writing task for international students was introduced.

---

Well, for international students, I think reading to write also indicates that reading is very important and indicates that the writing problem is not only the writing problem: it may be the reading problem. (Wang)

Wang pointed out that an integrated writing task would assess additional writing constructs to those evaluated in an independent writing task.

And a reading-to-write task could tell them that writing is not only about vocabulary and grammar. It's also about like background knowledge ... Maybe more might write more logically and clearly. So, I'm not sure if there is a correlation there, but I think it may help if students have more background knowledge. (Wang)

He also mentioned the usefulness of an integrated writing task for teaching.

And so, reading to write would be, I think, very useful, at least in teaching. I think it would be very useful. In testing, I think it's more like, it's more authentic to the writing that we do at university. So, I think that's one way to tell international students what the real life is at the university. And it can better prepare them for their university studies. (Wang)

Lizzy pointed out that the second writing task was more common and relevant to her undergraduate study experiences than the first one.

Whereas the second task, it's exactly what you're going to be doing at uni, like kind of what's the word like, getting to like synthesising it, you know, comparing arguments, contrasting different opinions, different ideas. That's exactly what you do when you do assignments. (Lizzy)

She thought that synthesis writing skills were crucial for various subjects in university study. She suggested that students should be better prepared in this area to be more likely to succeed in their studies.

I think all the subjects like history and politics subjects, all that kind of stuff, it's really, really important to be able to have that skill, to be able to take different ideas and compare them. (Lizzy)

Lizzy pointed out that an integrated writing task resembled what students would have to do in their university programs.

So it's testing them on the things that they are going to be doing later on. And I think being able to read it, like read and then use those arguments and pull them out; I think that's a really key skill that'll benefit them. (Lizzy)

She also thought that an opinion essay task was less beneficial for students in terms of their preparedness for university study.

So I guess if they know that that's what is in the IELTS test, in their revision for it, rather than writing things for like, like opinion pieces, like if they're reviewing lots and lots of stuff about like you know, like writing essays on their own opinion, um, I think that wouldn't be as beneficial as writing these types of essays, because these are the skills you have to be doing when you get to uni stuff like, um, looking at sources and quoting people and how to reference things. (Lizzy)

Sam believed an integrated writing task was relevant to university study for various reasons.

Because that's the nature of writing at a university. You have to look at source materials and then comment on support your own essay with information from source material. So, I think it's a good like task actually.

---

I think if they're doing the IELTS tests in preparation to go into university, it's helpful because it is testing a similar skill to what they'll need to do at university.

So, I think just quickly and in relation to writing, I think obviously I learnt a lot about writing by doing and I think that's allowed me to see a lot of examples of writing.

And it'd probably resemble real-world academic writing when people think their tutors or lecturers will probably be marking the assignment of something for something. Yeah, but I think students don't have as clear an idea about what the tutors are looking for when they mark their assignments. (Sam)

### Perceived negative impacts of using an integrated writing task in IELTS

Despite overwhelming support for an integrated writing task, some cautions about its unintended impacts due to its inclusion in the IELTS test have been noted by test-takers. Such impact includes the possibility of making the test much more challenging for lower proficient students to get admitted into a university program. For example, Cui felt that introducing a reading-to-write task could make it more difficult for international students to gain the required bands for admission into their preferred courses: "But they might have difficulty in meeting the criteria of the entrance score". In addition, Teng mentioned students with a low proficiency level may face challenges in completing an integrated writing task well.

This really is a rather long task to do so because I think I was given like 50 minutes or something like that. It's actually a long one. So, yeah, I guess students really have to be prepared for that. Yeah, maybe especially for students that don't have sufficient English skills, especially since I know some people like they read slowly. (Teng)

Liang cautioned that despite the potential benefits of an integrated writing task, test users should recognise that writing scores do not necessarily reflect trait language proficiency. That is, an integrated writing task may still not help international students overcome issues of academic writing problems.

But sometimes, we also need to acknowledge that although you can get a very high score on the test, it doesn't mean that you have developed language proficiency. Maybe they are just really good at taking the tests. (Liang)

According to Liang, international students need to be aware that although they may be able to achieve the required integrated IELTS writing score for admission to the university courses of their choice, such test achievement does not guarantee their academic success, especially when a substantive piece of writing is required for course fulfilment.

Ren felt that opportunities for international students to prepare for integrated writing tasks are essential for test preparation.

But the other thing that I think, ah, is how, you know, I mean, how feasible it is. Because a lot of students do not even know how to cite. And it might take a little bit, a little while for them to know what synthesis works and how good synthesis of certain opinions works in some ways and how to balance their opinion and others' opinions. I think this is quite difficult for students if they do not have any prior knowledge. But I think one way to address this problem is to provide students with sample articles. And if they have enough time to, for example, discuss sample articles, a good one or, you know, not that good ones. And, you know, coupled with practice, I think it's quite meaningful. And I think it's worth, you know, it is worth a shot. (Ren)

---

In summary, quantitative and qualitative analysis of test-takers' data suggest that they believed an integrated writing task was useful and relevant to their future academic study. This task would have a long-term impact on students' writing skills that could be applied in a real-world educational setting.

## 6.7 Research question 7: What are raters' perceptions toward an integrated writing task for assessing academic writing?

This research question investigates how the raters perceive the usefulness of the rating meetings before the actual rating and what they consider essential to ensure the validity and usefulness of this writing task (addressing consequential and scoring validity). Three themes have been identified in relation to the content analysis of raters' interviews: (1) perceived usefulness and advice on rater training; (2) key considerations for stakeholders in the use of an integrated writing task; and (3) feasibility of an integrated writing task in the official IELTS test.

### 6.7.1 Perceptions of the rater training and meeting sessions

Both raters reported that the rater sessions were useful and helped them understand the rating processes for the integrated writing task. The sessions allowed them to understand how the other rater rated the essays and to share their viewpoints and reasons for giving the scores they gave. The raters also had an opportunity to discuss expectations about rating instructions, etc.

I thought the sessions were very helpful and they pretty much confirmed the agreements between me and [Rater 2]. Anyway, they confirmed how high standard our ratings are. You know, like most of our ratings were either the same or just half a band out. And I think it just is a testament to the consistency of the way we rate. And I think it's really good. (Rater 1)

That process is always, always useful I think. Yeah, like no matter how confident you are. Still, yes, still hearing someone else's thinking process is always going to be helpful. (Rater 2).

Regarding rating the integrated essays, Rater 1 pointed out that it helped him iron out any issues he had in rating.

Like when we were talking about that one, that one candidate who just copied a whole bunch of text and just threw it in there and that brought up a little bit of an issue. Then we were able to discuss it and at least find some ground about how we're supposed to write something like that. (Rater 1)

Rater 2 pointed out that the rater meetings in this project were different from those in official IELTS training, as both raters were treated as the experts to determine the rating standards of integrated responses.

I totally get why you couldn't have done it differently. Yeah, yeah, yeah. Yeah. That, that's a bit more kind of like onus on us. OK, let's work out, let's work out what's good objective criteria for this thing. (Rater 2)

Rater 2's remark is significant for future rater training if an integrated writing task is to be used in an official IELTS test.

When I've had regular IELTS training, I, we have a lot of initial training and then we have what's called standardisation and re-certification every two years. I think the difference is there, that the scripts that we're given have been officially rated, like a

---

set of people in Cambridge, I think. Yeah, that's an agreed-upon standard. And so, the examiner who is trained in Sydney, knows exactly, OK, this script is five, six, five, six, like there's no debating that. That's the standard. And that's what everybody needs to adjust to, to internalise that standard. Yeah. So you're kind of like adjusting recalibrating yourself to fit the official standard. (Rater 2)

Accordingly, Rater 2 felt that even though she understood the purpose of the project's rater meetings, she found there was a lack of pre-rated essays by an expert panel of [Cambridge] examiners, which they could use to calibrate their scores. As such, Rater 2 felt that both Rater 1 and herself were "the only reference points we have for each other" or, as she mentioned above "a bit more kind of like onus on us". She expressed the challenge of having no other standards of reference as follows.

So, if he (Rater 1) says, I think this is a three, four, four, five, and then maybe we go, OK, let's agree for a four. But that's kind of just been defined by what I like. It's not actually based on anything more official like maybe an official team would say no; actually, it is a three. Yeah, that's a little bit tricky. (Rater 2)

Rater 2's comments are significant in that rater training in integrated writing evaluation needs to be carried out strictly according to official IELTS training procedures. This would also mean a range of integrated essays (from lowest to highest) would need to be rated by official experts so that raters can model their ratings, and a range of training and moderation activities should take place to achieve high-quality rating standards.

Overall, the three meetings with the raters were essential in the current study as they helped the raters adjust their pre-existing mindsets in rating essays. This was required in this project since the rating scales and the concepts used differed from those in the official IELTS Writing assessment.

### 6.7.2 Key considerations for stakeholders in the use of an integrated writing task

If an integrated writing task is to be implemented, key issues should be considered further. Both raters discussed them in relation to test-takers, test developers, and raters. Their ideas about these groups are interconnected.

#### *For test-takers*

The raters suggested that test-takers need to be clear on what an integrated writing task is about in a given IELTS context and how they should approach it. For example, how should they integrate the provided source texts with their own knowledge? This issue is discussed earlier about the nature of responses by weak and strong test-takers. Rater 1 also suggested that the test instruction should be explicit and clear in terms of the requirements for test-takers to use the provided source texts in their essays and the extent to which they can use the source texts (e.g., what percentage of their essays should refer to the source texts). Subsequently, this would also help raters deal with source text use in their responses. Rater 2 pointed out the importance of assisting test-takers become familiar with this kind of task, by providing a resource for test-takers to help them know what to do, so that they are better prepared to address an integrated writing task.

Yeah, I was going to say my overwhelming sense was just that because it's an innovation, the candidates lacked the training in this question. You know, I think if I were part of this experiment, like if I'd been given six months of training to deal with this kind of question, we would have received a totally different set of data. (Rater 2)

---

### For test developers

According to the two raters' reflections and shared experiences in this project, test developers need to consider the following. First, the test instructions should be very clear to the test-takers, especially regarding source text use by test-takers; for instance, how to deal with overuse and underuse of the source texts. This needs to be clarified with the assessors.

There need to be clearer instructions to the examiners on what to expect with using the source material and the impact of overusing source material on the other criteria. So how does that impact? So, if you overuse the source material, how does this impact answering the question? How does this impact grammar, especially if you're copying somebody else's words? Well, that's not your grammar. That's not your lexical resource. That is somebody else's. So how does that impact? (Rater 1)

Based on the two raters, the integrated writing instructions should explicitly ask test-takers to:

- read and comprehend the source texts as accurately as possible
- use the source texts in their essays and acknowledge them (e.g., quotations and paraphrasing)
- know the maximum use of source texts in an essay (e.g., no more than X% of the source texts or total word count).  
It needs to be clear that if they use more than 10% of the source material or 20%, they will be penalised because they need to. It should say use your knowledge and the source material to support your belief or your position. (Rater 1)
- include their own knowledge and examples in an essay.

Second, it is important to select the topic, prompt and source texts carefully. Rater 2 pointed out that source texts are important in determining test-takers' responses and raters' assessments.

Perhaps then for the source texts. Yeah. Maybe when they were being chosen, you'd be looking for things that were somehow, yeah, maybe a little less controversial than something like global warming. (Rater 2)

In the current project, the source texts caused some uncertainty and difficulty in acknowledging or referencing the source because a source text mentions people external to the text (secondary source). For example, in relation to Source Text 1, should an acknowledgement read "According to Ruth Curry" or "According to David Smith (2005), Ruth Curry ..."? It, therefore, needs to be clear to test-takers and raters what an appropriate or acceptable way to cite the source being used should be. Furthermore, it needs to be clear whether it is equally acceptable to say, for example, "According to passage 2, ...", "According to Sterling Burnett (2001), and "According to Burnett (2001), ..." Therefore, it is essential that test developers of an integrated writing task consider using source texts that allow clear citations and acknowledgements.

Perhaps it would be clearer for candidates if you had, uh, if the source text was just like one author from one place. It was just their opinion because it became quite complicated when you've got like Ruth Curry or Sharon Smith, who's from the University of Miami and quoted in the main author, David Smith, you know. Then that's been adapted from something else, like it becomes quite convoluted to know, like who my article referred to, you know. Whereas if you maybe just had something all from David Smith that would kind of simplify things for the candidate. Just to be able to say. And Smith also thinks this space, rather than having to deal with all these different names, because that makes it quite tricky. (Rater 2)

---

In relation to prompts, Rater 1 pointed out an *issue of limited access to resources and unfairness* for a particular group of test-takers.

But naturally, if you have a question about. Um... Let me think, a question about the environment, naturally, students living in India, in a village. They may not be very educated and they don't have access to that kind of knowledge or that vocabulary that someone living in like Switzerland might, so I imagine that the source texts would kind of be a little bit maybe democratising. Do you know what I mean? You know, like, I thought it would even things out, but actually, I don't think it did. Like, I think sometimes students probably did have enough ideas on vocabulary in their heads, but instead of drawing on it, they just copied out chunks, which ended up disadvantaging them. (Rater 1)

Third, it is important to develop scoring rubrics for differentiating responses at different proficiency levels. In the current project, analytic scoring rubrics were used and perhaps were more complex than holistic scoring rubrics. Rater 1 discussed this issue and acknowledged that it was complex and required time to work out through various response samples.

I mean, it needs to, it needs to be clear in the rating scales, as well. So for example, like [band] 2, like it could be something like overuse. It produces very little original work or ideas or something like that I would say ... I guess with the [band] 3 here, it's quite tricky. That's the thing, it is tricky. But I mean you'd have to sort of, take a look at ... you'd have to take a look at several examples and see how they impact the response. You'd have to see what a band 1 is and see an example of that and then like, oh, okay, so for example, a band 1 for using the source is not using the source at all. Okay, so that's why I write it as band 2. (Rater 1)

Rater 1 also pointed out that when analytic scoring is used for integrated writing assessment, the use of a band in the source text could impact other scoring criteria, so this needs to be carefully considered in the scoring rubric development.

A band 2 could be, okay, they've attempted to use it, but that's all they've done; they haven't done anything else, so that could be that. But that would affect everything around it, because if all they have used is that source material, they haven't answered the question. They haven't used their own language. They haven't organised the information the way it should be, and they haven't used their own grammar. So that should impact. So, you know. And then band 3 could be, okay, there is a little bit of their own language, just bits and pieces, but the majority of it is the text and so on. And you move up the scale that way. Do you know what I mean? (Rater 1)

### *For raters*

The two raters generally found the evaluation of the integrated writing task essays more challenging and demanding than the evaluation of the independent writing task essays. Compared to other assessment criteria, assessing source text integration was most challenging. Rater 2 explained this as follows:

When you're reading the candidate's response, there might be phrases that you're like, oh, like that's a nice phrase. That's good vocabulary. But like, actually, it's just been directly copied or equally, you know, something might not be that impressive. But actually, it's a really good attempt at paraphrasing a really complex idea in the source text. So, you need to be super familiar with the content of the source text to be able to recognise that and to recognise exactly how this source text has been used. (Rater 2)

---

Both raters pointed out that they did not judge test-takers based on their positions (e.g., whether they thought climate change was real). This confirms that both raters were aware of bias in assessment.

Because I think, you know, like you can go either way on it. I judge them on how good their argument is. And yeah. So that was what I did. (Rater 1)

Rater 2 also cautioned that an issue of unintended/subconscious bias from raters may be present due to a topic effect.

I didn't find this personally, but I can imagine perhaps an inexperienced examiner might get it. They might get swayed by the opinions rather than the language, because obviously IELTS is strictly assessing language. You're not in someone's opinion, but I can imagine, if some, if a candidate with the right you know, "I'm a climate change denier and like, climate change is false and here's the evidence from source two". I wonder if some examiners might influence them, even if subconsciously. (Rater 2)

Rater 2 was asked if technology such as Turnitin could be used to capture text test-takers copied (text similarity) and the percentage that they copied. While this might be helpful, she thought it could distract the normal and natural reading of an essay. It could also somehow influence her rating.

Like, it's like all of it comes up highlighted? Naturally, that would have an impact. It's quite tricky. I think it is tricky because then you're somehow using a kind of, you know, an artificial intelligence almost. It's like, you know, normally it's just the examiner and the text. And so with Turnitin, yeah. Even if it's subjective, it's still this, like, intermediary, you know, that's like another ... somehow like somehow getting in the way. I don't know. (Rater 2)

When assessing writing, a borderline issue occurs when an essay may be either one or other point in the rating scales. Raters were asked if they experienced this, and how they dealt with it. Rater 1 reported that he did share this when considering the language issues and development of ideas as discussed above.

At a borderline, is the argument good enough? Is there enough from the candidate to prove their point or is it missing something? (Rater 1)

However, Rater 2 found she experienced dealing with borderline issues more at the level of source text use. Her questions may be useful when considering the nature of rater training for assessing integrated writing task essays.

I wasn't totally clear for myself, like: What an ideal response looks like? What percentage of the response should be from the source material? And did that need to be spread through the response or did that not matter? How good did the referencing need to be like? Did it matter? For example, they might have used the source material really, really skilfully, but just left out the acknowledgement. And then you're like, OK, yeah, like, do I need to penalise them for that or like, is that unfair? (Rater 2)

Both raters believed that rater training is critical for successfully implementing an integrated writing task into IELTS Writing. A new writing task always requires the orientation of the raters.

It does, like, it makes you kind of orient your head differently from the regular IELTS. Yeah, it's just, like, it's clear that one hundred per cent of what's on the page has come from the candidate's head. So, I'll need to like to try to ascertain if something's been memorised. Whereas here, if you're kind of balancing, OK, what's come from

---

the client's head, what what's original and what might have been memorised. And then also what comes from the source text and what comes in a natural, skilful way and what's perhaps then copied verbatim. But you're balancing all of those different things, which is. Yeah, it's different. Like, it's more complicated. I think so. (Rater 2)

Furthermore, to help raters understand what is involved in an integrated writing task, the raters agreed that one of the best strategies is to ask raters to take a test sample.

I think the first thing that the rater should do, and you did it with me, is that I think everybody should attempt to do it by themselves because once it gets right, it certainly gives you an idea of what you know, what the test-takers are going through. So, yeah, I think that's the first thing. The second thing is you really need to be familiar with the source material and you do tend to get familiar with it as you mark. (Rater 1)

Rater 1 pointed out that another challenge raters will face is the need for raters to remember information from the source texts. In the current study, this was demanding at the beginning of the rating, but it got easier as the raters progressed through the marking.

Initially, yes, but after you read so many, you know what the source is saying and you don't have to go back. But, yeah, initially, that's why at the beginning it was a bit slower because I'm like, OK, what are they saying? (Rater 1)

Both raters underlined the fact that it would have been more challenging for raters if managing essay marking had not been thought through systematically and strategically.

Like, I mean, if it was an IELTS-style set up, there would be hundreds of questions around the world. I think that would be very challenging to mark, because you have to familiarise yourself with one and then all of a sudden, a new task comes up and then you have to familiarise yourself with that source material. So that would take more time to mark an IELTS. (Rater 1)

I guess. The tricky thing, like a real obstacle as the examiner would be at the moment, if I mark, if I'm marking like 50 IELTS task 2 papers, I might have like 20 different questions. But obviously, each question is only like a sentence long. It only takes me a second to familiarise myself with the topic, the task. Whereas if you're dealing with the source text, that obviously requires a whole other process for the examiner in terms of familiarisation, so it would be fine if I had 50 papers for one source text. But imagine if I have 50 papers and 20 different source texts like that, the work would just be too much ... That would be quite a challenge for examiners if they had to look at many different sources. (Rater 2)

The interviews suggest that introducing an integrated writing task in the official IELTS test would make IELTS Writing more challenging for raters as they would have to refer to the source texts and keep abreast of many integrated prompts during their essay evaluation. Therefore, it would be desirable for raters to assess essays based on the same integrated prompt. Without a clear rating management protocol, rating quality could be an issue.

Otherwise, you know, you're going to be compromising something, otherwise you're compromising speed, either speed or accuracy. You know, you might be fast, but then you're going to start missing things. (Rater 2)

The most I marked at one time was 35. That was the absolute most. I mean that's very little compared to IELTS because in IELTS I can do about 100, but in IELTS as well, I don't have to concentrate as much on this. Like there's no source.

---

Plus, in a standardised test, I'm also not making notes. I'm not writing down those notes so. This is also that.

### 6.7.3 Raters' perceptions toward the feasibility of an integrated writing task for an official IELTS

The raters were asked about the feasibility of including an integrated writing task in an official IELTS test. The raters had different perspectives about its inclusion. Rater 1 thought that introducing an integrated writing task would create a problem for raters, although he did understand the rationale behind it.

Well, I don't know, I think it would be viable. I mean, the idea is very clever. But at the same time. I see it causing a lot of problems in the sense that the IELTS system works in a particular way. And to change that system in the sense that you're adding something more for examiners to mark, and because examiners only have a certain amount of time to mark their work. So, to add, that would just make it even longer to add something that I have to, that they have to, read. (Rater 1)

Rater 1's major concern was practicality due to the additional cognitive load for raters, especially given that there could be several writing task prompts and source texts to handle in a single session.

Yeah, but you're adding the source material, so they have to read that source material, familiarise themselves with it, and then judge whether that material is used appropriately or not. So just doing that would, I mean, take an extra, say, two to three minutes, multiplied by one hundred tasks that they have to do. And then you have to add into the fact that. Tasks change as well. So, around the world, different test-takers are doing different tasks. (Rater 1)

Rater 2 believed that if both test-takers and raters were well prepared for this task, it would be viable and useful for international students.

That would definitely be helpful. Yeah, yeah. It's obviously a set of skills that these particular candidates haven't had the training for. But you know, if they had the training, then that training would definitely be helpful for them at a tertiary level, yeah. If I had training and, you know, like IELTS has the resources, you know, like, if there were other resources for them to know what to do. Yeah, then of course, yeah. (Rater 2)

In summary, the answers to this research question provided insights into the raters' perceptions towards the scoring of integrated essays, the challenges they faced, and viewpoints on including this essay task in the official IELTS Writing test.

---

## 7 Implications and conclusions

The current IELTS research project seeks to understand the nature of integrated writing performance as compared with independent writing performance. The independent writing task was similar to the current IELTS Writing Task 2. The integrated writing task asked test-takers to read two texts and write a response to address the essay question. The remaining sections are organised as follows. First, the key findings and implications are summarised and discussed in relation to the literature. Four types of validity (cognitive, context, scoring and criterion-related validity; Weir 2005) are considered in light of the findings. Key strengths and limitations of the current study are discussed, and recommendations for further research are next articulated.

### 7.1 Summary of the key findings

#### 7.1.1 Relationship between independent and integrated writing performance

It was found that the total independent score was moderately related to the total integrated score ( $r = 0.505$ ,  $R^2 = 0.26$ ,  $p < 0.01$ ). Qualitative analysis suggests that the quality of essay development, task achievement and cohesion and coherence might explain the shared variance. Both quantitative and qualitative findings indicate that in order for ESL students to be successful in an integrated writing task, they need to understand how to address an independent writing task, including how to structure an essay and support their argument and position to a given issue. It can be said that success in an integrated writing task can also be accounted for by other abilities that are not assessed in an independent writing task (e.g., reading comprehension skills, discourse synthesis skills). Researchers in integrated writing postulated that discourse synthesis in an integrated writing task requires higher-order processing between reading and writing, which has distinctive features from writing processes in an independent writing task (e.g., Grabe & Zhang, 2013; Plakans, 2015).

#### 7.1.2 Assessing integrated writing responses

The two raters reported they, for example, assessed whether test-takers answered the question, whether their essays were clear and precise, how they organised their essay structures, and how well they supported their arguments. Regarding the use of source texts in the integrated writing task, both raters reported they focused on text length and used ideas from the source texts. Both raters also said that they found several essays that either overused or underused the source texts. In particular, they observed that many test-takers used direct quotes from the source texts in their essays, but only a few effectively paraphrased the source texts for integration with their ideas. They also found that good integrated essays discussed both sides of the argument well. Gebril and Plakans (2014) analysed raters' evaluation processes when they assessed integrated essays. They found that the raters were locating the source texts, checking the content accuracy of the source texts being used and monitoring citation mechanics in essays. The raters engaged in similar assessment processes, as reported in Gebril and Plakans (2014).

#### 7.1.3 Differences between independent and integrated writing performance

It was found that test-takers performed significantly better in the independent writing task than in the integrated writing task (see e.g., Table 17). The statistical differences were found in a total score comparison and in each assessment criterion. The quantitative and qualitative analyses suggest that the integrated writing task was more difficult than the independent task. Brown et al. (2005) argued that integrated tasks are more demanding than a traditional stand-alone independent writing task.

---

Qualitative analysis suggests that performance differences could be explained by, for example, task familiarity (i.e., they were more familiar with an independent writing task than an integrated writing one). Task familiarity influences how test-takers write to meet the assessment criteria. Qualitative findings suggest other inhibiting factors affecting their writing performance, such as a lack of background knowledge and paraphrasing and citation skills. These factors were reported by participants across the integrated writing success levels, but more so among lower success groups. Some facilitative factors which more successful test-takers often reported include the ability to brainstorm or plan ideas for an essay, select and extract important information from the source texts for essay writing, and monitor and evaluate their writing. An opportunity to read their own text seems to be a mediator of their integrated writing processes (e.g., Gebriel & Plakans, 2014; Plakans, 2009b). Another explanation for poorer performance in the integrated writing task was related to test-takers' source text use, which was low across the participant groups (see also Cumming et al. 2005). Since source text use is at the heart of an integrated writing task, good or poor use of source texts in a given essay affects other assessment criteria (e.g., task fulfilment, cohesion and coherence). In an integrated writing task, not only do test-takers need to read the source texts to select and organise their essays (Plakans, 2015), but they also need to read their own writing to ensure good paraphrasing and text integration.

The quantitative and qualitative findings allow us to understand the criterion-related validity of the two writing tasks in that the constructs of the two tasks were quite different. Without appropriate and sufficient training or preparation on how to address an integrated writing task, it is likely that test-takers will not do well on this type of writing task.

#### 7.1.4 IELTS scores and integrated writing performance

Evidence of criterion-related validity was also addressed by examining the statistical relationships among reported IELTS proficiency scores and independent and integrated writing performances (Table 20). It was found that IELTS proficient scores (overall, reading and writing) were significantly correlated to both the independent and integrated writing performances. The correlation coefficient between overall IELTS proficiency scores was more strongly related to the independent writing performance than to the integrated writing performance ( $r = 0.57$  versus  $0.35$ ,  $p < 0.001$ , respectively). This was not surprising since an official IELTS test includes a writing task similar to the independent writing task used in the current study. A multiple regression analysis suggests that independent writing performance was the only significant predictor of integrated writing performance. A moderate regression coefficient might indicate that differences in writing constructs are being assessed in the integrated writing task, for example, discourse synthesis ability, which is not evaluated focally in the independent writing task. A path analysis (see Figure 3) also suggests that IELTS proficiency scores are indirectly related to integrated writing performance via independent writing performance.

#### 7.1.5 Integrated writing responses by success levels

In terms of the integrated writing success comparisons (see Tables 23 and 24), it was found that the three success groups significantly differed in their integrated writing performance. Qualitative analysis of the rater comments indicates some differences between the three groups, although there were some common features. Key raters' comments on their integrated writing were primarily related to issues of source text use (e.g., lack of source text use, overuse of the source texts and lack of acknowledgement or paraphrasing). The findings were consistent with Cumming et al. (2005) who found that low-ability writers had difficulty using source texts in their writing.

---

The analysis suggests that participants in Level 3 success were awarded higher scores in the source text use criterion than those in Levels 1 and 2. That is, although most of the test-takers performed better on the independent writing task than on the integrated writing task, different success groups illustrated differences in integrated writing quality. The findings were consistent with those of Gebril and Plakans (2013), who found significant differences in fluency, accuracy, source text use and complexity among different proficiency levels in an integrated writing task. The findings provide evidence of task and context validity of the integrated writing performance, which somewhat differed from the independent writing task.

#### 7.1.6 Cognitive processes in integrated writing

The study also provides evidence of cognitive validity in terms of similarities and differences in test-takers' reported use of writing and test-taking strategies in the two writing tasks. It also addresses the question of the perceived relevance and usefulness of an integrated writing task by the research participants. Test-takers reported they engaged in the cognitive activities that described both writing task engagement. The correlation between the writing and test-taking strategies in both the independent and integrated writing tasks was found to be strong (see Table 29). Reported test-taking strategies in both the independent and integrated writing tasks were positively related ( $r = 0.69, p < 0.001$ ). Reported reading and writing strategies in the integrated writing task were also significantly related ( $r = 0.75, p < 0.001$ ). Qualitative analysis suggests that key strategic processes explaining integrated writing performance include assessing the situations, understanding task requirements, outlining or essay planning, monitoring, and performance appraisal. There were various ways in which the participants approached both writing tasks. First, participants in a lower integrated success group (e.g., Cui and Yan in Level 1 group) might have some difficulty understanding the writing task requirements. However, they reported they read the task instructions and considered what to include in their essays. The quantitative and qualitative findings were in line with a study by Plakans (2009b), for example, who found that more successful test-takers engaged in better discourse synthesis processes than less successful test-takers, who tended to experience some reading comprehension difficulty and basic sentence construction.

#### 7.1.7 Test-takers' perceptions of an integrated writing task

In regard to the use of an integrated writing task in an official IELTS test, both quantitative and qualitative analyses suggest that test-takers perceived it to be valuable and relevant to their university study. For example, it allows them to improve their reading, critical thinking and paraphrasing skills, which are part of the writing task. The task is similar to several academic writing tasks they experienced while studying in Australia. Such academic tasks require them to synthesise multiple texts across multiple modes as part of their writing processes (Grabe & Zhang, 2013; Plakans, 2022).

#### 7.1.8 Raters' perceptions of an integrated writing task

The raters reported that the integrated writing task was more demanding and challenging to score as they needed to be aware of particular source texts and how they were used. That is, source texts formed another layer of their assessment processes. Similar to Gebril and Plakans (2014), both raters reported they experienced having to deal with some borderline issues with the level of source text use. The raters also recommended some further improvement in rater training, for example, by allowing raters to calibrate their assigned scores with those of experts. They also provided insightful recommendations for test-takers, test developers, and raters. The two raters agreed that an integrated writing task was an authentic academic writing task and would be useful for international students. However, Rater 1 pointed out that

---

acknowledgements and citations of source texts should go beyond a specific referencing style. Rater 2 was more optimistic about an integrated writing task than Rater 1 in that if test-takers had some training and preparations for an integrated writing task, their essays would differ from what they assessed in the current project. Rater 2 also pointed out that an integrated writing task was viable and practical for international students. Although Rater 1 agreed that an integrated writing task was useful for international students, he was more concerned with the feasibility of an integrated writing task in term of scoring, which was more demanding than an independent essay task. His concern was sensible in that, in reality, there would be more integrated essay questions for raters to score, and it took some time to become familiar with the source texts before they could start scoring. Accordingly, if an integrated writing task were to be implemented, more plans would be needed to maximise the quality and feasibility of an integrated writing task.

## **7.2 Strengths and limitations of the current study**

There are some strengths of the current study that are worthwhile mentioning. First, the data collection design in which a series of online test tasks and cognitive questionnaires were sequentially presented to the research participant on one data collection period allowed us to gain a closer understanding of the relationship between independent and integrated writing performance and related cognitive processes. Had the two writing tasks been given on two separate days or occasions, the same test-takers could have been lost. This would pose difficulty in examining the relationships between both writing task performances. Nonetheless, this data collection procedure also had some limitations. For example, test-takers spent about two hours completing writing tasks and questionnaires. They could experience fatigue that affects their subsequent writing, thinking and motivation. Accordingly, their essays might not best represent their best performance.

A second positive feature of the current study was that it employed an analytic scoring rubric to assess both writing tasks. Analytic scoring allows us to understand the specific quality of various aspects of writing. In the current study, task achievement and source text use were explicitly assessed in the integrated writing task, allowing us to obtain different scores for different aspects of writing. The total score based on two raters' scores also provided richer information about a given writing performance. However, using analytic scoring remains challenging in high-stakes, large-scale standardised testing situations, as this method is time-consuming and labour-intensive.

In addition to using the analytic scoring rubric, the current study employed two accredited IELTS raters to evaluate each test-taker's writing responses. Their scores were averaged to represent each test-taker's final score. Together with rater meetings and training sessions, this adoption of the scoring rubrics and two accredited IELTS raters helped ensure the quality of the writing scores used to infer writing performance on both writing tasks. Nevertheless, this is also a limitation in assessment practice since employing two raters to assess the same piece of writing could be impractical.

A significant limitation of the current study was the use of truncated data to address the research questions. As discussed in the Methodology section, most of the test-takers were at an intermediate level (i.e., homogeneous dataset) and were mainly Chinese students. There were fewer lower and higher proficiency levels in the current study. In statistical analysis, truncated data remained limited for realistically estimating a relationship between variables due to little variances. In future research, a larger sample size should be pursued to enhance the statistical validity of the findings. Finally, since the data did not include many ESL international students from other regions, the findings cannot be generalised to other populations and, hence, remain inconclusive.

---

A combination of the qualitative analysis, for example, through test-takers' interview data, provided an understanding of what test-takers did as they attempted to complete the two writing tasks. The qualitative analysis gives voices to individual test-takers to share their thoughts, stories, and experiences as they complete both writing tasks. Their experiences should be considered to inform test development. However, some limitations of qualitative analysis should be noted. For example, qualitative research is time-consuming as a qualitative analysis requires the researcher to read the transcripts several times to understand the data and extract aspects or themes from the participants' discussion. Second, more participants with lower success levels were needed, as only two were included. As discussed in the Methodology section, the key reason for not having many participants with a lower success level was that the interviews were conducted soon after their test-taking. Still, their writing responses were not yet rated. It might also be that people with higher language proficiency were more willing to participate in this type of study than those with a lower proficiency level. The current study might provide more insight had some interviews been conducted in the participant's first language. Some participants sometimes could not articulate their thoughts clearly or comfortably. Third, while the researcher explicitly acknowledged the subjectivity of the data analysis from the outset and the intent not to generalise the findings, qualitative analysis remains limited in that the ESL test-takers were from China, and only two Australian students were included. In further research, more participants from other contexts would provide other unique stories, writing processes, factors affecting their writing, and perceptions toward the relevance and usefulness of an integrated writing task.

The findings from rater data also helped address issues of the scoring validity of an integrated writing task. It also provided further empirical explanations for quantitative and qualitative findings, such as how the raters derived the scores statistically analysed with other variables and written comments on essays. Second, their viewpoints on differences among good, average, and poor integrated essay features were significant in helping fine-tune future scoring rubrics. Features of integrated essays are also useful for writing syllabuses to promote integrated writing skills among L2 English students.

Some limitations of rater data analysis included the time demanded to analyse the features of integrated essays of various success levels. Future studies need more test-takers with higher English proficiency levels as well because it will allow us to obtain more features of good integrated essays and confirm the identified features. Similar to the analysis of test-takers' interview data, analysing the interview transcripts was also time-consuming. The level of careful reading by the researcher for extracting significant concepts and exploring key issues related to the raters' standards and perceptions toward an integrated writing task is essential for the trustworthiness of the findings and interpretation.

It remains challenging to develop and use integrated writing tasks for assessing academic writing (Plakans, 2022; Gebriil & Plakans, 2014; Weigle & Malone, 2016). For instance, it is challenging to define the constructs of integrated writing that fit all academic contexts because integrated writing can vary in convention from language to language and context to context (see e.g., Hirvela, 2016; Hyland, 2016; Hyland & Shaw, 2016; Manchón, 2016). It is also challenging to identify common core topics and tasks for integrated writing (Plakans, 2022) and to develop high-quality prompts for integrated tasks (Weigle & Malone, 2016).

---

### 7.3 Insights into the nature of integrated writing in the current study

Analysis of quantitative and qualitative data to answer the research questions has led to understanding the writing requirements and processes for addressing an integrated writing task. First, it needs to meet a general academic writing expectation, similar to other kinds of writing. For instance, quality content and appropriate and adequate content knowledge (e.g., accuracy and relevance of evidence, claims, and inferences) are essential. Knowledge of academic language style (e.g., being formal or less casual, use of hedges, illustration of cohesion, acknowledgement of source texts) is associated with the accuracy of language use (e.g., the use of complete sentences, appropriate word choices, correct spelling, and mechanics) and the appropriacy of rhetorical expressions. Nonetheless, success goes beyond accurate grammatical accuracy, appropriate vocabulary, and essay length. In evaluating integrated writing, content in an academic text (e.g., source information, ideas) is expected to be logical, credible and relevant to an assessment task. Writers need to understand how to organise their writing and write coherently and how to express their own ideas (critical, creative, problem-solving), stance or voice in their writing, and synthesise and paraphrase source texts provided or used. The notion of voice in academic writing generally refers to the writer's position, proposition, viewpoint, attitude, or beliefs (Paltridge et al., 2009).

Second, integrated writing follows the writing processes of independent essay writing. Field (2004) describes such processes as macro-planning, organising, micro-planning, translating, monitoring, and revising.

- Macro-planning involves the gathering of ideas and the identification of contexts such as genre and readership.
- Organising includes sorting ideas, relating them to each other, and prioritising various writing aims.
- Micro-planning is related to dynamic planning for a specific part of the writing, such as a paragraph, and may include identification of the intended aim of that part.
- Translating occurs when the writer encodes and converts abstract ideas or concepts into meaningful linguistic forms while developing and finalising drafts.
- Monitoring occurs during and after texts have been constructed. It involves checking language accuracy, such as grammatical structures, spelling, and mechanic use (e.g., punctuation and capitalisation). Monitoring also includes evaluating the quality of texts.
- Revising is often a result of monitoring processes as writers adjust, correct, and improve specific sentences, paragraphs, or the whole text.

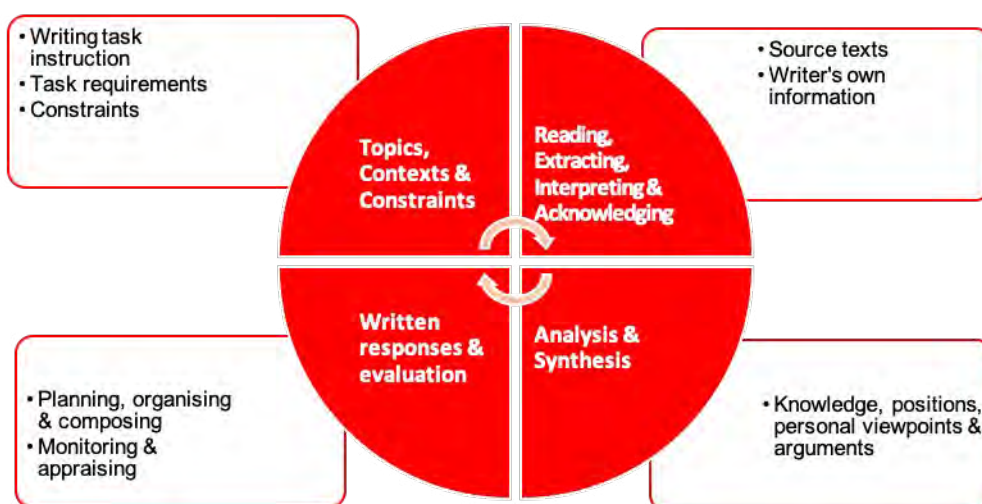
Third, integrated writing requires the writer to understand specific writing task requirements under a particular condition (e.g., time allowance, expected word counts, and lengths of source texts). That is, what a task asks them to do and address specifically. Unlike an independent writing task, integrated writing requires the writer to select and organise information from the provided source texts to combine it with their own information, such as viewpoints and examples (Jakobs, 2003). The writer needs to understand that it is not the source texts nor their own information that makes up integrated writing.

Fourth, successful integrated writing requires effective reading comprehension skills for the source texts and one's written responses (see Flower et al. 1990). Reading comprehension skills include:

1. receiving and interpreting written information (Urquhart & Weir, 1998)
2. remembering, memorising, inferencing, analysing, and utilising written information (Alderson, 2000)
3. extracting and integrating information from the text with what is already known. (Koda, 2005)
4. summarising, organising and synthesising information from multiple reading sources (Grabe & Stoller, 2020).
5. evaluating the writing products through reading (Hayes, 1996; Hayes & Flower, 1980; Hirvela, 2004, 2016; Hyland, 2016)

Figure 7 summarises the key insights based on the current study. The study has suggested that successful integrated writing requires more than writers' linguistic resources. Integrated writing ability requires writers to produce written responses directly addressing a given writing task cohesively and coherently (e.g., answering a question, positioning viewpoints, supporting them with evidence or data from source texts, and producing sufficient texts as specified by the writing task). Good integrated writing illustrates accurate and appropriate information extracted from the source texts (i.e., the ability to coordinate external resources with internal knowledge). It also represents how well writers strategically craft their texts through synthesis, paraphrasing, and acknowledgement of used source texts (e.g., via self-monitoring and appraisal).

**Figure 7:** Key features and requirements for successful integrated writing.



---

## 7.4 Implications for IELTS and future research

The current study has broadened an understanding of issues in assessing integrated writing through quantitative and qualitative methods. Subsequently, it has opened new avenues of research to address methodological and technical issues related to an integrated writing task. In particular, the study's outcomes have provided the IELTS organisation with empirical evidence of test-takers' performance variation between independent and integrated writing tasks and differences across proficiency levels. Studies that replicate the current project that include L2 English test-takers from different first language backgrounds are needed to shed more light on the nature of integrated writing performance and important assessment issues and technical considerations that help improve the validity and feasibility of this type of assessment.

The current project has provided several implications for L2 writing assessment, teaching and learning. Insights from the current study suggest that integrated writing is an authentic and multimodal writing activity that is affected not only by the knowledge of academic writing but also by the use of source texts that support an argument or position. It is different from the assessment of writing through an independent task. The current study has also suggested a potential positive washback of an integrated task for international students. For example, in the qualitative interviews with test-takers, participants reported learning about academic writing from completing the integrated writing task. They realised some strengths and weaknesses in their writing, and they learned about a topic about climate change that is useful for them.

If IELTS consists of an integrated writing task, international students can develop an integrated writing skill before their study commencement (i.e., promoting positive washback). Inclusion of this writing task can further enhance the reputation and standing of the IELTS Writing. It is beneficial and plausible for the IELTS organisation to consider including an integrated writing task in an official IELTS test. Of course, the IELTS developers can commit themselves to including an integrated writing task in IELTS Writing once they have considered the *viability and validity* of integrated writing assessment. They need to further address issues raised by the current findings (e.g., test preparation, task design, rater training, and scoring methods). Since technological affordances have influenced the way people write and the way we assess their writing, it is worth exploring how technology can be utilised to maximise both assessment and test-taker capacity. Writing assessment that is aligned with real-world academic writing will improve test-takers' transfer skills to their academic study (Plakans & Gabriel, 2013) and enhance test usefulness – a stronger validity argument (Bachman & Palmer, 2010; Chappelle, 2021; Kunnan, 2018).

## 7.5 Concluding remarks

The benefits of integrated writing assessment are far-reaching for all stakeholders, especially ESL international students, because it simulates authentic language use in a real-world academic setting. The need for an alignment between an academic writing test task and the real-world educational writing activity is a valid argument for promoting an integrated writing assessment task used for high-stakes language tests (Hirvela, 2016; Plakans & Gebiril, 2013). The current project has provided empirical evidence that an integrated writing task has a significant potential to provide a more complete picture of students' current proficiency levels that will, in turn, help them be more prepared for their university studies. The question is not whether an integrated writing task can be implemented in an official IELTS test, but rather, the question is when and how it can be best implemented.

---

## References

- Alderson, J.C. 2000. *Assessing reading*, Cambridge University Press, Cambridge.
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCEM). 2014. *Standards for educational and psychological testing*, American Educational Research Association, Washington, DC.
- Bachman, L, & Palmer, A. 1996. *Language testing in practice*, Oxford University Press, Oxford.
- Bachman, L, & Palmer, A. 2010. *Language assessment in practice*, Oxford University Press, Oxford.
- Banerjee, J, Franceschina, F, & Smith, AM. 2007. Documenting features of written language production typical at different IELTS band score levels, *IELTS Research Reports*, vol. 7, pp. 1–69. IELTS Australia and British Council.
- Barkaoui, K. 2016. 'What changes and what doesn't? An examination of changes in the linguistic characteristics of IELTS repeaters' Writing Task 2 scripts, *IELTS Research Reports Online Series*, vol. 3, pp. 1–55. IELTS Partners: British Council, Cambridge Assessment English, and IDP: IELTS Australia.
- Barkhuizen, G. 2015. Narrative inquiry in *Research methods in applied linguistics: a practical resource*, eds B Paltridge & A Phakiti, London, Bloomsbury, pp. 169–185.
- Brown, A. 2012. Interlocutor and rater training in *Routledge handbook of language testing*, eds G Fulcher & F Davidson, Routledge, London & New York, pp. 413–425.
- Brown, A, Iwashita, N, & McNamara, T. 2005. *An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks* (TOEFL Monographs Series # MS29), Educational Testing Services, Princeton, NJ.
- Brunfaut T, Harding L, & Batty AO. 2018. Going online: The effect of mode of delivery on performances and perceptions on an English L2 writing test suite, *Assessing Writing*, vol. 36, no. 1, pp. 3–18.
- Chapelle, CA. 2021. *Argument-based validation in testing and assessment*, Sage, London.
- Chapelle, CA, & Lee, H. 2022. Conceptions of validity in *Routledge handbook of language testing*, 2<sup>nd</sup> edn, eds G Fulcher & L Harding, Routledge, London and New York, pp. 17–31.
- Chapelle, CA, & Voss, E (eds). 2021. *Validity arguments in language testing: Case studies of validation research*, Cambridge University Press, Cambridge.
- Cheong, CM, Zhu, X, Li, GY & Wen, H. 2019. Effects of intertextual processing on L2 integrated writing, *Journal of Second Language Writing*, vol. 44, no. 1, pp. 63–75.
- Cho, Y, & Choi, I. 2018. Writing from sources: Does audience matter?, *Assessing Writing*, vol. 37, no. 1, pp. 25–38.
- Cotton, F, & Wilson, K. 2011. An investigation of examiner rating of coherence and cohesion in the IELTS Academic Writing Task 2, *IELTS Research Reports*, vol. 12, pp. 1–76. IDP: IELTS Australia and British Council.
- Cresswell, JW, & Plano Clark, VL. 2018. *Designing and conducting mixed methods research*, 3<sup>rd</sup> edn, Sage, London.

---

Cumming, A, Kantor, R, Baba, K, Erdosy, U, Eouanzoui, K, & James, M. 2005. Differences in written discourse in independent and integrated prototype tasks for the next generation TOEFL, *Assessing Writing*, vol. 10, no. 1, pp. 5–43.

Davis, L. 2022. Rater and interlocutor training in *Routledge handbook of language testing*, 2<sup>nd</sup> edn, eds G Fulcher & L Harding, Routledge, New York and London, pp. 322–338.

Department of Education, Australian Government. 2023. International student monthly summary and data tables. <https://www.education.gov.au/international-education-data-and-research/international-student-monthly-summary-and-data-tables>

Field, J. 2004. *Psycholinguistics: The key concepts*, Routledge, London.

Flower, L, Stein, V, Ackerman, J, Kantz, M, McCormick, K, & Peck, W. 1990. *Reading-to-write: Exploring a cognitive and social process*, Oxford University Press, Oxford.

Gebriel, A, & Plakans, L. 2013. Towards a transparent construct of reading-to-write assessment tasks: The interface between discourse features and proficiency, *Language Assessment Quarterly*, vol. 10, no. 1, pp. 1–19.

Gebriel, A, & Plakans, L. 2014. Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks, *Assessing Writing*, vol. 21, no. 1, pp. 56–73.

Golparvar, SE, & Khafi, A. 2021. The role of L2 writing self-efficacy in integrated writing strategy use and performance, *Assessing Writing*, vol. 47, 100504.

Grabe, W, & Stoller, FL. 2020. *Teaching and researching reading*, 3<sup>rd</sup> edn, Routledge, New York and London.

Grabe, W, & Zhang, C. 2013. Reading and writing together: A critical component of English for academic purposes teaching and learning, *TESOL Journal*, vol. 4, no. 1, pp. 9–24.

Grabe, W, & Zhang, C. 2016. Reading-writing relationships in first and second language academic literacy development, *Language Teaching*, vol. 49, no. 3, pp. 339–355.

Hayes, JR, & Flower, LS. 1980. Identifying the organization of writing processes in *Cognitive processes in writing*, eds LW Gregg & ER Steinberg, Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 31–50.

Hayes, JR. 1996. A new framework for understanding cognition and affect in writing in *The science of writing: Theories, methods, individual differences and applications*, eds CM Levy & S Ransdell, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 1–27.

Hirvela, A. 2004. *Connecting reading & writing in second language writing instruction*, University of Michigan Press, Ann Arbor, MI.

Hirvela, A. 2016. Academic reading into writing in *Routledge handbook of English for academic purposes*, eds K Hyland & P Shaw, Routledge, New York and London (pp. 127–138).

Hyland, K. 2016. *Teaching and researching writing*, 3<sup>rd</sup> edn, Routledge, London.

Ingram, D, & Bayliss, A. 2007. IELTS as a predictor of academic language performance, Part 1, *IELTS Research Reports*, vol. 7, pp. 1–68. IELTS Australia and British Council.

Ivankova, NV, & Greer JL. 2015. Mixed methods research and analysis in *Research methods in applied linguistics: A practical resource*, eds B Paltridge & A Phakiti, Bloomsbury, London, pp. 63–81.

- 
- Jakobs, E-M. 2003. Reproductive writing — writing from sources, *Journal of Pragmatics*, vol. 35, no. 4, pp. 893–906.
- Kane, MT. 2022. Articulating a validity argument in *Routledge handbook of language testing*, 2<sup>nd</sup> edn, eds G Fulcher & L Harding, Routledge, London and New York, pp. 32–47.
- Knoch U. 2022. Assessing writing in *Routledge handbook of language testing*, 2<sup>nd</sup> edn, eds G Fulcher & L Harding, Routledge, New York and London, pp. 236–253.
- Knoch, U, & Sitajalabhorn, W. 2013. A closer look at integrated writing tasks: Towards a more focused definition for assessment purposes, *Assessing Writing*, vol. 18, no. 4, pp. 300–308.
- Koda, K. 2005. *Insights into second language reading: A cross-linguistic approach*, Cambridge University Press, Cambridge.
- Kunnan, A. 2018. *Evaluating language assessments*, Routledge, New York.
- Lea, M, & Street, B. 1998. Student writing and faculty feedback in higher education: An academic literacies approach, *Studies in Higher Education*, vol. 23, no. 2, pp. 157–172.
- Lewkowicz, JA. 1994. Writing from sources: Does source material help or hinder students' performance? in *Language and learning: Papers presented at the annual international language in education conference*, eds N Bird et al., ERIC Document (ED 386 050).
- Lillis, T, & Tuck, J. 2016. Academic literacies: A critical lens on writing and reading in the academy in *Routledge handbook of English for academic purposes*, eds K Hyland & P Shaw, Routledge, New York and London (pp. 30–43).
- Mackey, A, & Bryfonski, L. 2018. Mixed methodology in *Palgrave handbook of applied linguistics research methodology*, eds A Phakiti, P De Costa, L Plonsky & S Starfield, Palgrave, London, pp. 103–121.
- Mickan, P, & Slater, S. 2003. Text analysis and the assessment of academic writing, *IELTS Research Reports*, vol. 4, pp. 59–88. IELTS Australia and British Council.
- Mickan, P, 2003. 'What's your score?' An investigation into language descriptors for rating written performance, *IELTS Research Reports*, vol. 5, pp. 128–155. IELTS Australia and British Council.
- Miles, MB, Huberman, AM, & Saldaña, J. 2014. *Qualitative data analysis: An expanded sourcebook*, 2<sup>nd</sup> edn, Sage, Los Angeles.
- Moore, T, & Morton, J. 1999. Authenticity in the IELTS academic module writing test: A comparative study of Task 2 items and university assignments. *IELTS Research Report*, vol. 2, pp. 64–106. IELTS Australia Pty Limited, Canberra.
- Mullins, G, Quintrell, N, & Hancock, L. 1995. The experiences of international and local students: three Australian universities, *Higher Education Research and Development*, vol. 14, no. 2, pp. 202–231.
- O'Loughlin, K, & Wigglesworth, G. 2003. Task design in IELTS academic writing task 1: The effect of quality and manner of presentation of information on candidate writing, *IELTS Research Reports*, vol. 4, pp. 89–130. IELTS Australia Pty Limited, Canberra.

---

Ohta, R, Plakans, LM, & Gebril, A. 2018. Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis, *Assessing Writing*, vol. 38, no. 1, pp. 21–36.

Paltridge, B, & Starfield, S. 2019. *Thesis and dissertation writing in a second language: A handbook for supervisors*, 2<sup>nd</sup> edn, Routledge, London.

Paltridge, B, Harbon, L, Hirsh, D, Shen, H, Stevenson, M, Phakiti, A, & Woodrow, L. 2009. *Teaching academic writing: An introduction for teachers of second language writers*, University of Michigan Press, Ann Arbor.

Paul, A. 2007. IELTS as a predictor of academic language performance, Part 2. *IELTS Research Reports*, vol. 7, pp. 1–35. IDP: IELTS Australia and British Council.

Payant, C, McDonough, K, Uludag, P, & Lindberg, R. 2019. Predicting integrated writing task performance: Source comprehension, prewriting planning, and individual differences, *Journal of English for Academic Purposes*, vol. 40, no. 87–97.

Pecorari, D. 2008. *Academic writing and plagiarism: A linguistic analysis*, Continuum, London.

Phakiti, A. 2007. *Strategic competence in EFL reading test performance*, Peter Lang, Frankfurt am Main.

Phakiti, A. 2016. Test-takers' performance appraisals, appraisal calibration, state-trait strategy use, and state-trait IELTS listening difficulty in a simulated IELTS Listening test, *IELTS Research Reports Online Series*, no. 6, pp. 1–140. IELTS Partners: British Council, Cambridge Assessment English, and IDP: IELTS Australia.

Phakiti, A. 2018. Confirmatory factor analysis and structural equation modelling in *Palgrave handbook of applied linguistics research methodology*, eds A Phakiti, P De Costa, L Plonsky & S Starfield, Palgrave, London, pp. 459–500.

Phakiti, A, & Li, L. 2011. General academic difficulties and reading and writing difficulties among Asian ESL postgraduate students in TESOL at an Australian university, *RELC Journal*, vol. 42, no. 3, pp. 227–264.

Plakans, L. 2008. Comparing composing processes in writing-only and reading-to-write tasks, *Assessing Writing*, vol. 13, no. 1, pp. 111–129.

Plakans, L. 2009a. Discourse synthesis in integrated second language writing assessment, *Language Testing*, vol. 26, no. 4, pp. 561–587.

Plakans, L. 2009b. The role of reading strategies in integrated L2 writing tasks, *Journal of English for Academic Purposes*, vol. 8, no. 2, pp. 252–266.

Plakans, L. 2010. Independent vs, integrated writing tasks: A comparison of task representation, *TESOL Quarterly*, vol. 44, no. 1, pp. 185–194.

Plakans, L. 2012. Writing integrated assessment in *Handbook of language testing*, eds G Fulcher & F Davidson, Routledge, Oxford, UK, pp. 249–261.

Plakans, L. 2015. Integrated second language writing assessment: Why? What? How?, *Language and Linguistics Compass*, vol. 9, no. 4, 159–167.

Plakans, L. 2022. Writing integrated tasks in *Routledge handbook of language testing*, 2<sup>nd</sup> edn, eds G Fulcher & L Harding, Routledge, New York and London, pp. 357–371.

---

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*, vol. 22, no. 2, pp. 217–230.

Plakans, L, Gebril A, & Bilki Z. 2019. Shaping a score: Complexity, accuracy, and fluency in integrated writing performances, *Language Testing*, vol. 36, no. 2, pp. 161–179.

Révész, A, Michel, M, & Lee, M. 2017. Investigating IELTS academic writing task 2: Relationships between cognitive writing processes, text quality, and working memory, *IELTS Research Reports Online Series*, vol. 3, pp. 1–44. IELTS Partners: British Council, Cambridge Assessment English, and IDP: IELTS Australia.

Riazi, AM, & Knox, JS. 2013. An investigation of the relations between test-takers' first language and the discourse of written performance on the IELTS Academic Writing Test, Task 2, *IELTS Research Reports Online Series*, vol. 2, pp. 1–89. IELTS Partners: British Council, Cambridge Assessment English, and IDP: IELTS Australia.

Sawaki, Y, Quinlan, T, & Lee, Y-W. 2013. Understanding learner strengths and weaknesses: assessing performance on an integrated writing task, *Language Assessment Quarterly*, vol. 10, no. 1, pp. 73–95.

Schreier, M. 2012. *Qualitative content analysis in practice*. Los Angeles, Sage.

Schreier, M. 2014. Qualitative content analysis in *Sage handbook of qualitative data analysis*, ed U Flick, Sage, London, pp.170–183.

Schumacker, RE, & Lomax, RG. 2015. *A beginner's guide to structural equation modelling*, 4<sup>th</sup> edn, Routledge, New York and London.

Shaw, P, & Pecorari, D. 2013. Source use in academic writing: An introduction to the special issue, *Journal of English for Academic Purposes*, vol. 12, no. 2, pp. A1–A3.

Street, BV, & May, S (eds). 2017. *Literacies and language education*, 3<sup>rd</sup> edn, Springer: Cham, Switzerland.

Swain, M, Huang, L, Barkaoui, K, Brooks, L, & Lapkin, S. 2009. The speaking section of the TOEFL iBT (SSTiBT): Test-takers' reported strategic behaviours, *TOEFL iBT Research Report*, RR 09–30.

Uitdenbogerd, AL, Lynch, K, Harland, J, Thevathayan, C, Hamilton, H, D'Souza, D, & Zydervelt, S. 2018. IELTS: Student and supervisor perceptions of writing competencies for a Computer Science PhD, *IELTS Research Reports Online Series*, No. 1. IELTS Partners: British Council, Cambridge Assessment English, and IDP: IELTS Australia.

Urquhart, S, & Weir, C. 1998. *Reading in a second language: Process, product and practice*, Longman, New York.

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.

Weigle, S. 2002. *Assessing writing*, Cambridge University Press, Cambridge.

Weigle, SC, & Malone, ME. 2016. Assessment of English for academic purposes in *Routledge handbook of English for academic purposes*, eds K Hyland & P Shaw, Routledge, New York and London, pp. 608–620.

---

Weir, C, O'Sullivan, B, Yan, J, & Bax, S. 2007. Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS Writing component: effects and impact, *IELTS Research Reports*, vol. 7, pp. 1–37. IDP: IELTS Australia and British Council.

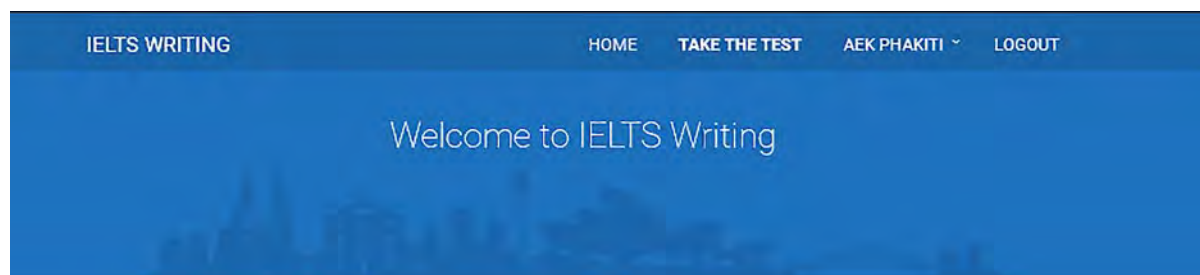
Weir, CJ. 2005. *Language testing and validation: An evidence-based approach*, Palgrave Macmillan, London.

Yu, G, He, L, & Isaacs, T. 2017. The cognitive processes of taking IELTS Academic Writing Task 1: An eye-tracking study, *IELTS Research Reports Online Series*, vol. 1, pp. 1–105. IELTS Partners: British Council, Cambridge Assessment English, and IDP: IELTS Australia.

Yu, G, Rea-Dickins, PM, & Kiely, R. 2011. The cognitive processes of taking IELTS Academic Writing Task 1, *IELTS Research Reports*, vol. 11, pp. 373–449. IDP: IELTS Australia and British Council

# Appendix 1: Screenshots of the online writing test

## Home screen



**Welcome to IELTS Writing.** This research project is led by Associate Professor Aek Phakiti ([aek.phakiti@sydney.edu.au](mailto:aek.phakiti@sydney.edu.au); Phone: +61293516312) from The University of Sydney. This study examines how people write in response to a given topic in English. There are two writing tasks in this test.

The first task is an independent writing task, which asks you to express your viewpoints about an issue using relevant examples from your own knowledge or experience. You have **40 minutes** to complete this task.

The second task is an integrated writing task, which asks you to express your viewpoints about a given topic using specific information that can be extracted from given two short texts, as well as relevant examples from your own knowledge or experience. You have **50 minutes** to complete this task.

For each task, there is a clock that indicates the remaining time to complete. You are required to submit the test after time has run out. After you have submitted your answer to each task, you will answer short questionnaire items about the test.

By participating in this, you will be eligible to win one of the three Apple iPads Air 2 (128GB) as well as receiving individualised feedback on your writing performance.

As part of the Ethics approval conditions, we are required to ask you to read the Participant Information Statement that is sent to you. By taking this test, you consent to taking part of this project.

Thank you for your participation.

[Take the test](#)

## General instructions for the IELTS independent writing task

### General Instructions for Independent Writing

In the writing test, you are presented with a point of view, argument or a problem. You are assessed on your ability to present a solution to the problem, present and justify an opinion, compare and contrast evidence and opinions, and evaluate and challenge ideas, evidence or arguments. You are also assessed on your ability to write in an appropriate style. You have **40 minutes** to complete this task, including drafting an essay outline and checking your writing before submitting. Write at least **250 words**.

Please note:

- **Do not close the browser** at anytime during the test
- You can have a paper to take notes and write an outline
- The test will commence as soon as you click on **START NOW**

START NOW

## General instructions for the IELTS integrated writing task

### General Instructions for Integrated Writing

In the integrated writing test, you are presented with a point of view, argument or a problem. In order to complete this task, you are to read the two short passages to obtain more information about the topic and use the readings to support your argument.

You are assessed on your ability to present a solution to the problem, present and justify an opinion, compare and contrast evidence and opinions, and evaluate and challenge ideas, evidence or arguments. You are also assessed on your ability to write in an appropriate style. You have **50 minutes** to complete this task, including reading the two texts, drafting an essay outline and checking your writing before submitting. You will be able to refer to the two reading texts at any time during this test. Write at least **250 words**.

Please note:

- **Do not close the browser** at anytime during the test
- You can take notes and write an outline
- The test will commence as soon as you click on **START NOW**

Start now

---

## Appendix 2: IELTS independent writing task

### General Instructions:

In the writing test, you are presented with a point of view, argument or a problem. You are assessed on your ability to present a solution to the problem, present and justify an opinion, compare and contrast evidence and opinions, and evaluate and challenge ideas, evidence or arguments. You are also assessed on your ability to write in an appropriate style. You have 40 minutes to complete this task, including drafting an essay outline and checking your writing before submitting. Write at least 250 words.

### Task:

#### Write about the following topic:

*"Some people say that the best way to improve public health is by increasing the number of sporting facilities. Others, however, say that this would have little effect on public health and that other measures are required."*

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

---

## Appendix 3: IELTS integrated writing task

### General Instructions:

In the integrated writing test, you are presented with a point of view, argument or a problem. In order to complete this task, you are to read the two short passages to obtain more information about the topic and use the readings to support your argument.

You are assessed on your ability to present a solution to the problem, present and justify an opinion, compare and contrast evidence and opinions, and evaluate and challenge ideas, evidence or arguments. You are also assessed on your ability to write in an appropriate style. You have 50 minutes to complete this task, including reading the two texts, drafting an essay outline and checking your writing before submitting. You will be able to refer to the two reading texts at any time during this test. Write at least 250 words.

### Task:

#### Write about the following topic:

*“Some people believe that global warming is damaging our planet. Others believe that global warming is not a serious problem.”*

*Which point of view do you agree with? Why?*

*Give reasons and support your writing with examples.*

Important

- The two passages should help you get some ideas about the topic. You should not spend more than 10 minutes reading the two texts and drafting your essay outline.
- You may go back to the passages to check information anytime.
- You can borrow ideas and examples from the text. However, you should mention the author's name or source text if you do so.
- Also, if you take exactly the same phrases or sentences mentioned in the passage, put them in quotation marks (“ ”).

**Reading (1):** *Scientists Say Global Warming is Undeniable* (Adapted from David Smith's (2005) “Scientists Say Global Warming Is Undeniable”)

Scientists have confirmed that climate change is being caused by human activity. A number of studies looking at the oceans and melting ice leave no doubt that it is getting warmer, people are to blame, and the weather is going to suffer. Tim Barnett, who is a famous global warming researcher, indicates that new computer models that look at ocean temperatures instead of the atmosphere show the clearest signal yet that global warming is well under way. Mr. Barnett said that earlier climate models based on air temperatures were weak because most of the evidence for global warming is not in the air. Other researchers found clear effects on climate and animals. For example, Ruth Curry, who is from an important oceanographic institute, said changes in the water cycle affect the ocean and, ultimately, climate. She said the changes were already causing droughts in the United States, and Greenland's ice cap. Sharon Smith of the University of Miami found melting ice was taking with it plants that are an important base of the food supply for many animals. And the disappearing ice meant the animals such as polar bears and seals were losing their homes. Given all these serious problems caused by global warming and the way humans have abused the earth, governments must act immediately to save our planet. The future of this planet depends on our actions and any delay would result in serious problems.

---

**Reading (2):** *Myths of Global Warming* (Adapted from Sterling Burnett's, (2001), "Myths of Global Warming")

There is no scientific agreement that global warming is a problem or that humans are its cause. Even if current predictions of global warming are correct, much of the environmental policy now proposed is based on wrong theories. First, there is a wrong belief that the earth is warming. While ground-level temperature suggests the earth has warmed between 0.3 and 0.6 degrees since 1850, reliable global satellite data show no evidence of warming during the past 18 years. In addition, scientists do not agree that humans affect global climate because the evidence supporting that theory is weak. Some people also think that the government must act now to stop global warming. However, a 1995 analysis by supporters of global warming theory concluded that the world's governments can wait up to 25 years to take action with no bad effect on the environment. In short, our policymakers need not act immediately. The government has time to gather more data, and industry has time to develop new ways of reducing its impact. Supporters of the theory of human-caused global warming also argue that it is causing and will continue to cause all environmental problems. Many famous scientists reject these beliefs. Sea levels are rising around the globe, though not equally. In fact, sea levels have risen more than 300 feet over the last 18,000 years. Contrary to the predictions of global warming theorists, the current rate of increase is slower than the average rate over the 18,000-year period.

## Appendix 4: The reflective questionnaire for the independent writing task

Gender: [ ] Male [ ] Female [ ] Others Age: \_\_\_\_\_

Nationality: \_\_\_\_\_

Your latest IELTS scores (if any, please complete below):

Overall IELTS score: \_\_\_\_\_ IELTS Reading score: \_\_\_\_\_

IELTS Writing score: \_\_\_\_\_

Directions: How true is each statement below for you? Rate the following descriptors to best describe the strategies you used.

1	2	3	4	5
Strongly disagree	Disagree	Neutral	Agree	Strongly disagree

### Part A: Writing Strategies

1.	I made an outline of what to include in my writing.	1	2	3	4	5
2.	I wrote using my own words.	1	2	3	4	5
3.	I organised my writing to meet the expected academic writing style.	1	2	3	4	5
4.	I made sure I supported my ideas with clear points, principles and examples.	1	2	3	4	5
5.	I checked to make sure I used appropriate words.	1	2	3	4	5
6.	I checked to make sure I didn't have any spelling mistakes.	1	2	3	4	5
7.	I checked to make sure I had produced sufficient words as required by the task.	1	2	3	4	5
8.	I used my knowledge and experience to support my points of view.	1	2	3	4	5
9.	I acknowledged sources of information in my writing.	1	2	3	4	5

### Part B: General Test-taking Strategies

10.	I made sure I understood the test instructions.	1	2	3	4	5
11.	I was aware of the time limitation of the test.	1	2	3	4	5
12.	I evaluated my performance while doing the test task.	1	2	3	4	5
13.	I considered whether my writing met the task requirements.	1	2	3	4	5
14.	I checked my writing before submitting.	1	2	3	4	5

## Appendix 5: The reflective questionnaire for the integrated writing task

Directions: How true is each statement below is for you? Use the following descriptors that best describe yourself.

1	2	3	4	5
Strongly disagree	Disagree	Neutral	Agree	Strongly disagree

### Part A: Reading Strategies

1.	I briefly skimmed the texts before I read more carefully.	1	2	3	4	5
2.	I considered what the author(s) meant in the texts.	1	2	3	4	5
3.	I used my prior knowledge or experience to help me understand texts.	1	2	3	4	5
4.	I took notes from the texts.	1	2	3	4	5
5.	I identified the main ideas in the given text.	1	2	3	4	5

### Part B: Integrated Writing Strategies

6.	I made an outline of what to include in my writing.	1	2	3	4	5
7.	I paraphrased the texts.	1	2	3	4	5
8.	I compared and synthesised multiple points of view from the written texts to help me write.	1	2	3	4	5
9.	I made sure I supported my ideas with clear points, principles and examples.	1	2	3	4	5
10.	I acknowledged source information from the given texts in my writing.	1	2	3	4	5

### Part C: General Test-taking Strategies

11.	I made sure I understood the test instructions.	1	2	3	4	5
12.	I knew the time limitation of the test.	1	2	3	4	5
13.	I evaluated my performance whilst doing the test task.	1	2	3	4	5
14.	I considered whether my writing met the task requirements.	1	2	3	4	5
15.	I checked my writing before submitting.	1	2	3	4	5

### Part D: Perceptions about this reading to write test

16.	I believe a reading-to-write test task is more authentic to assess academic writing than an opinion essay writing task.	1	2	3	4	5
17.	I believe my ability to write well academically is more appropriately assessed by a reading-to-write test task than by an opinion essay writing task.	1	2	3	4	5
18.	I believe that a reading-to-write test task would better prepare me to cope with academic writing in a real-life academic situations than an opinion essay writing task.	1	2	3	4	5
19.	I prefer an opinion essay writing task to a reading-to-write test task. (to be reverse coded)	1	2	3	4	5

---

## Appendix 6: Semi-structured interview questions for test-takers

### General information

1. Please describe your background.
2. Can you rate your academic writing skills? What are your strengths and weakness in your writing?
3. What do you think affects your academic writing?
4. What experience have you had with IELTS writing?

### Attitudes or beliefs about integrated writing

5. Do you think a reading-to-write task (integrated) would help you better prepare with your academic study at university?

### Independent writing scoring

6. What did you do when responding to this independent writing task?
7. What challenges did you face in responding to this independent writing task?
8. Were you satisfied with your performance on this task after you submitted?
9. What do you think you should further improve?

### Integrated writing scoring

10. What did you do when responding to this integrated writing task?
11. What challenges did you face in responding to this independent writing task?
12. Were there any of the features of the writing that you think impacted your performance in this task?
13. Were the given texts difficult for you to comprehend?
14. Did you consider the source texts as you responded to this integrated writing task?
15. How did you consider the source texts as you wrote in this integrated writing task?
16. Did you realise you copied directly any given texts without acknowledgement?
17. How did you control your writing in this aspect?
18. Were you satisfied with your performance on this task after you submitted?
19. What do you think you should further improve?

### Perceptions about the usefulness of the integrated writing task for international students

20. How would international students benefit from a reading-to-write task?

---

## Appendix 7: Semi-structured interview questions for raters

### General information

1. Tell me about yourself. What is your academic background? What do you do?  
Have you had any experience with ESL, EFL or EAP teaching?
2. What previous experience do you have with rating writing or academic writing?
3. How did you find the rater training/meeting sessions?

### Independent task:

4. What do you think are key characteristics or features of very good essays in the independent task? Can you recall any test-takers' writing?
5. What about key characteristics/features of average essays, poor and very poor essays? Can you recall any test-takers' writing?
6. What challenges do you normally face in rating the real IELTS independent task?
7. What challenges did you face in rating the independent task in this study?  
And using the scoring rubric?
8. Were there any of the features of the writing that you think impacted your rating more than others in the independent task? Or what writing issues (e.g., rhetorical/ideational) did you attend to?

### Integrated task:

9. What do you think are key characteristics or features of very good essays in the integrated task? Can you recall any test-takers' writing?
10. Key characteristics or features of poor or very poor essays in the integrated task?  
Can you recall any test-takers' writing?
11. What challenges did you face in rating the integrated task? And using the scoring rubric?
12. Did you consider the source texts as you rated? In what way? Did you assess the quality of source use (e.g., accuracy, relevance, adequacy, clarity, appropriacy, effectiveness or overuse of source)?
13. Did you find instances where writers' copied? If so, how did you detect this?  
How did it affect scoring?
14. What did you observe about common patterns of source use?
15. Were there any of the features of the writing that you think impacted your rating more than others?
16. Was the scoring rubric useful for rating? Could it be improved?  
What would you suggest?
17. Did you experience the borderlines (falling between two scores) of assigning scores? If so, what did you do?

### Perceptions and recommendations

18. What are your recommendations or strategies for raters when they rate the integrated task?
19. What do you think about including the integrated task in the real IELTS?  
What do you think are potential challenges? And how they may be addressed?
20. Do you think a reading-to-write task (integrated) would help international students better prepare with their academic study?
21. Do you have any other comments, issues or concerns you would like to raise about the integrated task?

## Appendix 8: Analytical rubric for the independent writing task

Aspect	6 (Very good user)	5 (Good user)	4 (Competent user)	3 (Modest user)	2 (Limited user)	1 (Extremely limited/non-user)
<b>Task achievement</b>	Fully addresses all parts of the task; presents a fully developed position in answer to the question with relevant, fully extended and well-supported ideas.	Addresses all parts of the task; presents a clear position throughout the response; presents, extends and supports main ideas, but there may be a tendency to overgeneralise and/or supporting ideas may lack focus.	Addresses the task only partially; the format may be inappropriate in places; expresses a position but the development is not always clear and there may be no conclusions drawn; presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail.	Responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate; presents a position but this is unclear; presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported.	Does not adequately address any part of the task; does not express a clear position; presents few ideas, which are largely undeveloped or irrelevant.	Answer is completely unrelated to the task; does not attend; does not attempt the task in any way; writes a totally memorised response.
<b>Coherence and cohesion</b>	Uses cohesion in such a way that it attracts no attention; skilfully manages paragraphing.	Logically organises information and ideas; there is clear progression throughout; uses a range of cohesive devices appropriately although there may be some under-/over-use; presents a clear central topic within each paragraph.	Presents information with some organisation but there may be a lack of overall progression; makes inadequate, inaccurate or over use of cohesive devices; may be repetitive because of lack of referencing and substitution; paragraphing may be inadequate/confusing.	Presents information and ideas but these are not arranged coherently and there is no clear progression in the response; uses some basic cohesive devices but these may be inaccurate or repetitive; may not write in paragraphs.	Does not organise ideas logically; may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas.	Fails to communicate any message.

Aspect	6 (Very good user)	5 (Good user)	4 (Competent user)	3 (Modest user)	2 (Limited user)	1 (Extremely limited/non-user)
<b>Lexical resource</b>	Uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'.	Uses a sufficient range of vocabulary to allow some flexibility and precision; uses less common lexical items with some awareness of style and collocation; may produce occasional errors in word choice, spelling and/or word formation.	Uses a limited range of vocabulary, but this is minimally adequate for the task; may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader.	Uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task; has limited control of word formation and/or spelling; errors may cause strain for the reader.	Uses only a very limited range of words and expressions with very limited control of word formation and/or spelling; errors may severely distort the message.	Can only use a few isolated words.
<b>Grammatical range and accuracy</b>	Uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'.	Uses a variety of complex structures; produces frequent error-free sentences; has good control of grammar and punctuation but may make a few errors.	Uses only a limited range of structures; attempts complex sentences but these tend to be less accurate than simple sentences; may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader.	Uses only a very limited range of structures with only rare use of subordinate clauses; some structures are accurate but errors predominate, and punctuation is often faulty.	Attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning.	Cannot use sentence forms at all.

Source: Adapted from the IELTS Writing Task 2 Rubric (public) the British Council, IDP: IELTS Australia and the University of Cambridge ESOL Examination

## Appendix 9: Analytical scoring rubric for the integrated writing task

Aspect	6 (Very good user)	5 (Good user)	4 (Competent user)	3 (Modest user)	2 (Limited user)	1 (Extremely limited/ non-user)
<b>Task achievement</b>	Fully addresses all parts of the task; presents a fully developed position in answer to the question with relevant, fully extended and well-supported ideas.	Addresses all parts of the task; presents a clear position throughout the response; presents, extends and supports main ideas, but there may be a tendency to overgeneralise and/or supporting ideas may lack focus.	Addresses the task only partially; the format may be inappropriate in places; expresses a position but the development is not always clear and there may be no conclusions drawn; presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail.	Responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate; presents a position but this is unclear; presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported.	Does not adequately address any part of the task; does not express a clear position; presents few ideas, which are largely undeveloped or irrelevant.	Answer is completely unrelated to the task; does not attend; does not attempt the task in any way; writes a totally memorised response.
<b>Source text</b>	Successfully selects important and relevant information from the given texts; coherently and accurately synthesise information from the two given texts; appropriately and accurately acknowledges source texts.	Generally good in selecting relevant information from the given texts; may have some minor omission, inaccurate, vagueness or imprecision of some content from the two given texts; generally good in acknowledging source texts.	Contains some important and relevant information from the given texts; conveys only vague, global, unclear or somewhat imprecise points made from the two given texts; somewhat acknowledges source texts.	Contains some relevant information from the given text; presents significant omission, or inaccurate important ideas from the two given texts; misrepresents information from the two given texts; acknowledges source texts; limited to no acknowledgment of source texts. Over-dependence on input material/source text.	Provides little or no meaningful or relevant information from the given two texts.	No evidence in use of information from the given two texts; or merely copies sentences verbatim from the two given texts.
Aspect	6 (Very good user)	5 (Good user)	4 (Competent user)	3 (Modest user)	2 (Limited user)	1 (Extremely limited/ non-user)

<b>Coherence and cohesion</b>	Uses cohesion in such a way that it attracts no attention; skilfully manages paragraphing.	Logically organises information and ideas; there is clear progression throughout; uses a range of cohesive devices appropriately although there may be some under-/over-use; presents a clear central topic within each paragraph.	Presents information with some organisation but there may be a lack of overall progression; makes inadequate, inaccurate or overuse of cohesive devices; may be repetitive because of lack of referencing and substitution; paragraphing may be inadequate/confusing.	Presents information and ideas but these are not arranged coherently and there is no clear progression in the response; uses some basic cohesive devices but these may be inaccurate or repetitive; may not write in paragraphs.	Does not organise ideas logically; may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas.	Fails to communicate any message.
<b>Lexical resource</b>	Uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'.	Uses a sufficient range of vocabulary to allow some flexibility and precision; uses less common lexical items with some awareness of style and collocation; may produce occasional errors in word choice, spelling and/or word formation.	Uses a limited range of vocabulary, but this is minimally adequate for the task; may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader.	Uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task; has limited control of word formation and/or spelling; errors may cause strain for the reader.	Uses only a very limited range of words and expressions with very limited control of word formation and/or spelling; errors may severely distort the message.	Can only use a few isolated words.
<b>Grammatical range and accuracy</b>	Uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'.	Uses a variety of complex structures; produces frequent error-free sentences; has good control of grammar and punctuation but may make a few errors.	Uses only a limited range of structures; attempts complex sentences but these tend to be less accurate than simple sentences; may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader.	Uses only a very limited range of structures with only rare use of subordinate clauses; some structures are accurate but errors predominate, and punctuation is often faulty.	Attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning.	Cannot use sentence forms at all.

Source: Adapted from the IELTS Writing Task 2 Rubric (public) the British Council, IDP: IELTS Australia and the University of Cambridge ESOL Examination and TOEFL Integrated Rubric by Educational Testing Service

## Appendix 10: Coding schemes for test-taker interviews

Abbreviated Code	Full code	Explanation	Example
<b>Inde-process</b>	Independent writing processes	This code refers to test-takers' reported writing or thinking processes during an independent writing task. These include <u>test-taking strategies</u> such as understanding task instructions and requirements, knowing how to start and submit the test; <u>writing strategies</u> such as paragraphing, using topic sentences, providing supports to viewpoints or positions; <u>metacognitive strategies</u> such as planning, outlining, brainstorming, monitoring and evaluation.	"I first read the instructions." "I think about what to include in my essay." "I include a topic sentence at the beginning of a paragraph." "I made some revisions." "I did not have enough time to double-check my essay."
<b>Influ-factors</b>	Influential factors	This code refers to factors, issues or aspects that test-takers report to influence their writing such as ideas, concepts, quality of writing. Influential factors include either inhibiting or enhancing factors.	"I [don't] have relevant background knowledge in this topic." "I had a difficulty to understand task requirements". "I am running out of time." "I know basic academic writing convention."
<b>Inte-process</b>	Integrated writing processes	This code refers to test-takers' reported writing or thinking processes during an integrated writing task. Similar to those in "Inde-process", it includes <u>test-taking strategies</u> such as understanding task instructions and requirements; <u>reading strategies</u> such as extracting main ideas or specific information from the source texts; <u>writing strategies</u> such as paragraphing, using topic sentences, supporting arguments or positions; <u>metacognitive strategies</u> such as planning, goal-setting, monitoring and evaluation	"I try to familiarise myself with the task." "I focused on what the task asked me to do." "I scanned through both texts quickly." "I came back to the first text to see what I can use". "I did not write an outline on a paper, but I had an idea what to include in my mind." "I prefer to paraphrase than using direct quotes." "I checked my writing as I wrote." "I wanted to double-check my writing before submitting." "I forgot to acknowledge the source text."
<b>Rel-useful-Inte</b>	Perceived relevance and usefulness of an integrated writing task	This code refers to test-takers' reported or perceived <u>[ir]relevance or [un]usefulness</u> of an integrated writing task. It includes explanations why they think it is or is not relevant or useful.	"I think an integrated task is useful to help me prepare for my future study." "It helps me understand the value of reading skills that help my writing." "I don't think this task is needed."
<b>Self-assess-write</b>	Self-assessment of current writing ability	This code refers test-takers' self-assessment of the own writing ability, for example, a perceived level of their current writing.	"I think my writing is at an intermediate level." "My writing is not very good."

<b>Task-challenge</b>	Challenges or difficulties in task completion	This code refers to experienced or perceived challenges during task completion in both writing tasks.	<p>"It was difficult for me to concentrate."</p> <p>"I don't know this topic at all. Therefore, it was hard for me to come up with good vocabulary or examples."</p> <p>"I have never done this type of writing test before. I am not sure how to address this task."</p> <p>"I am running out of time. I could not complete the last sentence."</p>
<b>Write-weak</b>	Perceived writing weaknesses	This code refers to test-takers' perceptions on their writing weaknesses, for example, grammatical accuracy, vocabulary range, topical knowledge.	<p>"I think I am not good at ..."</p> <p>"My problem in writing ..."</p>
<b>Write-strength</b>	Perceived writing strengths	This code refers to test-takers' perceptions on their writing strengths, for example, grammatical accuracy, vocabulary range, topical knowledge.	<p>"I am good at ..."</p> <p>"My writing strengths include ..."</p> <p>"I am not worried about X issues in my writing."</p>

## Appendix 11: Coding schemes for raters' integrated writing comments

Abbreviated Code	Full code	Explanation	Example
<b>Clear-posit</b>	Clear position	This code refers to comments about essays that present a clear position. It is stated explicitly in their essays.	"A very solid argument with a clear position presented." "A clear position taken by this candidate, who aligns themselves with Text 2."
<b>Inef-source</b>	Ineffective/ inaccurate/improper source text use	This code refers to comments about a lack of acknowledgements of the source texts being used. They can be about indirect or broad use of source texts without referencing or citations.	"There is use of source material, but it is not acknowledged." "Last paragraph is inspired by Source (2) however does not explicitly reference the source." "... the source texts which are used, but are not adequately acknowledged."
<b>Insuf-resp</b>	Insufficient responses	This code refers to comments about the length being short or under-length, insufficient to assess.	"The candidate responds in a minimal way." "159 words only. No paragraphing." "A short response with relevant ideas but not much development." "The response is short with limited development."
<b>Inte-source-own</b>	Integrated source texts with own knowledge	This code refers to comments about an effective use of source text integration with test-takers' own knowledge or source in essays.	"A good attempt at including both source material and original language." "The writer uses a combination of their own knowledge/opinions and the source material to develop and argument."
<b>No-source</b>	No, little or minimal use of source texts	This code refers to comments about absence or no source texts in test-takers' writing at all, very little or minimally (i.e., lack of clear evidence)	"No evidence source texts have been used." "Very minimal use of the sources in addition to the response being short." "There is only one reference to the source material, which is irrelevant to the point being made."
<b>Over-source</b>	Over-reliance on source texts	This code refers to comments about over-reliance or overuse of source texts in essays. This could be a mere summary of the source texts as essays. There is evidence of repetitiveness of source text use.	"The other paragraphs were copied." "Almost all the essay uses the source material." "A third of it is the source." "There is an over-reliance on source material." "The candidate summarises the articles in the beginning of the essay."

<b>Posit-arg-dev+</b>	Position/argument/development present	This code refers to comments about essays that present positions or arguments, but they may not be as clear. There is evidence of essay development.	<p>"There is a position that provides support using some of the ideas from the text."</p> <p>"An argument is there with the candidate using mainly their knowledge and a small part of the source material."</p>
<b>Some-rel-source</b>	Some relevant source text use	This code refers to comments about evidence in essays that present some evidence of selection and engagement of source text use that is relevant to the task. Test-takers may attempt to paraphrase source texts in their own words. However, their source text use may still be incomplete or underdeveloped.	<p>"Good attempts throughout to paraphrase source material (though this does result in basic errors and imprecision)."</p> <p>"The sources used are somewhat relevant but there does not seem to be much connection to the point the writer is making."</p>
<b>Sound-argu</b>	Sound/clear argument	This code refers to comments about essays that present sound or clear arguments, essays that include reasonable, convincing, support from either source text use or own sources.	<p>"A sound argument is given for global warming."</p> <p>"A clear, convincing response, with good attempts to include examples from the source texts."</p>
<b>Under-dev-posit</b>	Under/unclear development/position	This code refers to comments about essays that are underdeveloped such as a lack of position or argument. Ideas may be confusing, incohesive or unsubstantiated.	<p>"The candidate provides an argument although without clear direction with the supporting details."</p> <p>"There is an argument but there needs to be more cohesion between the main ideas presented and supporting ideas."</p> <p>"The body of this essay requires further extension and development."</p>

## Appendix 12: Coding schemes for rater interviews

Abbreviated Code	Full code	Explanation
<b>Comment-IELTS</b>	Comments on current IELTS Writing tasks	This code refers to raters' comments about the appropriateness and usefulness of the current IELTS Writing tasks.
<b>Comment-source</b>	Comments on source text use	This code refers to raters' comments on source text use by test-takers. This includes poor, average or appropriate source text use.
<b>Essay-feat</b>	Essay features	This code refers to raters' comments on essay features based on their expertise and experiences as an IELTS examiner, as well as essay features they found in the current project. Essay features can be about poor, average, good or excellent essays. Their comments can be about independent and/or integrated essays.
<b>Feat-inte-write</b>	Features of observed integrated essays	This code refers to raters' comments or perspectives about various features or characteristics of poor, average or good integrated essays.
<b>Influ-factors-inte</b>	Influential factors on integrated essays	This code refers to raters' perceptions or viewpoints on issues or factors that affect test-takers' responses to the integrated writing task. These factors may be positive or negative to essay performance.
<b>Percep-inte-write</b>	Perception toward an integrated writing task	This code refers to raters' comments or viewpoints about an inclusion of an integrated essay task in an official IELTS test. This includes what they see as usefulness, relevance, benefits and caveats.
<b>Rater-training</b>	Rater training/meetings	This code refers to raters' discussion, perception and comments about the rater training and meeting sessions prior to actual scoring.
<b>Rating-Chal</b>	Rating challenges	This code refers to raters' comments or expressions about challenges they face when rating either or both independent and integrated essays.
<b>Rating-proc</b>	Rating processes	This code refers to raters' reported rating processes and criteria they use. It includes how much time they spend on each scoring criterion or essays.
<b>Sug-inte-use</b>	Suggestions for integrated task use	This code refers to raters' suggestions or recommendations on how an integrated writing task can be appropriately designed, implemented and used. This includes their statements about instructions, source texts, instructions for test-takers, raters and rater training.
<b>Use-score-rub</b>	Use of scoring rubrics	This code refers to raters' reported use or issues in scoring rubrics for either or both independent and integrated essays.