

LTTC-CRELLA Collaboration Project RG-03

Examining the context and cognitive validity of
the GEPT Advanced Writing Task 1:
A comparison with real-life academic writing
tasks

Sathena H. C. Chan
Rachel Y. F. Wu
Cyril J. Weir

**Examining the context and cognitive validity of the GEPT
Advanced Writing Task 1:
A comparison with real-life academic writing tasks**

**LTTC-CRELLA
Collaboration Project RG-03**

**Sathena H. C. Chan
Rachel Y. F. Wu
Cyril J. Weir**

LTTC-CRELLA Collaboration Project RG-03

Examining the context and cognitive validity of the GEPT Advanced Writing Task 1:

A comparison with real-life academic writing tasks

Published by The Language Training and Testing Center
No. 170, Sec.2, Xinhai Rd., Daan Dist., Taipei City, Taiwan 10663

© The Language Training and Testing Center, 2014

All rights reserved. No parts of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the prior written permission of The Language Training and Testing Center.

First published April 2014

Foreword

It is with pleasure that the LTTC brings to you the LTTC-CRELLA Collaboration Project. The study reported in this volume was the result of an eighteen-month collaborative research conducted by the LTTC and the Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire. The study was funded by the LTTC, while personnel from both CRELLA and LTTC were jointly involved in completing the project. The research reflects LTTC's commitment to build and maintain mutually beneficial relationships with the academic community.

This project was the extension of the 2011-2012 LTTC-GEPT Research Grants Project RG-01 (Phase 1). In the Phase 1 study, the criterion-related validity of the GEPT Advanced Reading and Writing Test was investigated, with a focus on the cross-test comparability and the predictive power of the test scores. The study revealed moderate to strong positive correlations between the GEPT Advanced Reading and Writing and IELTS. It also showed significant correlations between GEPT Advanced Reading and Writing tests and real-life academic performances.

As a follow-up study, this research compared the contextual and cognitive validity of Advanced Writing Part 1 (summary from verbal input) with real-life academic writing tasks. Textual analysis and questionnaire survey results revealed that the overall task setting and features of the GEPT as well as the cognitive processes elicited by the GEPT test task were comparable to those of the real-life input texts. Full research reports can be downloaded at <http://www.lttc.ntu.edu.tw/RReports.htm>.

The GEPT, developed more than a decade ago by the LTTC to serve as a fair and reliable testing system for EFL learners, has gained wide recognition in Taiwan and abroad. We believe that through consistent collaboration between the external research community and the LTTC research team, the GEPT will continue to refine its quality and achieve wider recognition at home and overseas.

Complete and more up-to-date information about the GEPT is available at

<https://www.lttc.ntu.edu.tw>.

A handwritten signature in black ink, appearing to read 'Hsien-hao Liao'.

Hsien-hao Liao
Executive Director
LTTC

Author Biodata

Dr. Sathena H. C. Chan is a post-doctoral research fellow at the Centre for Research in English Language Learning and Assessment (CRELLA), the University of Bedfordshire. She is the principal investigator for different test validation research studies and test review/revision projects for examination boards and educational organisations in the UK and worldwide. Her research interests include integrated reading-into-writing test development and validation, cognitive validation in language testing and criterial features of writing performance at different levels.

Dr. Rachel Y. F. Wu is the Head of Testing Editorial Department at the Language Training and Testing Center, Taipei. She has been closely involved in the research and development, test production and validation of the General English Proficiency Test. She also serves on the GEPT Examination Committee. Her research interests include reading assessment, language test development and validation, and methodological approaches to linking examinations to the *CEFR*.

Prof. Cyril J. Weir is currently Powdrill Professor in English Language Acquisition and the Director of the Centre for Research in English Language Learning and Assessment (CRELLA) at the University of Bedfordshire. He has a PhD in language testing. He is the author of a wide range of educational publications, including *Communicative Language Testing*, *Understanding and Developing Language Tests* and *Language Testing and Validation: An evidence-based approach*. He is also the co-author of *Examining Writing*, *Examining Reading*, *Evaluation in ELT*, *Reading in a Second Language*, *Measured Constructs*, and the co-editor of six *Studies in Language Testing* (SiLT) volumes. He has taught short courses and carried out consultancies in language testing, evaluation and curriculum renewal in over fifty countries worldwide. Current interests include language construct definition, the history of English language tests in the UK and validation. He is a member of the Editorial Board of *Language Testing* and is co-editor for the *Studies in Language Testing* series published by Cambridge University Press.

摘要

◆ 研究團隊與研究目的

本研究計劃由 LTTC 與 CRELLA, the University of Bedfordshire 研究團隊共同執行，比較「全民英檢」高級寫作測驗第一部分與大學學術寫作作業的相似度，藉以檢視「全民英檢」高級寫作測驗中的語境效度(context validity)和認知效度(cognitive validity)。

◆ 研究問題

第一階段

1. 就語境特徵而言，「全民英檢」高級寫作測驗第一部分與大學學術寫作作業是否相似？
2. 就涉及的認知過程而言，「全民英檢」高級寫作測驗第一部分與大學學術寫作作業是否相似？

第二階段

3. 參加「全民英檢」高級正式測驗的台灣考生，在進行寫作第一部分時涉及哪些認知過程？

◆ 受試者

第一階段

160 位就讀英國一所大學的學生。

第二階段

192 位在台灣參加正式「全民英檢」高級測驗的考生。

◆ 研究結果摘要

1. 就題目要求而言，「全民英檢」高級寫作第一部分與大學作業大致相同，例如，二者所涉及的語言功能大致相同。然而，大學作業的主題較偏向「學術」或「專業」性質，而「全民英檢」的主題較偏社會議題。
2. 「全民英檢」與大學作業皆要求受試者根據數篇文本的內容寫作，而且閱讀材料的語境特徵相近。但是「全民英檢」的閱讀材料主要為論說文，大學作業閱讀材料則包括論說文和說明文。
3. 「全民英檢」與大學作業閱讀材料的詞彙複雜度、句構複雜度和內容組織難度相當，但是「全民英檢」的材料所陳述的概念較具體。
4. 大學作業所涉及的寫作認知過程與「全民英檢」大致相同，例如，受試者寫大學作業和考「全民英檢」時都需要搜尋與題目相關的資訊、建立資訊間的關聯，並進行精讀等。

Abstract

This present study investigates the context and cognitive validity of the GEPT Advanced Writing Task 1. Regarding its context validity, this study applied both expert judgement and automated textual analysis to examine the degree of correspondence between the overall task setting and input text features of the GEPT task and those of the target academic writing tasks in real life university courses in the UK. Regarding its cognitive validity, this study examined the cognitive processes elicited by the GEPT task as compared to the real-life academic writing tasks through a cognitive process questionnaire. In the main study (Phase 1), data were collected from 160 students who were studying at a British University. Additional data were collected from 192 test takers under live test conditions in Taiwan in a follow-up study (Phase 2).

The demonstration of a close similarity between test and real-life conditions in the findings of this study supports the context and cognitive validity of the GEPT Advanced Writing Task 1. The GEPT task resembled the *overall task setting* and *input text features* of the real-life tasks in a number of important ways. The difficulty level of the GEPT input texts and the real-life input texts was comparable in terms of most of the automated textual analysis indices utilised by this study. There was no significant difference in 14 out of the 17 indices of lexical complexity, syntactic complexity and degree of coherence. Results of the real-life cognitive processing data identified eleven cognitive validity parameters: (1) task representation and macro-planning; (2) revising macro plan; (3) connecting and generating; (4) selecting relevant ideas; (5) careful global reading; (6) organising ideas in relation to input texts; (7) organising ideas in relation to writer's own text; (8) low-level editing during writing; (9) low-level editing after writing; (10) high-level editing during writing and (11) high-level editing after writing. The GEPT task was able to elicit from high-achieving participants most of the cognitive processes they employed in the real-life tasks. In terms of the underlying structure of the cognitive processes, exploratory factor analyses on the GEPT data (Phase 1 and Phase 2) showed that although a few individual items yielded complex loadings, the GEPT task was largely able to elicit from the participants the same underlying cognitive processes as the real-life tasks did. In particular, the eleven individual factors across the five cognitive phases yielded by the GEPT live test data resembled closely those factors yielded by the real-life data.

According to the results of this study, GEPT Advanced Writing Task 1, an integrated reading-into-writing test, demonstrates both context and cognitive validity. The results have important implications for university admissions officers and other stakeholders; in particular they demonstrate that the GEPT Advanced Writing Task 1 is a valid option when considering writing tests for academic purposes.

Table of Contents

Acknowledgements	1
1. Scope of the study	3
2. Theoretical framework	3
2.1 Theoretical support for contextual and cognitive validation	3
2.2 Previous studies relevant to the context and cognitive validity of writing tests for academic purposes	4
3. Research Questions	9
4. Methodology	9
4.1 The tasks	9
4.2 Research design: the context validity of GEPT Advanced Writing Task 1	11
4.2.1 Participants (expert judges)	11
4.2.2 Data collection methods and instruments	11
4.2.3 Data collection procedures	14
4.2.4 Method of data analysis	15
4.3 Research design: cognitive validity of the GEPT Advanced Writing Test Task 1	16
4.3.1 Participants	16
4.3.2 Writing Process Questionnaire	16
4.3.3 Data collection procedures	17
4.3.4 Method of data analysis	18
5. Results	19
5.1 The contextual validity of GEPT Advanced Writing Task 1	19
5.1.1 Results from expert judgement – overall task setting	19
5.1.2 Expert judgement - input text features	24
5.1.3 Automated textual analysis – the linguistic complexity of the input texts	27
5.2 The cognitive validity of GEPT Advanced Writing Task 1	30
5.2.1 Defining the target cognitive parameters in the real-life context	30
5.2.2 Investigating the cognitive validity of GEPT Advanced Writing Task 1	47
5.3 The cognitive processes elicited under the GEPT live test condition	64
5.3.1 Comparisons of the cognitive processes elicited by the GEPT task: university context in the UK vs. live test conditions in Taiwan	64
5.3.2 The underlying constructs elicited by GEPT Advanced Writing Task 1 under live test conditions in Taiwan	65

5.3.3 Summary	71
6. Conclusion	72
6.1 Summary of the main findings.....	72
6.2 Limitations of the study	75
6.3 Implications of the findings	75
Reference	78
Appendix.....	83
Appendix 1: Glossary of the Contextual Parameter Proforma	83
Appendix 2: Feedback Evaluation Questionnaire.....	84
Appendix 3: Writing Process Questionnaire.....	86
Appendix 4: Comparison of the processes elicited by the two real-life tasks	89
Appendix 5: Rejected factor solutions (real-life).....	90
Appendix 6: Rejected factor solutions (GEPT)	91

Acknowledgements

This research project was based on close collaboration between CRELLA and the LTTC. We would like to express our gratitude to the LTTC, which has funded this research and provided valuable test materials and data for analyses. Without the LTTC's support, this study would not have been possible. We would also like to extend our gratitude to Dr. Fumiyo Nakatsuhara (CRELLA) for her invaluable advice at various stages of this study, and Mr. Chia-Lung Lee, a statistician from the LTTC, for his analysis of the Phase 2 questionnaire data.

1. Scope of the study

Research demonstrating the construct validity of the GEPT Advanced Level Test is crucial in order for the test to gain acceptance from educational institutions as a valid measure of international students' proficiency in English for academic purposes. This research followed up Weir, Chan and Nakatsuhara's (2013) study on the criterion-related validity of the GEPT Advanced Test. Their study showed moderate to strong positive correlations between the GEPT Advanced Reading and Writing Test and IELTS. In addition, the GEPT Advanced Reading and Writing Test scores accounted for 27.89% variance of real-life academic performance at a correlation of 0.529 ($p < 0.01$). Weir et al (2013:18) stressed the importance of conducting follow-up studies to investigate the validity of GEPT Advanced Reading and Writing Test in terms of the remaining construct validity components.

The primary aim of this study is to validate the GEPT Advanced Level Writing Test in terms of a) context validity and b) cognitive validity in two phases.

- Phase 1 investigates contextual and cognitive validity by comparing the GEPT Advanced Level Writing Task 1 with real-life writing tasks assigned to undergraduates on Business programmes¹ in a UK university.
- Phase 2 is a follow-up investigation of the cognitive validity of the GEPT Advanced Level Writing Task 1 conducted under live test conditions in Taiwan.

In complementing the findings from Weir et al.'s (2013) study, the present study seeks to provide evidence on which to build a more complete validity argument for the use of the GEPT Advanced Level Writing Test as a test for university admission². Drawing upon Weir's (2005) socio-cognitive framework for test validation, this study gathers context validity evidence by using systematically defined key contextual parameters which define the overall task setting and textual features of the source texts of the GEPT Advanced Level Writing Test Task 1. Cognitive validity evidence is gathered by identifying the cognitive load imposed by task setting and the actual cognitive processes elicited by the task.

2. Theoretical framework

2.1 Theoretical support for contextual and cognitive validation

This study focuses on two components of Weir's (2005) socio-cognitive framework for test development and validation – context validity and cognitive validity. Context validity and cognitive validity are a priori components of test validation to be obtained before live testing. The a priori evidence of test validation enables test stakeholders to define and evaluate the integral nature and quality of a test. For writing tests, context validity addresses the appropriateness of the linguistic and content demands set by a writing test task in comparison with the contextual parameters of the target writing tasks in real life. Cognitive validity in writing tests, on the other hand, measures 'how closely a writing task represents the cognitive processing involved in writing contexts beyond the test itself, i.e. in performing the task in real life' (Shaw and Weir, 2007:34). The framework has already been extensively used by different professional testing bodies worldwide. For instance, Cambridge English Language Assessment has used the framework to gather validity evidence for its English examinations (e.g. Geranpayeh & Taylor (eds), 2012 - Examining Listening; Khalifa & Weir, 2009 - Examining Reading; Shaw & Weir, 2007 - Examining Writing; Taylor (ed), 2011 - Examining

¹ Business and administrative studies contained the highest proportion of international students (36%) (UKCISA, 2012)

² The cognitive validity of the GEPT Advanced Level Writing Test Task 2 is being investigated by Yu and Lin (forthcoming).

Speaking). Chan (2013) extended the application of the framework to validation of integrated reading-into-writing tests. This study will investigate the context and cognitive validity of the GEPT Advanced Task 1 using the parameters refined in the above mentioned studies.

2.2 Previous studies relevant to the context and cognitive validity of writing tests for academic purposes

What contextual features should a valid writing test task for academic purposes operationalise?

For writing tests, context validity addresses the appropriateness of the linguistic and content demands set in the test task in comparison with the contextual parameters of the writing tasks in the target language use context (Weir, 2005; Shaw & Weir, 2007). Bachman and Palmer (1996: 44) defined target language use domain as '*a set of specific language use tasks that the test taker is likely to encounter beyond the test itself, and to which we want our inferences about language ability to generalise*' (1996: 44). The importance of developing test tasks which are representative of the target language use domain is repeatedly stated in the literature (Bachman, 1990; Bachman & Palmer, 1996; Weigle, 2002; Weir, 1983, 1990 and 2005).

Researchers have surveyed the tasks that are required of students in educational contexts in recent decades (e.g. Bridgeman & Carlson, 1983; Carson, 2001; Horowitz, 1986a, 1986b; Johns, 1993; Leki & Carson, 1994; Weir, 1983). By surveying teachers in 190 academic departments across undergraduate and postgraduate levels in Canada and USA, Bridgeman & Carlson (1983) found that *description and interpretation of non-verbal input* and *comparison and contrast plus taking a position* were the two task types perceived as the most typical by teachers. Hale et al (1996) analysed actual writing tasks assigned in 162 undergraduate and graduate courses in several disciplines at eight universities. They found that the most common real-life tasks across disciplines and levels were *short tasks* (i.e. *writing tasks which require students to produce an output about half a page long*), *essays*, *summaries*, *reports with interpretation* and *research papers*. Similarly, based on an analysis of writing tasks in 38 faculties, Horowitz (1986a, 1986b) found that reading was essential in the most common academic writing task types. The common tasks he identified included *synthesis of multiple sources*, *connection of theory and data*, *report*, *research report* and *summary*. Among these types, *synthesis of multiple sources* was the most typical across the 38 faculties. More recently, Cooper & Bikowski (2007) analysed 200 graduate course syllabi from 10 academic departments with follow-up interviews at one university, and found that *library research papers* and *project reports* were the most commonly assigned writing tasks across different disciplines.

The findings of these studies clearly show that most real-life academic writing tasks require students to write drawing upon external materials. However, reading and writing have largely been regarded as two independent constructs in most language tests. Horowitz (1991) argues that there is a fundamental discrepancy regarding the use of source texts between most independent writing test tasks and real-life writing tasks. Similarly, Moore and Morton (1999, 2005), in their study to compare the contextual parameters between real-life academic writing tasks and IELTS Academic Writing Task 2 (i.e. an impromptu argumentative essay task), suggested that the major gap between real-life tasks and academic writing test tasks was in the use of secondary source in the writing process.

Whilst these studies indicate that integrated reading-into-writing tasks appear to have a stronger claim to context validity, especially in the academic context (Plakans, 2008, 2010; Weigle, 2004; Weir et al, 2013), there is a lack of research which systematically compares the contextual features of such integrated reading-into-writing test tasks and real-life academic writing tasks. The GEPT Advanced Level Writing Task 1 as an integrated task type involving multiple reading inputs and therefore has the potential to satisfy the need for greater validity in the assessment of test takers' academic writing ability. This study, therefore, aims to generate empirical evidence regarding the contextual parameters set in the GEPT Advanced Level Writing Task 1, and to investigate the extent to which the test task resembles the contextual features of real-life academic writing tasks.

The features of the GEPT Advanced Level Writing Task 1 and real-life academic writing tasks were analysed using the following context validity parameters, which are considered to be those most relevant to the present study.

Table 1: Context validity parameters in reading-into-writing tests for academic purposes (adapted from Chan, 2013; Khalifa and Weir, 2009; Shaw and Weir, 2007)

Context validity parameters of academic writing tasks with integration of reading materials
Overall task setting (productive demands)
• Clarity of purpose
• Topic domain
• Genre
• Cognitive demands
• Language functions to perform
• Clarity of writer-reader relationship
• Clarity of marking criteria
Input text features (receptive demands)
• Input format
• Verbal input genre
• Non-verbal input genre
• Discourse mode
• Concreteness of ideas
• Explicitness of textual organisation
• Cultural specificity
• Linguistic complexity <ul style="list-style-type: none"> ○ Lexical complexity ○ Syntactic complexity ○ Degree of cohesion

Section 4.2 will provide details of the research methods used to measure these parameters when applied to the GEPT Advanced Level Writing Task 1 and the real-life academic writing tasks in this study.

What cognitive processes should a valid writing test task for academic purposes elicit from test takers?

The investigation of the context validity of the GEPT Advanced Level Writing Task 1 will determine how closely the test task simulates the target tasks in real life in terms of the

demands of the task setting, with its specified input and expected output. There is also a need to collect evidence of the cognitive processes employed by test takers whilst they complete the GEPT Advanced Level Writing Task 1, and to investigate the extent to which these processes, elicited by the test task, resemble the processes that would normally be employed by writers in a real-life academic context.

The knowledge transforming approach to writing

Academic writing is widely regarded as a knowledge transforming process (Flower et al., 1990; Spivey, 1984, 1990, 1997; Weigle, 2002; Weir, Vidakovic, & Galaczi, 2013). Scardamalia & Bereiter (1987), from a pedagogical perspective, proposed the models of *knowledge telling* and *knowledge transforming* to differentiate between the characteristics of the writing processes of novice elementary school writers at one end, and those which characterise more advanced college, undergraduate and graduate writers at the other end of a continuum of writing expertise. The knowledge telling approach refers to a rather linear text generating process involving 'telling' existing knowledge and information available from memory which has been automatically activated by the cues provided in the writing task (Scardamalia & Bereiter, 1987). In contrast, the knowledge transforming approach is a complex problem-solving process. More advanced writers tend to have a high awareness of the conflicts between available internal and external resources and their writing goals. Scardamalia & Bereiter (1991) found that advanced writers put explicit effort into establishing task representation (i.e. an initial understanding of the writing task), setting writing goals, and constantly evaluating the available resources against their goals and constraints. Scardamalia & Bereiter (1987) argued that such an approach to writing leads to knowledge transformation, which can be in the form of an enhanced understanding of the subject knowledge or well-developed opinions about a particular topic. Critically, this process leads to the generation of novel ideas rather than the "retelling" of existing information. Nevertheless, it is important to note that the difference in the two writing approaches is also arguably influenced by the type of task. Researchers have argued that the integrated reading-into-writing task type is more likely to elicit the knowledge transforming approach from writers than the independent writing-only task type (Chan, 2011; Flower et al., 1990; Spivey, 1984, 1990, 1997; Weigle, 2002; Weir, Vidakovic, & Galaczi, 2013)

Writing consists of recursive multiple processes

In the writing literature, a considerable amount of research has been conducted in an attempt to establish the cognitive processes involved in writing. Different writing models (e.g. Hayes and Flower, 1980; Hayes, 1996; Grabe and Kaplan, 1996; Field, 2004; Shaw and Weir, 2007) have shown that writing is not a linear stand-alone process but involves multiple recursions of a particular set of processes. A highly influential model of writing was proposed by Hayes & Flower (1980) informed by verbal 'think-aloud' protocols. Hayes and Flower proposed that writing is an extended, goal-directed, problem-solving exercise which involves multiple recursions of *planning*, *translating* and *reviewing*. An updated version of the model (Hayes, 1996) expanded the number of the internal and external components which may impact on the writing processes. Internal factors include working memory, long-term memory resources and the motivation of the writers whereas external factors include the physical environment of the task (e.g. the text read so far, the writing medium) and the social environment of the task (e.g. the audience, other texts read while writing).

More recent models, which build upon psycholinguistic theory, offer a clearer account of how writing processes are influenced by internal factors, when compared to the above models. Kellogg (1996, 1999, 2001) made a strong argument for the importance of working memory

in writing. He proposed that the individual processes of writing draw upon different components of working memory, rather than seeing working memory as a unitary facility. Based upon Kellogg's (1996) model and Levelt's (1989) model of speaking, Field (2004, 2011) proposed a model which accounts for the phases that a writer goes through when they produce a text (as a productive language skill). Field proposed that writing, as a productive skill, involves the phases of conceptualisation, organisation, encoding (grammatical, lexical, graphic), execution and monitoring. An important issue for language testing is identifying which phases and processes are relevant for test development and validity. Building upon Kellogg's and Field's models, Shaw & Weir (2007) considered five processes: (1) *macro-planning*, (2) *organisation*, (3) *micro-planning*, (4) *translation*, and (5) *monitoring and revising* to be those most relevant to the discussion of the cognitive validity of writing tests. They argued that valid writing tests should elicit from test takers those core cognitive processes involved in real-life writing. The models discussed above provide a useful blueprint for researchers investigating the cognitive processes employed by writers. However, the processes of integrating reading materials into writing have largely been excluded from these models, even though some models (e.g. Hayes, 1996) have pointed out the essential role of reading as part of the writing process.

Whilst there is a rich body of research investigating models of reading and writing, there is very limited discussion regarding how the reading processes interact with the writing processes when a writer writes based on reading materials. Field (2004, 2008, 2013) proposed that receptive skills (i.e., listening and reading) involve phases of input decoding, lexical search, parsing, meaning construction, and discourse construction. From the perspective of language testing, Khalifa & Weir (2009) expanded the model of reading to include the processes of, in an ascending order of cognitive demands, word recognition, lexical access, syntactic parsing, establishing propositional meaning, inferencing, building a mental model, creating a text level representation and creating an intertextual representation. They also provided a detailed account of how these reading processes are tested in standardised reading tests at different levels. However, there is little discussion in the literature about how these reading processes fit into a model of academic writing.

Despite a lack of comprehensive models of writing from sources, a number of research studies have investigated the 'unique' processes involved in writing from sources (i.e. the processes which are typically not involved in reading comprehension or writing-only tasks). Spivey and colleagues (e.g. Mathison & Spivey, 1993; Spivey & King, 1989; Spivey, 1990, 1997, 2001) conducted a series of studies to investigate the processes involved in different writing tasks which require the use of reading materials. This body of work involved discourse synthesis which is defined as 'a process in which readers read multiple texts on a topic and synthesize them' (Spivey & King, 1989: 11). The findings of the studies showed that when writing from external reading materials, a writer transforms a new representation of meaning from multiple texts to their own text through three core processes: a) *selecting* relevant content from multiple texts, b) *organising* the content according to the writing goals and c) *connecting* the content from different sources and *generating* links between these ideas. The results of these studies indicate that reading-into-writing activities place higher cognitive demands on students than reading comprehension processes per se.

Academic writing might be more accurately understood as an integration of receptive (reading) and productive (writing) skills. Both receptive and productive language skills involve multiple cognitive phases, and each phase involves multiple processes. A major challenge for language testing is how to model these phases and processes within a test

validation framework. A series of studies have identified the most appropriate cognitive processes for independent writing examinations (Shaw & Weir, 2007) and independent reading examinations (Khalifa & Weir, 2009) for adult users at intermediate level upwards, i.e. B2-C2 in terms of the CEFR (the proficiency levels university students need to be at). From a language testing perspective, Chan (2013) synthesised the earlier models of writing, reading and discourse synthesis to propose a cognitive validity framework for integrated reading-into-writing tests for academic purposes (see Table 2).

Table 2: Cognitive validity parameters in reading-into-writing tests for academic purposes

Cognitive phases involved in academic writing with integration of reading materials
Conceptualisation (Kellogg, 1996, Field, 2004, 2011) is the first phase of productive skills where the writer develops an initial mental representation of a writing task. Processes involved at this phase include task representation (Flower et al, 1990) and macro-planning (Hayes and Flower, 1980; Shaw & Weir, 2007)
Meaning and discourse construction is a higher-level phase from the model of receptive skills (Field, 2004, 2013). Meaning and discourse construction is a phase when the writer contextualises abstract meanings based on the contextual clues provided in the writing task and their own schematic resources (background knowledge) (Field, 2004, 2013) and integrates the information from these different sources into a discourse representation (Brown & Yule, 1983). Processes involved at this phase include careful reading (Khalifa & Weir, 2009), search reading (Khalifa & Weir, 2009; Spivey, 1990, 1997), and connecting ideas from multiple sources and generating new meaning (Spivey, 1990, 1997).
Organising (Shaw & Weir, 2007; Spivey 1990, 1997) is a phase where the writer 'provisionally organises the ideas, still in abstract form, in relation to the text as a whole and in relation to each other (Field, 2004, 329)'.
Monitoring and revising (at high and low level) (Hayes and Flower, 1980; Kellogg, 1996; Shaw & Weir, 2007) is a phase where the writer checks the quality of the text. After monitoring, a writer usually revises the unsatisfactory parts of the text (Field, 2004, 330).

The cognitive processes elicited by GEPT Advanced Level Writing Task 1 and the real-life academic writing tasks were analysed based on these cognitive validity parameters, which are considered to be most relevant to the context of the present study. Micro-planning and translating are two important phases in writing. However, when compared to other processes, micro-planning and translating seem to be more difficult to report reliably unless through directed verbal protocols (Field, 2004). Previous studies which investigated these two processes usually focused on these individual processes solely under experimental settings (see Kellogg, 1994 for a review). For these reasons they were not investigated in this study. In addition, lower-level reading processes (i.e. decoding, lexical search, and parsing), which also play a fundamental role in any reading-into-writing activities, were not investigated in this study because students at the undergraduate level already have high automaticity in these lower-level reading processes. A writing process questionnaire was adapted from Chan (2013) to assist the participants in this study to self-report the extent to which they employed the processes of each of the cognitive phases. Section 4.3 will provide more details of the research method of investigating the cognitive validity of GEPT Advanced Level Writing Task 1 and the structure of the questionnaire.

3. Research Questions

The project consists of two phases. Phase 1 was the main study of both contextual and cognitive validity conducted in the UK while Phase 2 was conducted to investigate whether comparable results from Phase 1 RQ2 can be obtained under live test conditions in Taiwan. It should be noted that, as described in the previous section, Phase 1 was conducted in a target language use context (i.e. a UK university) whereas Phase 2 was conducted in the live test conditions of GEPT in Taiwan. The general comparability of the results for the cognitive processes elicited in both contexts (see Table 40 p.63) suggest that any differences occasioned by variability in the contextual parameters of the two test situations was for the most part negligible. Cognitive processes elicited by the same language test(s) would be expected to be largely generic where critical contextual parameters are not markedly different across settings.

Phase 1:

RQ1: How closely do the contextual features of the GEPT Advanced Level Writing Test Task 1 resemble those of the real-life academic writing tasks that students would normally encounter in a UK university?

RQ2: How closely do the cognitive processes elicited by the GEPT Advanced Level Writing Test Task 1 resemble those elicited by the real-life academic writing tasks?

Phase 2

RQ 3: What are the cognitive processes employed by Taiwanese test takers to complete the GEPT Advanced Level Writing Test Task 1 under live test conditions in Taiwan?

4. Methodology

The present study was designed to establish the context and cognitive validity of GEPT Advanced Level Writing Test Task 1 through a mixed methods approach. Firstly this section presents how the two academic writing tasks were selected as being representative of real-life academic tasks, and describes the basic features of these tasks. It then describes the research design with respect to the investigation of context validity and cognitive validity of GEPT Advanced Level Writing Test Task 1 (Sections 4.2 and 4.3). Each of these research design sub-sections presents the details of participants, data collection methods and instruments, data collection procedures and methods of data analysis.

4.1 The tasks

Real-life academic writing tasks

In order to sample the predominant writing tasks that students are expected to encounter in the UK real-life academic context, eight module handbooks were collected from the Business School at a University in the UK. The specification and assessment plan of each module were assessed according to the following criteria:

- enrolment rate of the module (i.e. as modules were optional, a module was considered for inclusion in the study only if more than 50% of the cohort enrolled)
- individual writing assignments only (i.e. no group/collective assignments)
- the type of input source involved
- the type of response text required

As a result, two writing tasks (real-life task A and real-life task B) were selected from two different modules to represent the real-life academic tasks that student would normally encounter in a UK academic context. Table 3 describes the basic features of these two real-

life tasks. For both tasks, students are expected to process both verbal and non-verbal input while performing the task.

Table 3 Basic features of the two real-life tasks

Real-life tasks	Task instructions	Input	Output
A	Write an essay on a given topic <ul style="list-style-type: none"> - Summarise salient issues - Discuss the issues with justified personal views 	<ul style="list-style-type: none"> • Verbal (a stimulus article) • Non-verbal (e.g. diagrams, graphs, etc) • Students are expected to use other input texts of their choice 	5000 words
B	Write a report to forecast the business of a company <ul style="list-style-type: none"> - Describe data - Discuss and justify ways of analysis - Make recommendations 	<ul style="list-style-type: none"> • Verbal (a description of the company) • Non-verbal (e.g. numeric dataset, charts, graphs etc.) • Students are expected to use other input texts of their choice 	2000 words

A small corpus of the real-life input texts

As indicated in Table 1, the context validity of integrated reading-into-writing concerns not only the overall task setting, but also the features of input texts. A small corpus of real-life input texts was built by sampling from the texts that students read to complete these two selected real-life tasks by the following steps:

1. 100 student scripts were collected from each of the selected real-life tasks, totalling 200 students' scripts.
2. The bibliography of each script was examined.
3. The ten most cited source texts from each of the two real-life tasks were identified.
4. Three extracts (from the beginning, middle and end of each text) were obtained from each of the twenty selected source texts. Each extract was about 500 words long.
5. A small corpus of real-life input texts, containing 60 extracts from the 20 most cited source texts, was computed.

GEPT Advanced Writing Task 1

The General English Proficiency Test (GEPT) Advanced, developed and administered by the LTTC, targets English learners at the CEFR C1 level. This test corresponds to Taiwan's English education framework, meets the specific needs of English learners in Taiwan for self-assessment, and provides institutions with a reference for evaluating the English proficiency levels of their students. Test-takers who pass this level have English language abilities which enable them to communicate fluently with only occasional errors related to language accuracy and appropriateness, and to handle academic or professional requirements and situations (LTTC, 2013). In particular, GEPT Advanced Writing aims to assess test takers' ability to:

- summarize articles on general and professional topics
- write well-organized and coherent essays, with appropriate lexical and grammatical usage
- express their opinions on a range of topics and discuss them in depth

GEPT Advanced Writing Task 1 is the first task of the Writing component. Table 4 below presents its basic features (for more information about GEPT, see LTTC, 2013).

Table 4 Basic features of the GEPT Advanced Writing Task 1

Test task	Test instructions	Time allowance	Input	Output	Function	Level
A	<ul style="list-style-type: none"> • Write a comparative essay summarising the main ideas from the verbal inputs and stating own viewpoint 	60 minutes	2 articles without non-verbal input	At least 250 words	Criterion-referenced level specific test	CEFR C1

4.2 Research design: the context validity of GEPT Advanced Writing Task 1

4.2.1 Participants (expert judges)

Sixteen judges with experience in language testing and/or language teaching were recruited to form an expert panel to evaluate the overall task setting of the real-life tasks and the GEPT task. Additionally, two judges were recruited from the same panel to evaluate the features of all input texts sampled in the study. All judges have at least 3-years experience in either language testing or language teaching, or both. Over half of the judges have more than ten-years experience in one of these fields.

4.2.2 Data collection methods and instruments

4.2.2.1 Contextual parameter proforma for expert judgement

A Contextual Parameter Proforma (see Table 5), which aims to facilitate an operationalisable analysis of the contextual features of reading-into-writing tasks by a group of judges, was used. The Proforma consists of two sections: *overall task setting* and *input text features*. Items 1 to 7 address the overall task setting which includes *genre*, *purpose*, *topic domain*, *cognitive demands*, *intended reader*, *knowledge of criteria* and *language functions to perform*. Items 8 to 14 address input text features, which cannot be effectively analysed by automated textual analysis tools (these tools are described in Section 4.2.2.2 below). The variables include *input format*, *genre*, *discourse mode*, *concreteness of ideas*, *explicitness of textual organisation* and *cultural specificity*.

Table 5 Contextual Parameter Proforma

Part 1 - Overall task setting																
1. Purpose	1 Unclear		2		3		4		5 Clear							
2. Topic Domain (Please circle a rating for each domain)	Personal				Social				Academic				Professional			
	Not at all 1	2	3	4	5	Not at all 1	2	3	4	5	Not at all 1	2	3	4	5	
3. Genre	Essay				Report		Case Study		Summary				Others (Please specify):			
4. Cognitive demands	1. Telling personal experience / viewpoints					2. Summarising / organising given ideas					3. Transforming given ideas into new representations					
5. Language functions to perform (you may choose more than 1)	Classifying				Citing sources		Describing		Defining		Evaluating					
	Persuading				Predicting		Recommending		Reasoning		Summarising					
	Synthesising (to combine different parts of texts to form a new text with own interpretations)				Expressing personal views		Illustrating visuals		Others (Please specify):							
6. Clarity of intended reader	1 Unclear		2		3		4		5 Clear							
7. Knowledge of criteria	1 Unclear		2		3		4		5 Clear							

Part 2 - Input text features											
8. Input format	Single verbal		Single non-verbal		Multiple verbal		Multiple non-verbal		Multiple verbal and multiple non-verbal		
	Others (Please specify):										
9. Verbal input genre	Book Chapter		Journal article		News / Magazine article		Proposal		Report		Review
	Others (Please specify):										
10. Non-verbal input	Table		Graph		Diagram		Picture				
11. Discourse mode (Consider the primary purpose of the text)	Narrative		Descriptive		Exposition		Argumentative				
12. Concreteness of ideas	1 Abstract		2		3		4		5 Concrete		
13. Explicitness of textual organisation	1 Implicit		2		3		4		5 Explicit		
14. Cultural specificity	1 Neutral		2		3		4		5 Specific		

4.2.2.2 Automated textual analysis tools

In addition to expert judgement, automated textual analyses were performed to analyse a range of textual features of the input texts in this study. Automated textual analysis has been regarded as a more systematic and efficient way to assess textual features than the more traditional expert judgement method, especially when a large number of texts are involved.

Many researchers have used automated textual analytic tools to evaluate the features of different types of texts such as L1 students' scripts (e.g. Crossley & McNamara, 2010), L2 students' scripts (e.g. Crossley & McNamara, 2012), reading materials (e.g. Crossley, Louwerse, McCarthy, & McNamara, 2007; Green, 2012), undergraduate reading texts (e.g. Green et al., 2010), reading texts in language tests (e.g. Green et al., 2012; Wu, 2012), and L2 test takers' scripts (e.g. Weir, 2012).

In this study, two automated textual analysis tools were used - CohMetrix Version 2.1 (Graesser, McNamara, Louwerse, & Cai, 2004) and VocabProfile version 3 (Cobb, 2003). CohMetrix was designed to explore attributes of cognitive language use. Graesser, McNamara & Kulikowich (2011) argued that CohMetrix's automated indices measure 'deep-level factors of textual coherence and processing' (223). VocabProfile (Cobb, 2003) is another popular textual analysis tool which provides a profile of texts in terms of different vocabulary frequency bands based on BNC (The British National Corpus, 2007) (e.g. the most frequent 1000 words) and different types of vocabulary (e.g. academic words based on Coxhead, 2000). Both tools have been used in the testing literature. For instance, Green et al. (2010) compared IELTS reading texts and undergraduate texts at British universities, Green (2012) investigated reading texts targeted at different levels of the Common European Framework of Reference for Languages (Council of Europe, 2001), Green et al. (2012) investigated the features of reading texts in the Cambridge CAE examination, and Wu (2012) compared Cambridge Main Suite and GEPT Taiwan examinations at the B1 and B2 levels. Weir (2012) investigated features of the test takers' scripts of the Test of English for Academic Purposes (TEAP) in Japan.

While CohMetrix and VocabProfile allow researchers to automate a large number of textual indices in an objective and reliable way, the results have to be interpreted with caution. Researchers have argued that not all indices produced are equally useful or interpretable. Green et al (2012) criticised the fact that some of the indices seem to overlap and Green (2012) attempted to identify those indices which are helpful to distinguish texts between adjacent CEFR levels. Drawing upon previous studies, especially those looking at reading texts (e.g. Green et al., 2010; Green et al., 2012 and Wu, 2012), the usefulness of the all CohMetrix and VocabProfile indices were examined in a pilot analysis. 30% of the real-life input texts were analysed in the pilot analysis. Based on the results of the pilot study, 13 CohMetrix and 4 VocabProfile indices were selected to analyse the features of the input texts and draw comparisons between the real-life and reading-into-writing test tasks (See Table 6 for a glossary of the selected indices).

The reasons for excluding the rest of the indices are summarised below:

- they overlapped with other indices – they showed very similar or no difference to results of other indices;
- they were difficult to interpret in terms of text complexity;
- they were not applicable to the data in this study due to the sampling procedures of the real-life texts; and/or
- they were not useful or effective in determining the complexity of a text.

Table 6 Glossary of the selected indices

Contextual index	Definition (Extracted from the official documents of the two tools)	Automated analysis tool
Lexical		
High frequency words (K1)	The ratio of words which appear in the first most frequent 1000 BNC (2001) wordlist to the total number of words per text	VocabProfile
High frequency words (K2)	The ratio of words which appear in the second most frequent 1000 BNC (2001) wordlist to the total number of words per text	VocabProfile
Academic words	The ratio of words which appear in the Academic Wordlist (Coxhead, 1998) to the total number of words per text	VocabProfile
Low frequency words (Offlist)	The ratio of words that do not appear in either the most frequent 15000 BNC wordlist to the total number of words per text	VocabProfile
Log frequent content words	The log frequency of all content words in the text	Cohm 46
Average syllables per word	The mean number of syllables per content word, a ratio measure	Cohm 38
Type-token ratio (content words)	The number of unique words divided by the number of tokens of these words	Cohm 44
Syntactic		
Average words per sentence	The mean number of words per sentence	Cohm 37
Sentence syntax similarity	The proportion of intersection syntactic tree nodes between all sentences	Cohm 56
Mean number of modifiers per noun-phrase	The mean number of modifiers per noun-phrase	Cohm 41
Mean number of words before the main verb	The mean number of words before the main verb of the main clause in sentences	Cohm 43
Logical operator incidence	The incidence of logical operations (i.e. connectives), such as <i>and</i> , <i>or</i> , <i>not</i> , <i>if</i> , <i>then</i> , etc	Cohm 26
Cohesion		
Adjacent overlap argument	The proportion of adjacent sentences that share one or more arguments (i.e. noun, pronoun, noun-phrase) or has a similar morphological stem as a noun	Cohm 16
Adjacent overlap stem	The proportion of adjacent sentences that share one or more word stems	Cohm 17
Adjacent overlap content word	The proportion of content words in adjacent sentences that share common content words	Cohm 58
Proportion of adjacent anaphor references	The proportion of anaphor references between adjacent sentences	Cohm 18
Adjacent semantic similarity (LSA)	The measure of conceptual similarity between adjacent sentences	Cohm 27

4.2.3 Data collection procedures

The **overall task setting** of the two real-life tasks and the GEPT Advanced Level Writing Test Task 1 were analysed by expert judgement by using Part 1 (*overall task setting*) of the

Contextual Parameter Proforma (see Table 5). The judges were trained prior to the panel meeting with the adapted Familiarisation and Specification procedures (Council of Europe, 2009). The expert judgement session involving three sets of tasks (i.e. two real-life tasks and one GEPT task) was carried out as below.

1. The researcher explained the Contextual Parameter Proforma (Part 1 - overall task setting only) to the judges. An explanation sheet of the analytical categories was provided (see Appendix 1). The judges sought clarification of any unclear points.
2. The judges were grouped into pairs.
3. The judges were assigned to analyse one of the tasks (i.e. two real-life tasks and one GEPT task) individually and filled in the Proforma. The order of the tasks assigned to each pair was counter-balanced. After they had completed the individual analyses, they discussed their responses in pairs. They were asked to record the reasons for any disagreement. They then filled in another Proforma to record their agreed responses. The judges handed in their responses (both individual and pair) to the researcher.
4. The judges analysed the other two tasks one by one following Step 3.
5. The judges completed the Feedback Evaluation Questionnaire (see Appendix 2).

The **input text analysis** involved 1) ten sample texts for each of the two real-life tasks, and 2) two passages from ten testlets of the GEPT Advanced Level Writing Test Task 1, totalling 20 real-life input texts and 20 GEPT passages in all. Results from the expert judgement session and automated textual analysis were used to analyse the input text features of the two real-life and test tasks. Following the same procedures detailed above, the two judges used Part 2 (*Input text features*) of the Contextual Parameter Proforma (see Table 5) to analyse all the sampled input texts from the two real-life input tasks and the GEPT task. The input texts were also analysed by automated textual tools. 60 extracts from the 20 real-life input texts and 20 passages from 10 testlets of GEPT Advanced Level Writing Test Task 1 were analysed using the 13 CohMetrix and 4 VocabProfile indices (see Table 6).

Overall, the judges reported that they were confident when responding to the items of the Proforma on a scale of 4 (4=very confident; 3=confident; 2=not confident; 1=not confident at all). All categories, apart from *topic domain*, had a mean above 3 (for details see Appendix 2). The judges commented that it was not straightforward to determine the topic domain of a reading-into-writing task. The decision might be influenced by a range of task features, such as the context described in the prompt, the suggested title of the output text, the common theme of the input texts, and the sources of the input texts. Future research is advised to provide sample tasks in each topic domain to help with the evaluation process.

4.2.4 Method of data analysis

The purpose of the contextual investigation in this study was to examine the extent to which the set of context validity parameters were represented in GEPT Advanced Level Writing Task 1 so as to generate context validity evidence for the task.

The expert judgement responses on the 14 categories regarding the overall task setting and input text features were reported. For the classification categories (i.e. Items 3-5, 8-11), results of the percentage of each option were presented. For the rating categories (i.e. Items 1-2, 6-7, 12-14), the mean and standard deviation on the five-point Likert scale were presented. Descriptive comparisons were made between the real-life tasks and the GEPT task.

Descriptive statistics instead of inferential statistics were used due to a small sample size. Graphic presentation of the data was provided for further illustration where necessary.

Regarding the automated textual analyses of the input text features, the mean and standard deviation of the 17 selected indices were obtained. Results for the GEPT Advanced Level Writing Test Task 1 were compared with the real-life textual indices. As the comparisons involved non-normally distributed data, Mann-Whitney tests were performed, where appropriate, for inferential statistics analyses between the conditions.

4.3 Research design: cognitive validity of the GEPT Advanced Writing Test Task 1

4.3.1 Participants

Phase 1: The cognitive validity of the GEPT Advanced Writing Test Task 1 was investigated in the context of a UK university. 160 Chinese university students (51.9% males and 48.1% females) studying on a full-time collaborative undergraduate programme at the Business School, the University of Bedfordshire, were recruited. They were pursuing one of four majors: Business Administration (25.6%), Advertising and Marketing Communications (51.9%), Human Resource Management (22.5%), Marketing and Accounting (30%). Their English proficiency was at the estimated level between CEFR B2 and C1 (see Table 7 for the details of their IELTS scores). All IELTS scores were effective at the time of the study, i.e. within 2 years of effectiveness, in line with the University's admission policy.

Table 7 Participants' IELTS scores

	IELTS Reading	IELTS Writing
Mean	5.93	5.53
Std. Deviation	0.59	0.51
Minimum	5.0	5.0
Maximum	7.5	7.0

Phase 2: The cognitive validity of the GEPT Advanced Writing Test Task 1 was investigated under live test conditions in Taiwan. 192 Taiwanese test takers were recruited by LTTC.

4.3.2 Writing Process Questionnaire

A Writing Process Questionnaire (adapted from Chan, 2013) was used in the study to assist the participants to self-report the processes that they employed on a) the two selected real-life writing tasks, and b) the GEPT Advanced Level Writing Test Task 1. Chan's questionnaire was developed based on Kellogg's (1996) and Field's (2004, 2008, 2011, 2013) models of different phases of receptive and productive skills and the work of Shaw & Weir (2007) on the cognitive validity parameters in independent writing tests, the work of Khalifa & Weir (2009) on the cognitive validity parameters in independent reading tests and Spivey's (1984, 1990, 1997, 2001) model of discourse synthesis. The questionnaire of 48 items (see Appendix 3) was designed to measure the processes involved across the five hypothesised academic writing cognitive phases: (1) conceptualisation, (2) meaning and discourse construction, (3) organising, (4) low-level monitoring and revising, and (5) high-level monitoring and revising

(see Table 2 for the definition of each cognitive phase and Table 8 for the structure of the questionnaire).

Table 8 Structure of the questionnaire

Phases of academic writing	Items ³	Reliability	No. of items
Conceptualisation	1.2, 1.3, 1.4, 1.5, 2.6, 2.13, 4.4, 4.6	0.522	8
Meaning and discourse construction	1.1, 2.1, 2.2, 2.4, 2.5, 2.7, 2.9, 2.12, 4.2, 4.3, 4.5	0.683	11
Organising	2.3, 2.8, 2.10, 2.11, 3.1, 3.2, 3.3, 3.4, 4.1	0.630	9
Low-level monitoring and revising	4.12, 4.13, 4.14, 4.16, 5.12, 5.13, 5.15, 5.16	0.841	8
High-level monitoring and revising	4.7, 4.10, 4.11, 4.8, 4.9, 4.14, 5.7, 5.10, 5.11, 5.8, 5.9, 5.14	0.738	12
Total			48

4.3.3 Data collection procedures

Phase 1

The data concerning the UK university students' processes on GEPT Advanced Level Writing Test Task 1 was collected at the beginning of the term. The test task was administered to the participants under strict test conditions during their language classes, following the instructions provided by the test providers, i.e. LTTC. Immediately after the participants had completed the test task, the questionnaire was used to prompt the participants to self-report the extent to which they employed different cognitive processes when completing the task. The data concerning the university students' processes on the real-life tasks, on the other hand, was collected during the term. The questionnaire was administered online through an online survey tool called Survey Monkey one week before the submission week of each of the real-life tasks (in middle of the term for real-life task A and at the end of the term for real-life task B). The 160 participants in the main study, like other students in the Business School, were allowed to choose two to four selective modules, depending on the structure of their programme. As a result, a total of 160 questionnaires were collected on the GEPT task and 143 on the real-life academic tasks (70 on real-life task A and 73 on real-life task B).

Phase 2

The data concerning the Taiwanese test takers' processes on the GEPT Advanced Level Writing Test Task 1 was collected by LTTC under live test conditions, following normal standardised test administration procedures. A total of 192 questionnaires were collected.

³ All items have an item-total correlation to its corresponding phase at a level of 0.3 or above.

4.3.4 Method of data analysis

The purpose of the cognitive investigation in this study was to examine the extent to which the processes elicited by GEPT Advanced Level Writing Test Task 1 resemble the processes which students would normally employ on real-life academic writing tasks. Additionally, a follow-up Phase 2 study was conducted to confirm whether similar evidence of the cognitive validity of GEPT Advanced Level Writing Test Task 1, as collected under the research setting in the UK, could be obtained under live test conditions in Taiwan.

The 303 questionnaires collected in Phase 1 and 192 collected in Phase 2 were computed for statistical analysis. Descriptive statistics of individual questionnaire items from each of the real-life tasks were obtained.

Phase 1 data

The Mann-Whitney U test, which is a non-parametric independent-sample test, was performed on individual questionnaire items to compare the results of the two real-life tasks. Based on the fact that the means of the majority of the questionnaire items showed no significant difference between the two tasks, the data sets collected from the two real-life tasks were analysed together in the subsequent analyses.

The real-life data was then submitted to exploratory factor analysis (EFA) to investigate the underlying structure of the cognitive processes involved in each of the five academic writing phases⁴ elicited by the real-life tasks. Additionally, another set of Mann-Whitney U tests was conducted to examine if these EFA-generated cognitive parameters could reflect a difference in how high-achieving and low-achieving participants approached the tasks.

After that, a set of Wilcoxon signed ranks tests, which are non-parametric related-sample tests, were performed on the processing data from the GEPT and real-life tasks to examine the extent to which the processes elicited by the GEPT Advanced Level Writing Test Task 1 are comparable to those elicited by the real-life tasks. The analyses were conducted to compare the processes employed by the higher-achieving participants, low-achieving participants and the whole population of the participants under test and real-life conditions.

Exploratory factor analysis (EFA) was then performed on the GEPT data to investigate the underlying structure of the cognitive processes involved in each of the five cognitive phases. The analyses were to reveal the extent to which the underlying structure of the processes elicited by the GEPT Advanced Level Writing Test Task 1 resemble the target real-life cognitive constructs.

Phase 2 data

The data on the GEPT task collected under live test conditions was compared to the data on the GEPT task collected under the research setting in the UK using Mann-Whitney U tests. Exploratory factor analysis (EFA) was also performed on the GEPT (live test) data to investigate the underlying structure of the cognitive processes involved in each of the five cognitive phases. The analyses were to confirm that similar evidence of the cognitive validity

⁴ As explained earlier, this study builds on Field's models of receptive and productive language skills, and considers five of the cognitive phases to be most relevant to the discussion of the cognitive validity of academic writing. The exploratory factor analysis conducted was not to build an overall model of the cognitive phases involved in academic writing, but to examine the number of distinctive cognitive processes and the underlying structure of these cognitive processes involved in these five phases of academic writing.

of GEPT Advanced Level Writing Test Task 1 can be obtained under live test conditions as reported under the research setting in the UK.

5. Results

5.1 The contextual validity of GEPT Advanced Writing Task 1

5.1.1 Results from expert judgement – overall task setting

Based on the sixteen judges' responses in pairs to the Contextual Parameter Proforma Items 1- 7, the overall task setting of the two real-life tasks and GEPT Advanced Writing Task 1 are presented and discussed below.

Genre

Regarding the genre of the output text of the two real-life tasks, the judges' responses showed total agreement on their responses. For real-life task A, students were expected to produce an essay whereas for real-life task B, students were expected to produce a report (real-life task A hereafter will be called the essay task, real-life task B the report task). Regarding the genre of the GEPT task, seven pairs of the judges considered the genre to be essay. However, one pair decided it was 'essay and summary'. Pair 1 argued that 'essay can be anything'. In their opinion, although the output is 'an essay', the task actually required test takers to write a summary in more specific terms.

Purpose of the task

Regarding the effectiveness of the communicative purpose set in the task, i.e. 'a reason for completing the task that goes beyond a ritual display of knowledge for assessment' (Shaw & Weir, 2007: 71), the judges' responses fell toward the positive end of a five-point Likert scale for all three tasks but the GEPT task seemed to have presented a slightly clearer purpose than the real-life tasks did (see Figure 1). Pair 5 commented that the communicative purpose presented in the two test tasks were straightforward and unambiguous. The purpose of the real-life essay task was perceived to be the least clear among the three, though the rating was still above four out of five. Pair 2 commented that apart from following the instructions, the 'real' communicative purpose to achieve on the real-life essay task was not very clear. Nevertheless, although the communicative purpose presented on the two real-life tasks seemed to be slightly less transparent than the test tasks, students may well receive further explanations from the lecturer. As the test takers would not receive any verbal explanation of the test task under test conditions, it is essential for the test tasks to present a clear communicative purpose. The GEPT task sampled in this study did very well in this regard.

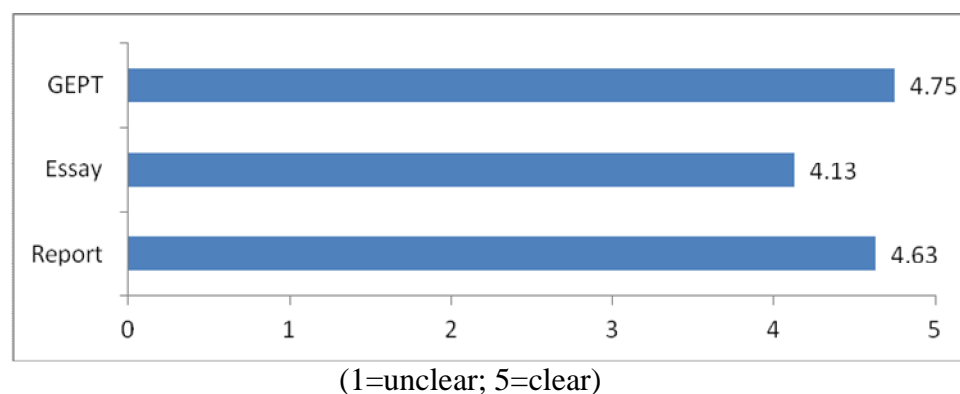


Figure 1 Clarity of the purpose of the tasks

Topic domain

The judges were asked to rate the extent to which each task falls into each of the four topic domains, i.e. professional, academic, social and personal. The results show that all tasks fell into more than one topic domain (see Figure 2)

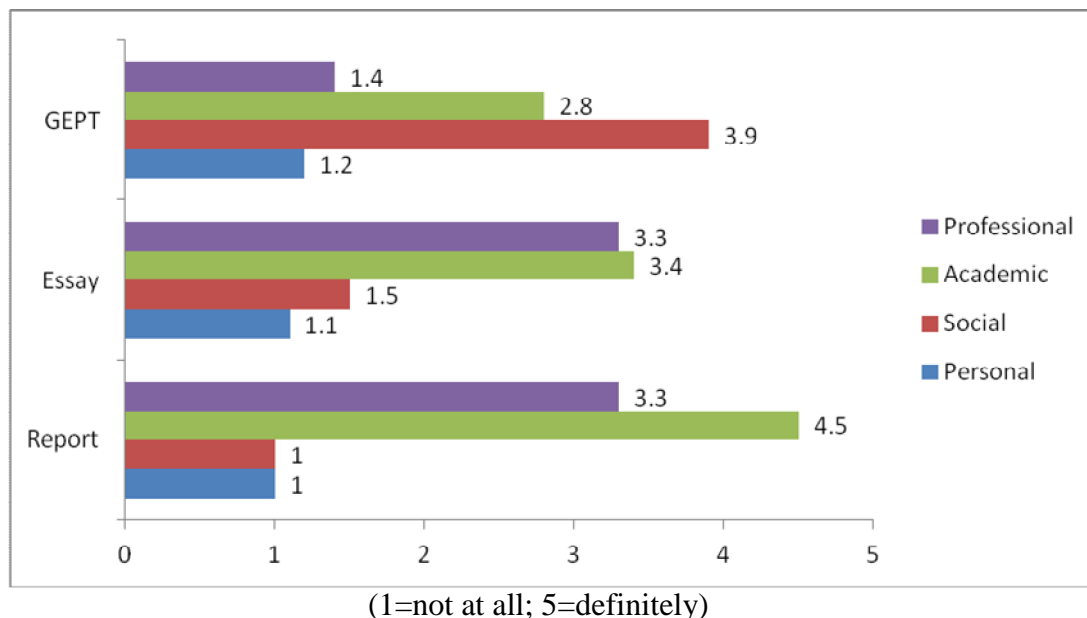


Figure 2 Topic domains of the tasks

Based on the judges' response, the topic domain of the essay task was largely *professional*, i.e. Business in the context of the study, and *academic*. The topic domain of the report task was regarded as primarily *academic*, followed by *professional*. Agreeing with the literature (Khalifa & Weir, 2009), the *personal* and *social* domains did not play an important role in the academic writing context. Both the real-life tasks were predominantly in the *professional* and *academic* domains. The GEPT task was in the *social* and, to a lesser extent, *academic* domains. The social domain refers to the contexts connected with general social interaction in a public domain, one usually adopted in language tests of general proficiency. The topic of the GEPT task was about whether it is worth saving endangered languages. While GEPT is a test of general proficiency across different levels, GEPT Advanced also serves as a means of measuring the English language ability of Taiwanese applicants who wish to pursue further studies overseas (LTTC, 2013). The social domain does not seem to be entirely appropriate from this perspective. It is, however, understandable why the test task did not contain very specific content. One advantage of reading-into-writing tasks is that they can provide an equal access of background knowledge of the topic to prevent bias against test takers. If the content provided in the input texts is too specific, it may impose a background effect on test takers (Ackerman, 1990; Clapham, 1996; Kellogg, 1987). Unlike ESP tests, writing tasks in EAP language tests should not require a high level of specific knowledge (Douglas, 2000).

Cognitive demands

The cognitive demand of a reading-into-writing task depends largely on the expected 'scope' of the interaction between input and output (Douglas, 2000: 65). Building upon the literature review (Bereiter & Scardamalia, 1987; Galbraith & Torrance, 1999; Purves et al., 1984), the level of cognitive demand of a writing task can be broadly divided into three levels:

1. Telling/retelling content

2. Organising/reorganising content
3. Transforming content

The judges were asked to determine the level of the cognitive demands the tasks impose on writers by considering the nature of the reading processes required and in what way writers should draw upon the input texts. The results are shown in Figure 3.

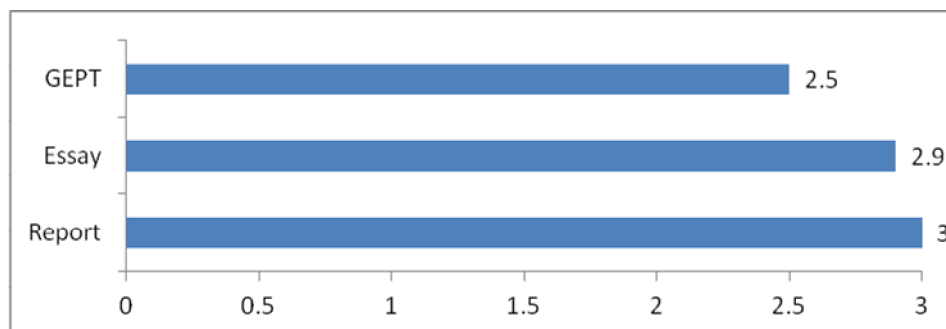


Figure 3 Cognitive demands of the tasks

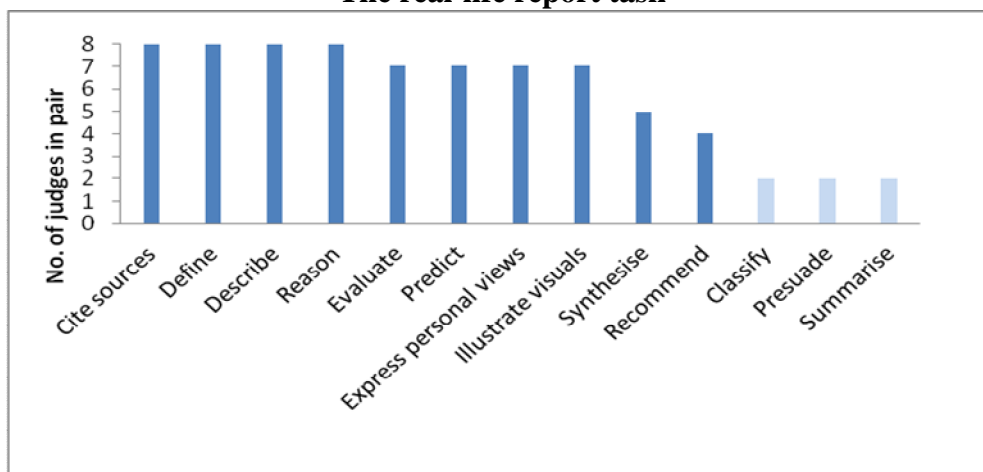
As shown in Figure 3, both the two real-life tasks and the GEPT task were mapped towards the knowledge-transforming end of the cognitive demand's continuum (Scardamalia & Bereiter, 1987). The real-life report task scored an average of 3 out of a scale of 1 (telling/retelling content), 2 (organising/reorganising content) and 3 (transforming content), the real-life essay task an average of 2.9. On the other hand, the GEPT task scored an average of 2.5. The findings suggest that the real-life academic writing tasks were knowledge transforming tasks. In order to complete a knowledge transforming task, writers are usually expected to employ high-level processes, such as planning rhetorical goals, integrating ideas from different sources and transforming ideas (the actual processes elicited by the tasks are discussed in Section 4.2). The GEPT task was apparently slightly easier than the real-life tasks in terms of the cognitive demands. It required the test takers to transform the ideas by selecting, organising and summarising relevant ideas from the input sources as well as evaluating different points of view. However, the GEPT task sampled in this study did not seem to require test takers to interpret, evaluate, and apply ideas in context to the extent that the real-life tasks did. Perhaps this is not surprising given the constraints on an exam essay as compared to the wider possibilities of transformational activities in university writing tasks.

Language functions to be performed

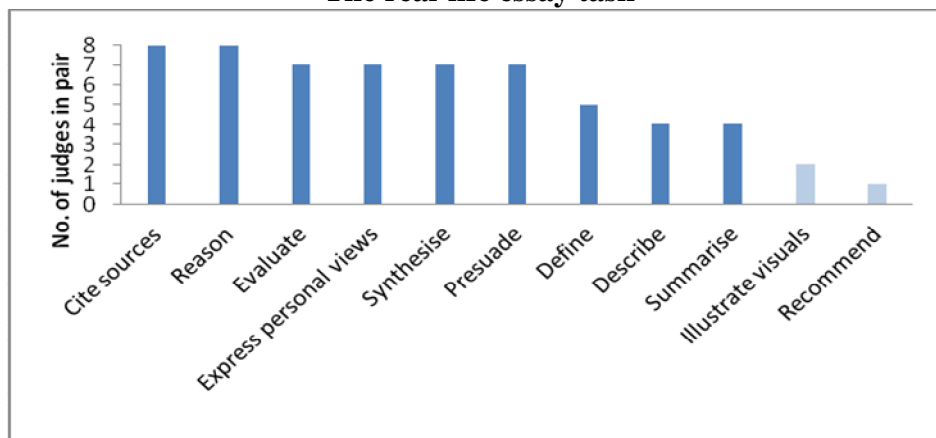
The judges were asked to analyse the language functions that the writers are expected to perform on the tasks. The findings show that the real-life tasks required students to perform more language functions than the GEPT task. The most important functions on the report task (i.e. those determined by four or more pairs of the judges) included *describe*, *define*, *reason*, *cite sources*, *evaluate*, *predict*, *express personal views* and *illustrate visuals*, followed by *synthesise* and *recommend*. Two pairs of the judges identified *classify*, *persuade* and *summarise*. The most important language functions to be performed on the essay task were *reason*, *evaluate*, *persuade*, *cite sources*, *synthesise* and *express personal view*, followed by *define*, *describe* and *summarise*. Two pairs of the judges identified *illustrate visuals*, one pair identified *recommend*. The GEPT task apparently required fewer language functions. The most important functions on the task were *cite sources*, *summarise*, *express personal view*, *evaluate* and *reason*. However, two pairs of judges identified *describe*, *persuade*, *recommend*, and *synthesise*, one pair identified *define* and *predict* (Further language functions, such as *describe*, are targeted by Task 2 of the GEPT Advanced Writing Test).

The findings reveal that some language functions, e.g. *cite sources*, *describe*, *reason*, *evaluate*, *express personal views*, and *synthesis*, were important for both real-life tasks. This indicates the importance of these 'core' language functions to be tested in writing tests for academic purposes. It is worth noting that these core functions were also considered to be those expected in the GEPT task (although only two of the pairs of judges identified *synthesise*, eight pairs identified *summarise*). Further evidence will be useful to check whether the expected functions are actually carried out by students (see Weir & Wu 2006).

The real-life report task



The real-life essay task



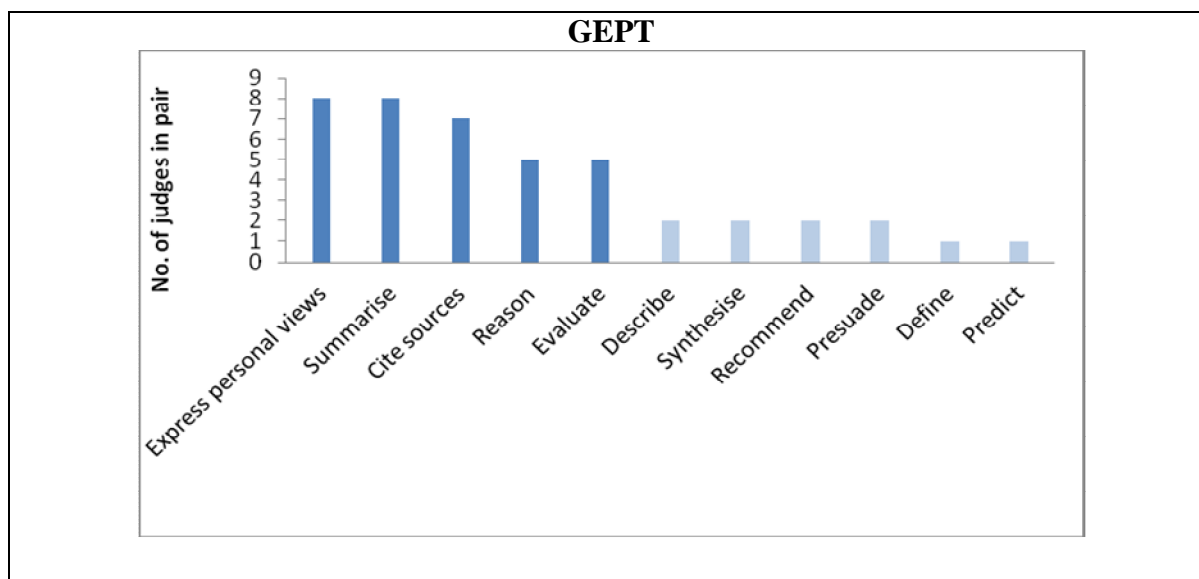


Figure 4: Language functions expected to perform on the tasks

Clarity of writer-reader relationship

With respect to the clarity of audience presented, the judges considered that there was room for improvement for both real-life tasks and the GEPT task (see Figure 5). Pair 5 explained that while it might be obvious to the students that the 'real' intended reader of the real-life tasks were the lecturers, both tasks did not provide specific information about the intended reader. Mature writers would consider the needs of the reader while they plan, write and edit their text (Scardamalia & Bereiter, 1987). A valid writing task should always clearly present the intended reader, e.g. self, well known other, distant other (Shaw & Weir, 2007). The judges gave an average of 3.13 out of 5 for the GEPT task (see Figure 5). However, there was some obvious disagreement in the focus group meeting (the rating ranged from 2 to 5). The context of the task was a national essay contest. While some judges considered the implied reader was clear, i.e. the judges of the writing contest, while other judges thought that no actual information was provided regarding the intended reader. It was unclear to them if the reader would be a single judge, a group of judges, or even a bigger community which could get access to the writing contest.

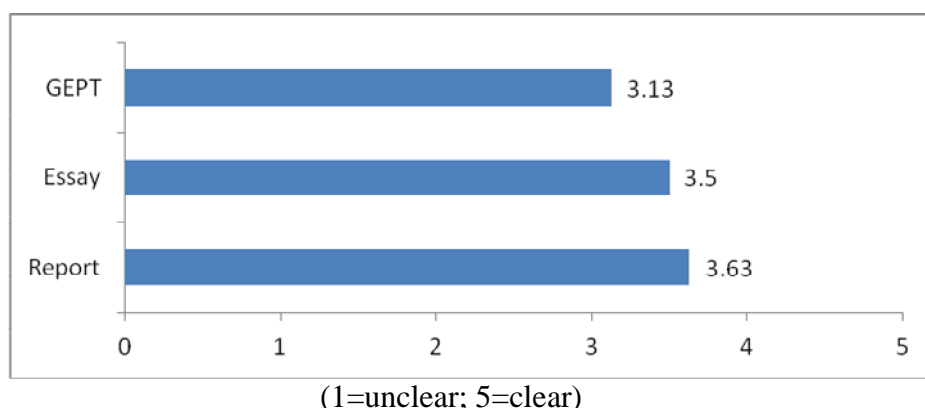


Figure 5 Clarity of audience of the tasks

Knowledge of criteria

The overall ratings concerning the provision of the knowledge about marking criteria on the tasks are presented in Figure 6. The judges felt that both real-life tasks provided students with very clear and detailed marking criteria. On the other hand, the judges gave an average rating

of 3.75 out of 5 for the GEPT task. The task stated that test taker's performance would be scored according to four criteria, namely, relevance and adequacy, coherence and organisation, lexical use and grammatical use. However, most judges considered some more specific descriptions of the criteria would be helpful. In real-life context, test takers of GEPT Advanced Test are provided with the detailed marking criteria before they take the test. The detailed marking criteria are published in the GEPT Handbook for Examinees and the GEPT Advanced Level Past Paper, as well as released on the GEPT website (LTTC, 2013).

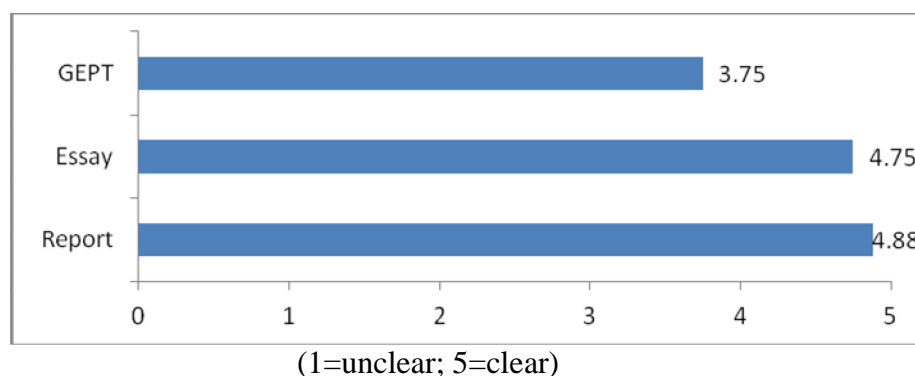


Figure 6 Provision of the knowledge of criteria

We have so far discussed the overall task setting of the GEPT and real-life tasks. The next section will look at the results regarding input text features analysed by expert judgement and automated textual analysis tools.

5.1.2 Expert judgement - input text features

As mentioned previously, 20 real-life texts and 20 GEPT input texts were sampled in the study. Two judges individually analysed all the input texts in terms of *input format*, *verbal input genre*, *non-verbal input*, *discourse mode*, *concreteness*, *textual organisation* and *cultural specificity* (i.e. item 8-14 of the Contextual Parameter Proforma).

For the classification criteria, the judges showed total agreement on *the input format*, *verbal input genre*, and *non-verbal verbal input type*. Their agreement rate on the *discourse mode* was 93%. The most divergent responses as regards discourse mode were obtained in the choice of options *expository* and *argumentative*. The individual judges explained that some real-life input texts seemed to serve both discourse modes. They were asked to identify together the primary discourse mode in those texts.

For the three criteria, i.e. *concreteness*, *textual organisation* and *cultural specificity*, the majority of the responses between the two judges were either with exact agreement (62%) or within one scale point (35%). About 3% of the responses between the judges were with a difference of two scale points. Their responses, if different, were averaged.

Input format and verbal input genre

The input format of the GEPT task is standardised. It contains two reading passages. (Non-verbal inputs are used in the GEPT Advanced Writing Task 2: Summarising main ideas from non-verbal input and providing solutions). For the two real-life tasks, verbal input was more dominant in the essay task while verbal and non-verbal inputs were more important in the report task. 80% of the essay input texts was solely verbal and 20% contained both verbal and

non-verbal information. In contrast, 30% of the report input texts was verbal and 70% was verbal and non-verbal. None of the input texts contained solely non-verbal input.

Regarding the distribution of the verbal input genres, the real-life tasks consisted of a wider range of different genres than the test tasks. News / magazine articles (50%) and journal articles (30%) were the most frequently occurring genres read by the participants on the real-life essay task while book chapters (60%) were the dominant genre for the real-life report task. Each GEPT task contained 2 input texts. All input texts from the 10 testlets of GEPT collected in this study were more difficult to associate with these genres and seemed to belong to a rather non-specific text created specifically for the exam, perhaps a simplified version of the essay genre.

Discourse mode

The discourse mode of the input texts would have a direct impact on the task difficulty. Brewer (1980: 225) argued that processing descriptive texts would require the reader to build a visual and spatial cognitive structure; narrative texts would require creating a mental representation of a series of occurring events; and expository texts would require the cognitive processes of constructing induction, classification and comparison. With respect to the primary discourse mode of the input texts, both real-life texts contained expository and argumentative texts. The report task contained more expository texts (70%) while the essay task contained more argumentative texts (60%). No input texts on both real-life tasks were considered as narrative or descriptive. In contrast, the input texts on the test tasks were dominated by single discourse mode. All texts collected from the GEPT task were identified as argumentative texts (100%).

Concreteness of the ideas, explicitness of the textual organisation and cultural specificity

With respect to the concreteness of the ideas in the input texts, the ideas in the GEPT input texts were considered more concrete than the ideas in the real-life input texts (see Figure 7). Topic domains may partially explain why the input texts of the GEPT task contained more concrete ideas than the real-life input texts. As discussed earlier, the real-life tasks were in the topic domains of *academic* and *professional*, whereas the GEPT task in this study was in the *social* and, to a less extent, *academic* domains. For example, the input texts of one tasklet of the real-life essay task were related to the phenomenon of the feminization in the public relations (PR) industry, another tasklet related to the advertisement strategies of John Lewis to focus on core family values. The input texts of one tasklet of the real-life report task were related to business-specific knowledge such as different techniques to predict the uncertain nature of business trends and graphics useful for modeling and forecasting time series. On the other hand, the input texts of the two test tasks are much less knowledge-specific. For example, one testlet of the GEPT task was about the reasons why saving the disappearing languages is important (e.g. every language has unique characteristics). However, there may be other factors contributing to the concreteness of ideas. For example, input texts are often closely edited as part of the test development procedure. As a result, some abstract ideas might be removed or made more concrete in the interests of clarity or accessibility. In comparison, the input texts of the real-life tasks tend to be unaltered extracts from primary sources, for instance, book chapters and articles.

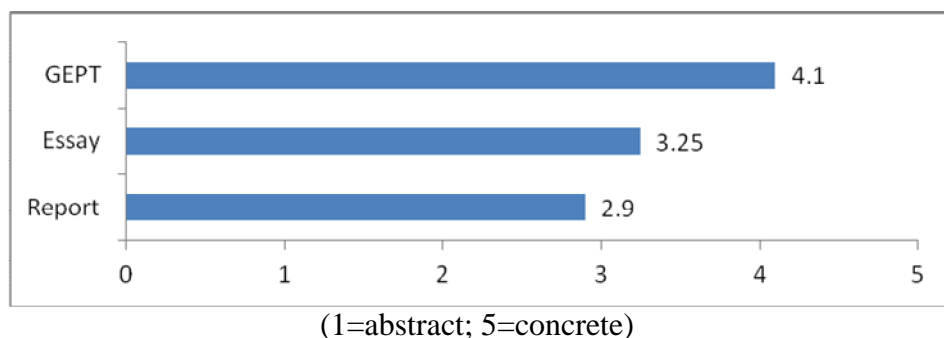


Figure 7 Concreteness of the input texts

Regarding the explicitness of the textual organisation of the input texts, the GEPT input texts were more explicitly organised than the real-life input texts. The judges felt that most of the GEPT input texts sampled in this study were organised mechanically into 3 to 5 paragraphs, each containing a main idea. And these paragraphs were sometimes too explicitly linked with the use of formulaic markers such as 'firstly', 'in addition' and 'lastly'. On the other hand, formulaic markers were found to be less frequent in real-life input texts. There was a higher demand for the students to figure out how each paragraph relates to each other, and to the whole text (i.e. the process of discourse construction).

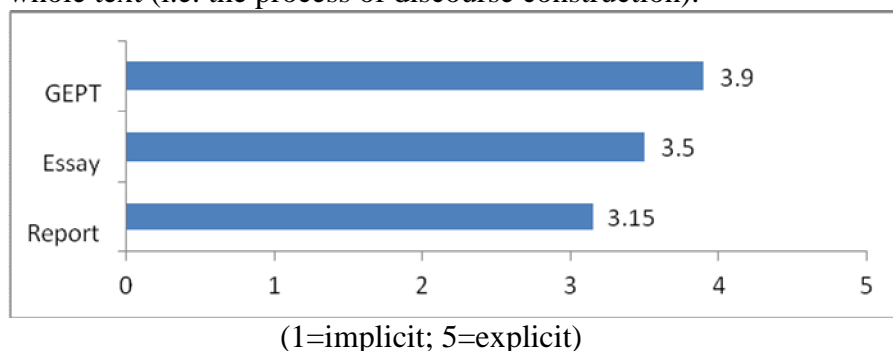


Figure 8 Explicitness of the textual organisation

Regarding the cultural specificity of the input texts, the GEPT input texts were considered more culturally specific than the real-life texts. Many of the GEPT input texts referred specifically to the Taiwanese context.

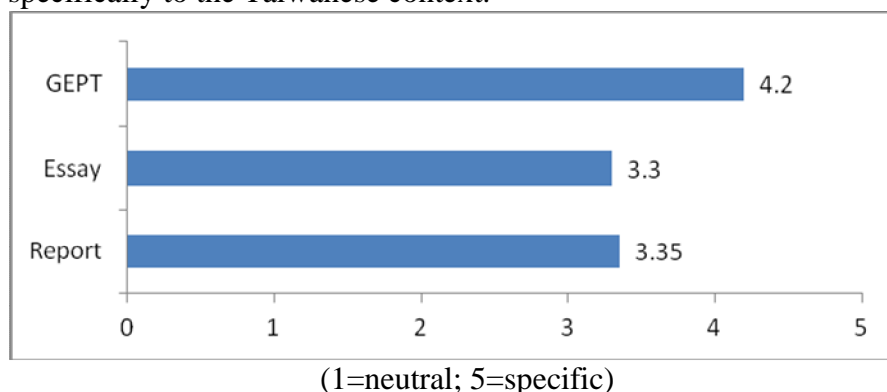


Figure 9 Degree of cultural specificity

5.1.3 Automated textual analysis – the linguistic complexity of the input texts

Regarding the automatic textual analyses, 60 extracts from the 20 real-life input texts, and 20 passages from 10 testlets of GEPT Advanced Writing Task 1 were analysed in terms of the lexical complexity, syntactic complexity and cohesion of the text. The three aspects of text were analysed by 17 selected indices generated by two automated tools: Coh-Metrix and VocabProfile (see Table 6). The differences between the real-life and GEPT Advanced Writing Task 1 input texts were analysed by Mann-Witney test (see Tables 9-11). Generally speaking, the difficulty level of sampled GEPT Advanced Writing Task 1 input texts was very comparable to the level of the real-life input texts. There was no significant difference in 14 out of the 17 indices obtained between the two conditions and with the exception of *Low frequency words (Offlist)* these differences were slight.

Lexical complexity

Lexical complexity has long been used to determine the difficulty level of reading texts in the second language learning context (Green, 2010). The lexical complexity of the input texts was analysed by 7 indices, namely, *high frequency words (K1)*, *High frequency words (K1+K2)*, *academic words*, *low frequency words (offlist)*, *log frequent content words*, *average syllables per word* and *type-token ratio (content words)* (see the results in Table 9). The first four lexical indices measure the frequency of all words in the texts, including the first 1000 and 2000 most frequent words in the BNC corpus, academic words and low frequency words (i.e. those are not included in frequency list of 15000 on the BNC). Whereas GEPT input texts contained similar proportions of high frequency words (K1 and K1+K2) as the real-life input texts, they contained slightly fewer academic words on average though the difference was non significant but interestingly more low frequency words. The difference in the proportion of low frequency words was significant ($z=-6.545$, $p<0.01$) and the mean difference was as large as 8% (real-life: 2.41%, GEPT: 10.41%). The offlist words on the GEPT inputs texts were mainly proper names of places and organisations/companies. The next three lexical indices concern the content words in the real-life source text. Content words, i.e. nouns, main verbs, adjectives and adverbs, are those containing conceptual meaning. The first index (frequency level) showed the frequency level of the content words in the real-life texts. It computes the log frequency of all content words in the text, ranging from zero to six. The lower the score is, the less frequent the content word is. The next index (average syllables) measures the average syllables per content word in the real-life input texts. The last lexical index (type-token ratio of content words) measures the type-token ratio of all content words in the input texts. The ratio reflects the proportion of unique content words which need to be decoded. There was not much difference in terms of the frequency of content words and the average syllables per word between the GEPT and real-life input texts. GEPT input texts had a slightly higher type-token ratio but the difference was not significant.

Table 9 Comparison of the lexical complexity of the real-life and GEPT input texts

	Real-life tasks (60 extracts from 20 texts) Mean	GEPT (20 texts) Mean	Mann- Whitney U	Wilcoxon W	z	Asymp. Sig. (2- tailed)
Lexical features						
High frequency words (K1)	77.20	76.54	556.000	766.000	-.489	0.63
High frequency words (K1+K2)	87.76	87.69	600.000	810.000	.000	1.00
Academic words	10.37	8.84	437.000	647.000	-1.811	0.07
Low frequency words (Offlist)	2.41	10.41	11.000	1841.000	-6.545	0.00
Log frequent content words	2.10	2.05	508.000	718.000	-1.022	0.31
Average syllables per word	1.70	1.72	591.000	2421.000	-.100	0.92
Type-token ratio (content words)	0.69	0.72	448.000	2278.000	-1.689	0.09
Adjacent semantic similarity (LSA)	0.23	0.25	488.000	2318.000	-1.245	0.21

Syntactic complexity

Syntactic complexity is believed to be an important indicator of the difficulty of a text. Researchers, for instance Crossley, Greenfield & McNamara (2008), have demonstrated that the more complex sentence structures a text contains, the more difficult it is for readers to process the text. Syntactic complexity is particularly important in determining the difficulty of a reading-into-writing task where higher-level reading processing (such as creating textual and intertextual representations) rather than low-level lexical decoding is targeted. The syntactic complexity of the real-life and GEPT input texts was analysed by five indices in this study. Table 10 below summarises the results.

Table 10 Comparison of the Syntactic complexity of the real-life and GEPT input texts

	Real-life tasks (60 extracts from 20 texts) Mean	GEPT (20 texts) Mean	Mann- Whitney U	Wilcoxon W	z	Asymp. Sig. (2- tailed)
Syntactic features						
Average words per sentence	21.38	20.49	514.000	724.000	-.956	0.34
Sentence syntax similarity	0.08	0.09	401.500	2231.500	-2.206	0.03
Mean number of modifiers per noun-phrase	1.03	0.91	336.000	546.000	-2.933	0.00
Mean number of words before the main verb	5.50	5.76	493.500	2323.500	-1.183	0.24
Logical operator incidence	45.12	43.76	560.500	770.500	-.439	0.66

The first syntactic index (average words per sentence) measures the average number of words per sentence in the real-life input texts. Generally speaking, the longer a sentence is, the more complex it is because it might contain more phrases and clauses. A text with many complex sentences is demanding to process because the reader needs to build many elaborate syntactic structures. The second index (syntax similarity index) measures how syntactically similar the sentences in the real-life input texts are, by calculating the proportion of nodes in the two syntactic tree structures that are intersecting nodes between all sentences and across paragraphs. It is easier to process a text with more syntactically similar sentences than with more syntactically different sentences due to a syntactic parsing effect. The next two syntactic indices concern the noun phrases and main clauses of the real-life input texts. They measure the mean number of modifiers per noun phrase and the mean number of words before the main verbs of the main clauses respectively. Noun phrases and main verbs in a text are believed to carry the key meaning of a text. Modifiers, e.g. adjectives, adverbs, or determiners, are used to describe the property of the head of a noun phrase or the main verb.

The last syntactic index deals with the density of logical operators (connectives) in the real-life texts. Logical operators can be used to explicitly express the relations among the ideas in a text. According to CohMetrix, texts with a high density of these logical operators are difficult. This is perhaps true for low proficiency readers who have problems processing complex sentences. Otherwise, many researchers, e.g. Brown & Yule (1983) and Green et al (2012), argue that the lack of connectives actually increase the difficulty of a text because the reader has to build the relationships between the ideas. The study considers the second viewpoint as appropriate for the context of this study.

As shown in Table 10, there was no significant difference in three syntactic indices (i.e. *average words per sentence*, *mean number of words before the main verb* and *sentence syntax similarity*) between the GEPT and real-life input texts. However, GEPT input texts had a significantly higher sentence syntax similarity index than the real-life input texts, and contained significantly fewer modifiers per noun-phrase than the real-life input texts. This suggests that the GEPT texts might be slightly easier than the real-life input texts in terms of syntactic complexity, although the actual mean differences were very small.

Degree of cohesion

Measurement of cohesion is used less frequently to determine the difficult level of a reading text than the measurements of lexical and syntactic complexity. However, the degree of cohesion of the input texts is particularly relevant to the discussion of the reading-into-writing tasks in the study. The more cohesive a text is, the easier it would be for the reader to build the textual representation because a cohesive text contains 'explicit features, words, phrases or sentences that guide the reader in interpreting the substantive ideas in the text, in connecting ideas with other ideas and in connecting ideas to higher level global units, e.g. topics and themes' (Graesser et al., 2004: 193). It is certainly easier for the reader to create textual representation of a more cohesive text. However, one reason why the measurement of cohesion is less popular in determining the text difficulty is because the cohesion of a text may not be reflected directly by the occurrence of cohesive devices. Kennedy & Thorp (2007) argued that, especially concerning a more advanced level, an overt occurrence of cohesive devices does not necessarily improve the cohesion of a text. Therefore, the results have to be interpreted with caution. The cohesion of the real-life input texts were analysed by five indices in this study. Table 11 summarises the results.

Table 11 Comparison of the degree of cohesion of the real-life and GEPT input texts

	Real-life tasks (60 extracts from 20 texts) Mean	GEPT (20 texts) Mean	Mann- Whitney U	Wilcoxon W	z	Asymp. Sig. (2- tailed)
Cohesion						
Adjacent overlap argument	0.55	0.60	520.500	2350.500	-.884	0.38
Adjacent overlap stem	0.58	0.65	500.500	2330.500	-1.106	0.27
Adjacent overlap content word	0.10	0.09	490.500	700.500	-1.217	0.22
Proportion of adjacent anaphor references	0.25	0.28	457.000	2287.000	-1.589	0.11
Adjacent semantic similarity (LSA)	0.23	0.25	488.000	2318.000	-1.245	0.21

The first three indices (adjacent overlap argument, adjacent overlap stem and adjacent overlap content word) measure the proportion of adjacent sentences sharing one or more arguments (i.e. nouns, pronouns, noun-phrases), stems and content words respectively. The occurrence of repeated arguments, stems or content words would make the text more

cohesive and hence easier to be processed. These previously-occurring ideas would ease the demand on the reader to process new ideas. The next index (anaphor reference adjacent) measures the proportion of anaphor references between adjacent sentences. It is easier for the reader to resolve the anaphor reference when the referent is in an adjacent sentence, rather than at a distance of a few sentences. The last selected index (adjacent semantic similarity) quantifies how conceptually similar each sentence is to the next sentence by comparing the Latent Semantic Analysis (LSA) dimensions of their lexical items. The higher the score was, the more conceptually similar the adjacent sentences are with each other. A high proportion of adjacent sentences with conceptually-related words can help the reader to build the textual representation, e.g. the themes of the text. The degree of cohesion in GEPT and the real-life input texts was similar. There was no significant difference in the cohesion indices obtained between the GEPT and real-life input texts.

In short, regarding the comparisons of the linguistic complexity of the GEPT and real-life input texts, the results showed more similarities than discrepancies. Out of the 17 indices, only 3 indices obtained significant differences between the GEPT and real-life input texts. The GEPT input texts had a significantly greater density of *low frequency words*, mostly proper nouns, than the real-life input texts. This would probably increase the difficulty of processing the texts if the test takers were not familiar with these proper nouns. On the other hand, the GEPT input texts had a significantly higher *sentence syntax similarity index* than the real-life input texts, and contained significantly fewer *modifiers per noun-phrase* than the real-life input texts. Apart from these three indices, the linguistic complexity of the GEPT and real-life input texts was comparable.

5.2 The cognitive validity of GEPT Advanced Writing Task 1

5.2.1 Defining the target cognitive parameters in the real-life context

The previous sub-section presented and discussed the results of the context validity of GEPT Advanced Writing Task 1 used to assess academic writing ability. This sub-section presents and discusses the results of its cognitive validity which is concerned with the extent to which a test elicits from test takers cognitive processes that correspond to the processes that they have to use in the target language context (Glaser, 1991; Shaw & Weir, 2007). There are two major steps involved in investigating cognitive validity. First we need to define the target cognitive processes to be measured in a writing test by investigating the processes that students employ in a real-life context. Secondly we need to investigate the extent to which these target cognitive processes can be elicited by the test tasks.

For Phase 1 of the study, a total of 303 questionnaires were collected from the real-life and test conditions in the UK: 70 questionnaires on the real-life essay task, 73 on the real-life report task, and 160 on GEPT Advanced Writing Task 1. For Phase 2 of the study, 192 questionnaires were collected on GEPT Advanced Writing Task 1 in Taiwan.

This sub-section begins with the results pertaining to the cognitive processes performed by the participants under real-life conditions. Descriptive statistics of individual questionnaire items from each of the real-life tasks and the comparison of the cognitive processes employed on the two real-life tasks are reported in Section 5.2.1.1 Results from the exploratory factor analysis (EFA) of the number and underlying structure of the cognitive parameters involved in each of the five hypothesised academic writing cognitive constructs are reported in Section 5.2.1.2. In the context of language tests, it is important to collect evidence to show if the

cognitive parameters can distinguish more proficient writers from less proficient writers. A comparison of the cognitive processes employed by high-achieving and low-achieving participants in the real-life context is presented in Section 5.2.1.3.

After defining the target cognitive validity parameters, Section 5.2.2 reports the results comparing the cognitive processes elicited between the real-life and testing conditions collected at Phase 1 of the study. Section 5.2.2.1 compares the cognitive processes elicited by GEPT Advanced Writing Task 1 and the real-life tasks (whole group and in groups of high-, medium- and low-achievers). Section 5.2.2.2 reports the results from the exploratory factor analysis (EFA) of the underlying structure of the cognitive processes elicited by GEPT Advanced Writing Task 1 and discusses the extent to which the results resembles the target cognitive parameters identified in the study.

Section 5.2.3 then reports the findings of Phase 2 of the study, to confirm whether similar evidence of the cognitive validity of GEPT Advanced Writing Task 1, as collected under a research setting in the UK, can be obtained under the live test condition in Taiwan. Comparisons of the cognitive processes elicited by GEPT Advanced Writing Task 1 under the two conditions are reported in Section 5.2.3.1 and results of EFA in Section 5.2.3.2.

5.2.1.1 Significant differences between the two real-life tasks in terms of individual questionnaire items

First, the means and standard deviations of the rating (*4=definitely agree; 3=mostly agree; 2=mostly disagree; 1=definitely disagree*) of the 48 items from the two sets of real-life data were obtained. Mann-Whitney U tests were performed to investigate if the participants employed the individual items differently between the two real-life tasks. The results showed that apart from 3 items (i.e. Item 1.4, 4.1 and 4.14), the differences in all items between the two real-life tasks were non significant (The results of the items with significant difference are presented in Table 12; the results of all items are provided in Appendix 4). This indicates that the participants rated the extent to which they employed most items in a similar way on the two real-life tasks, and the actual differences were relatively slight.

Table 12 Significant differences between the two real-life tasks in terms of individual items

		Essay			Report			Mann-Whitney U	Wilcoxon W	z	p
		Median	Mean	Std Dev	Median	Mean	Std Dev				
1.4	I was able to understand the instructions for this writing task very well.	3.00	3.03	.636	3.00	3.25	.703	2091.500	4576.500	-2.105	.035
4.1	While I was writing, I sometimes paused to organise my ideas.	3.00	3.16	.673	3.00	2.86	.751	2040.500	4741.500	-2.292	.022
4.14	I checked the possible effect of my writing on the intended reader while I was writing.	3.00	3.06	.814	3.00	2.75	.846	2059.500	4760.500	-2.162	.031

As shown in Table 12, the participants reported that they understood the instructions for the report task better on a scale from 1 (*definitely disagree*) to 4 (*definitely agree*) than the essay task. This agrees with the results in the contextual analyses, reported in Section 5.1.1, where the expert judges regarded that the report task presented clearer information about the communicative purpose and the intended reader than the essay task.

In addition, participants rated the extent to which they *paused to organise their ideas* and *checked the possible effect on the intended reader during writing* higher on the essay task than on the report task. The essay task seemed to have elicited from the participants a higher awareness of the needs of the intended reader than the report task. The results in the contextual analyses may offer an explanation. According to the judges, the essay task required students to *persuade* whereas the report task did not. The marking schemes of the two real-life tasks may offer more details. The report task was scored based on four categories: (1) examination of the data and description of the nature of the dataset; (2) discussion and justification of the techniques chosen; (3) reasons for rejecting the inappropriate techniques; and (4) discussion of other relevant issues. The essay task was scored based on: (1) problem definition and structure of the text; (2) information identification (the number of sources, relevance to the topic, reliability of the sources); (3) critical reason; and (4) persuasion and influencing. In comparison to the report task, the quality of the essay is more dependent on the persuasiveness of the content. This may be why the participants checked the possible effect on the intended reader significantly more on the essay task than the report task.

For the remaining 45 items, the results of the Mann-Whitney U tests showed no significant difference in the participants' rating between the two real-life tasks (see Appendix 4). With regards to individual questionnaire items, the participants seemed to have employed the processes very similarly between the real-life essay and report tasks.

5.2.1.2 The underlying structure of the cognitive processes elicited by the real-life academic writing tasks

As mentioned earlier, the 48 questionnaire items were developed to quantify the cognitive processes involved in each of the five hypothesised academic writing cognitive constructs drawn from the literature (Field, 2004, 2008, 2011, 2013; Shaw & Weir, 2007):

- 1) conceptualisation;
- 2) meaning and discourse construction;
- 3) organisation;
- 4) low-level monitoring and revising; and
- 5) high-level monitoring and revising

In order to define the target cognitive validity parameters of reading-into-writing tests for academic purposes, the underlying structure of the cognitive processes involved in each of these hypothesised academic writing constructs was examined by Exploratory Factor Analysis (EFA). The findings will provide statistical evidence of (1) how many distinct cognitive processes loaded on each academic writing cognitive construct, and (2) the extent to which each distinct cognitive process loaded on the academic writing construct. Considering the three reasons given below, it was decided to analyse the data collected from the two real-life tasks together in the subsequent factor analyses.

(1) The means of 97% of the questionnaire items showed no significant difference between the two tasks;

(2) The primary purpose of the study was to define the cognitive constructs measured by predominant real-life academic writing tasks, and the examined the extent to which they are measured by GEPT Advanced Task 1. Provided that the participants reported using individual processes similarly between the two real-life tasks, analysing the data as a whole group would improve the generalisability of the results; and

(3) Analysing the data together would increase the size of the data which is beneficial for factor analysis.

First of all, the *Kaier-Meyer-Olkin (KMO) measure of sampling adequacy* and the *Bartlett's test of sphericity* were performed to test the real-life processing data's appropriateness for factor analysis. The results showed that the data passed both tests, indicating its suitability for factor analysis. In addition, the data was analysed by *Kolmogorov-Smirnov* test with regards to its normal distribution. The results showed that the real-life processing data was not normally distributed ($p < 0.01$). Following Fabrigar et al's (1999) recommendation for non-normally distributed data in factor analysis, the *Principal Axis Factor Method* was performed to extract the initial factors. The *eigenvalues*⁵ and *scree plot*⁶ were examined for an initial indication of the possible number of factors extracted by the data. Rotated solutions of the factor loadings, which avoid maximising the variance accounted by the first factor, were obtained by an oblique promax rotation. Oblique rotation method is often used in social science and language studies because it allows the factors to be correlated. To determine the ultimate number of underlying factors to be extracted, a few possibilities were interpreted and

⁵ Factors with an eigenvalues below 1 need to be dropped.

⁶ The point (also called an elbow) where there is a sudden drop of the steepness of the curve indicates that the factors on its left are significant.

evaluated based on both statistical results and theoretical rational. The underlying structure of the five hypothesised academic writing phases elicited on the real-life tasks are presented and discussed below.

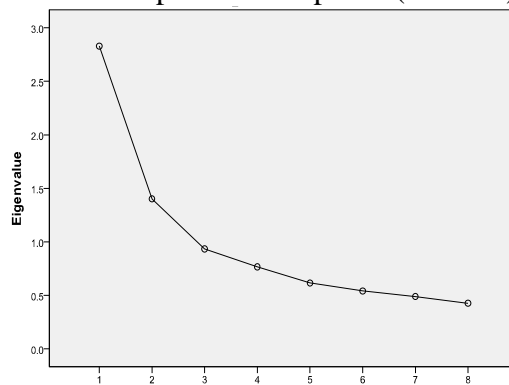
The underlying structure of the conceptualisation phase (real-life)

Conceptualisation is the first phase of productive skills (Kellogg, 1996, Field, 2004, 2011) where a writer develops an initial task representation, i.e. an initial understanding of the rhetorical situation of the writing task (Flower, 1990) and where a writer sets macro-plans, i.e., to establish writing goals in different aspects such as intended readership, genre, content and style (Shaw & Weir, 2007). Lower-level reading processes, i.e. decoding, lexical search, and parsing, were not sampled in the questionnaire of this study because students at the undergraduate level are judged to have already mastered high automaticity in these lower-level reading processes.

The conceptualisation phase was measured by 8 questionnaire items in this study. The initial factor extraction for the conceptualisation phase elicited in the real-life conditions yielded two factors with eigenvalues greater than 1.0. The scree plot also suggested a two-factor solution (see Table 13).

Table 13 Eigenvalues and scree plot for the conceptualisation phase (real-life)

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.829	35.363	35.363
2	1.401	17.517	52.880
3	.934	11.670	64.550
4	.765	9.567	74.117
5	.615	7.690	81.806
6	.541	6.757	88.563
7	.489	6.112	94.675
8	.426	5.325	100.000



The two-factor solution was accepted. The rotated two-factor solution for the conceptualisation phase and the interacted correlations of the factors are presented in Table 14. The percentage in brackets indicates the extent to which each factor accounts for the variance. As shown in Table 14, Factor 1 contains the majority of the 8 items, which include three *macro-planning* processes (i.e. Item 1.2, 1.3, 1.5) with regards to the relevance and adequacy of content, purpose of the task and effect on intended reader, and three *task representation* processes at different stages of the writing process (i.e. Item 1.4, 2.6, 4.4). The factor was named *task representation and macro-planning*. Factor 2 contains two items only, both relating to the process of changing macro plans at different stages of the writing production, one after reading the source texts (Item 2.13), the other while writing the first draft (Item 4.6). The factor was named *revising macro plan*.

Table 14 Pattern and interfactor correlations matrix for the conceptualisation phase (real-life)

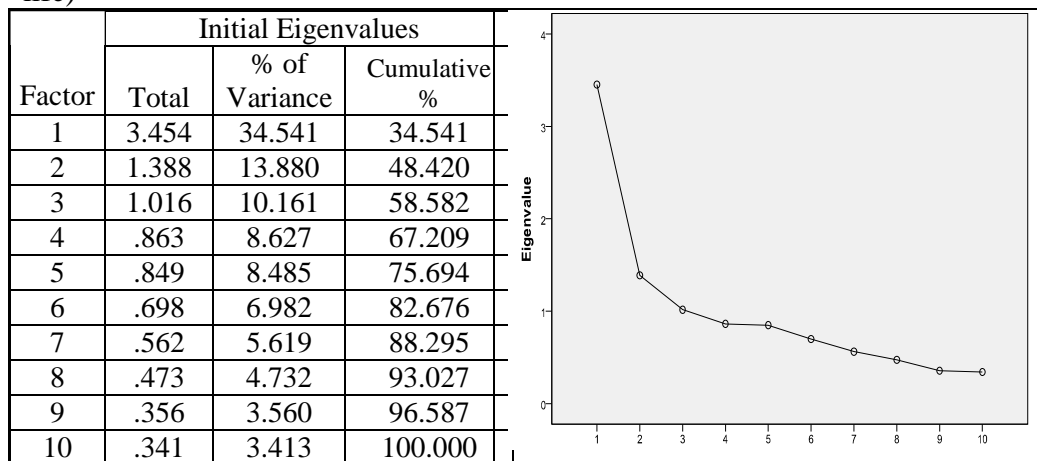
		F1 Task representation and macro-planning (34%)	F2 Revising macro plan (19.09%)
1.2	I thought of what I might need to write to make my text relevant and adequate to the task.	.806	
1.5	After reading the prompt, I thought about the purpose of the task.	.688	
1.4	I was able to understand the instructions for this writing task very well.	.588	
1.3	I thought of how my text would suit the expectations of the intended reader.	.519	
4.4	I re-read the task prompt while I was writing.	.449	
2.6	I read the task prompt again while I was reading the source texts.	.445	
2.13	I changed my writing plan while reading the source texts.		.804
4.6	I changed my writing plan while I was writing.		.583
Interfactor correlations			
Factor 1 (Task representation and macro-planning)		1.000	
Factor 2 (Revising macro plan)		.061	1.000

The results suggest that, as indicated by Factor 1, task representation and macro-planning processes were employed by the participants together in a similar way when they conceptualised an understanding of the writing task and established their macro plans to complete the task in the real-life academic context. On the other hand, Factor 2 empirically supports Hayes & Flower's (1983) writing model that planning is a not a one-off process employed in the beginning of the writing process, but a recursive process employed throughout the writing production process. The results show that participants revised their macro plans at different stages of the writing production in real-life academic writing. Interestingly, as shown in Table 14, the two factors correlated weakly ($r=0.061$). In other words, the process of *revising macro plan* at later stages of the writing production is apparently a writing construct distinctive from the other conceptualisation processes.

The underlying structure of the meaning and discourse construction phase (real-life)

Meaning and discourse construction is a higher-level phase where the writer (1) contextualises abstract meanings based on the contextual clues provided in the writing task, (2) identifies what information (which could be retrieved from long-term memory or selected from input texts) is relevant to the context, and (3) identifies how information from different sources connects to each other and to the task (Field, 2013; Spivey, 1997). The *meaning and discourse construction* phase was measured by 11 questionnaire items in this study. The initial factor extraction for the *meaning and discourse construction* phase elicited in real-life academic conditions produced three factors with eigenvalues greater than 1.0. The scree plot suggested two- or three- factor solutions (see Table 15).

Table 15 Eigenvalues and scree plot for the meaning and discourse construction phase (real-life)



The rotated two-factor and three-factor solutions were compared. The two-factor solution (provided in Appendix 5 Table 1) showed that Factor 1 includes reading processes to select relevant ideas and some processes of *connecting and generate* while Factor 2 includes careful reading processes and a process of generating new ideas or better understanding. However, three items loaded on both factors. Therefore the solution was rejected. The three-factor solution, on the other hand, showed a clearer distinction between the factors. Based on the initial three-factor solution (see Table 16), Factor 1 includes mostly *connecting and generating* items, while Factor 2 and 3 include *reading* items. However, Item 4.5 and 1.1 loaded on more than one factor. The fact that these two reading items loaded on more than one factor implies that these items might have contributed to other underlying constructs. Another reason could be that participants employed these reading processes differently from other reading processes. Future study should investigate into this issue. They were dropped from the factor analysis. The final rotated three-factor pattern and interfactor correlations matrix is presented in Table 17.

Table 16 Pattern matrix for the meaning and discourse construction phase (real-life): initial three-factor solution

Items		F1	F2	F3
2.12	I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.	.739		
4.5	I selectively re-read the source texts while writing.	.588	.498	
2.9	I linked the important ideas in the source texts to what I know already.	.548		
4.2	I developed new ideas while I was writing.	.407		
4.3	I made further connections across the source texts while I was writing.	.274		
2.5	I read some relevant part(s) of the texts carefully.		.866	
2.7	I took notes on or underlined the important ideas in the source texts.		.569	
2.4	I searched quickly for part(s) of the texts which might answer the question		.462	
2.1	I read through the whole of each source text carefully.			.979
1.1	I read the whole task prompt (i.e. instructions) carefully to understand each word in it.	.391		.451
2.2	I read the whole of each source text more than once.			.414

Table 17 Pattern and the interfactor correlations matrix for the meaning and discourse construction phase (real-life)

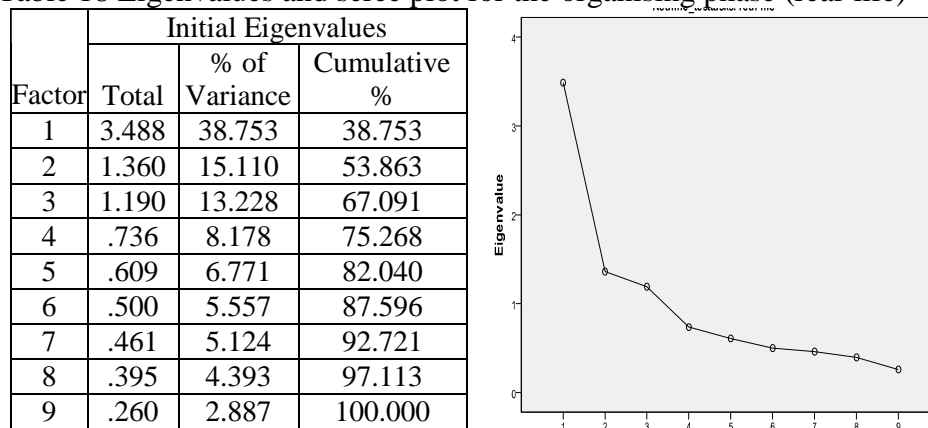
		F1 Connecting and generating (33.27%)	F2 Selecting relevant ideas (12.53%)	F3 Careful global reading (9.57%)
2.12	I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.	.665		
2.9	I linked the important ideas in the source texts to what I know already.	.528		
4.2	I developed new ideas while I was writing.	.514		
4.3	I made further connections across the source texts while I was writing.	.274		
2.5	I read some relevant part(s) of the texts carefully.		.767	
2.7	I took notes on or underlined the important ideas in the source texts.		.715	
2.4	I searched quickly for part(s) of the texts which might answer the question		.383	
2.1	I read through the whole of each source text carefully.			.846
2.2	I read the whole of each source text more than once.			.509
Interfactor correlations matrix				
Factor 1 (Connecting and generating)		1.000		
Factor 2 (Selecting relevant ideas)		.595	1.000	
Factor 3 (Careful global reading)		.377	.223	1.000

Researchers argued that meaning and discourse construction involves high-level processes such as selecting relevant information, connecting ideas from different sources, and building a consistent discourse pattern (Field, 2003, 2008; Spivey, 1990, 1997). The results here show that the *meaning and discourse construction* phase elicited on the real-life tasks in this study involved three distinct yet correlated cognitive processes. As shown in Table 17, Factor 1 includes four items of connecting ideas from different sources and generating new representations. The factor was named *connecting and generating*. Factor 2 includes three reading items of identifying ideas which are relevant and important to the context of the writing task. The factor was named *selecting relevant ideas*. Factor 3 includes two global careful reading items. The factor was named *global careful reading*. According to the interfactor correlation matrix (see Table 17), Factor 2 (*selecting relevant ideas*) and Factor 3 (*careful global reading*) correlated weakly at 0.223. This supports the hypothesis that selective and search reading skills are different from global careful comprehension skills (Khalifa & Weir, 2009; Weir, Yang, & Jin, 2000). The results also reveal that Factor 1 (*connecting and generating*) correlated with Factor 2 (*selecting relevant ideas*) more than with Factor 3 (*careful global reading*). Both results indicate a need to test selective reading skills in writing tests for academic purposes.

The underlying structure of the organisation phase (real-life)

Organisation is a phase 'where the writer provisionally organises the ideas, still in abstract form, (a) in relation to the text as a whole and (b) in relation to each other (Field, 2004, 329)'. The organisation construct was measured by 9 questionnaire items in this study. The initial factor extraction for the organisation construct produced three factors with eigenvalues greater than 1.0. The scree plot suggested two- or three- factor solutions (see Table 18). The rotated two- and three-factor solutions were compared. The three-factor solution (provided in Appendix 5 Table 2) showed that Factor 1 is related to organising main ideas. Factor 2 is related to organising the structure of the texts. However, Factor 3 is difficult to interpret. In addition, 2 items loaded on two factors. Therefore this solution was rejected.

Table 18 Eigenvalues and scree plot for the organising phase (real-life)



The two-factor solution, on the other hand, provides a clearer distinction between the factors. Based on the initial two-factor solution (see Table 19), Factor 1 includes mostly the processes of organising ideas from the source texts while Factor 2 includes mostly the processes of organising ideas in relation to the writer's own text. However, Item 3.4 did not load on either of the factors at a level of 0.3 or above. It was dropped from the analysis.

Table 19 Pattern matrix for the organising phase (real-life): initial two-factor solution

Items		F1	F2
2.10	I worked out how the main ideas in each source text relate to each other.	.906	
2.11	I worked out how the main ideas across the source texts relate to each other.	.880	
2.8	I prioritised important ideas in the source texts in my mind.	.704	
2.3	I used my knowledge of how texts like the source texts are organised to find parts to focus on.	.482	
3.2	I recombined or reordered the ideas to fit the structure of my text.		.836
3.1	I organised the ideas for my text before starting to write.		.795
3.3	I removed some ideas I planned to write.		.688
4.1	While I was writing, I sometimes paused to organise my ideas.		.513
3.4	I tried to use the same organizational structure as in one of the source texts	<.3	<.3

After the removal, the rotated two-factor solution was extracted (see Table 20).

Table 20 Pattern and interfactor correlations matrix for the organising phase (real-life)

		F1 Organising ideas in relation to input texts (34.73%)	F2 Organising ideas in relation to new text (16.60%)
2.11	I worked out how the main ideas across the source texts relate to each other.	.984	
2.10	I worked out how the main ideas in each source text relate to each other.	.761	
2.8	I prioritised important ideas in the source texts in my mind.	.564	
2.3	I used my knowledge of how texts like the source texts are organised to find parts to focus on.	.437	
3.3	I removed some ideas I planned to write.		.691
3.2	I recombined or reordered the ideas to fit the structure of my text.		.627
3.1	I organised the ideas for my text before starting to write.		.441
4.1	While I was writing, I sometimes paused to organise my ideas.		.314
Interfactor correlations matrix			
Factor 1 (Organising ideas in relation to input texts)		1.000	
Factor 2 (Organising ideas in relation to new text)		.533	1.000

Factor 1 involves four items of organising ideas in relation to a single input text or in relation to multiple input texts. The factor was named *organising ideas in relation to input texts*. Factor 2 involves four items of organising ideas in relation to the writer's own text. The factor was named *organising ideas in relation to own text*.

In the writing literature, the ability to organise ideas according to the communicative purpose of the task has been regarded as an important writing skill (e.g. Hayes & Flower, 1983; Shaw & Weir, 2007). In the reading-into-writing literature, the ability to organise has also been regarded as important when writers use external sources (e.g. Flower et al., 1990; Plakans, 2009a, 2009b; Spivey & King, 1989; Spivey, 1984).

The results of this study showed that the organisation phase elicited by the real-life tasks involves two distinct underlying cognitive processes - *organising ideas in relation to input texts* and *organising ideas in relation to own text*. This supports Spivey's notion that organising is a core process of writing from sources – a writer constructs their text by organising the ideas he/she selected from the input texts. The results of this study revealed that the process of organising ideas in relation to textual and intertextual representation of the input texts is distinctive from the process of organising ideas in relation to the writer's own text. The pattern matrix showed that the two types of organising processes correlate moderately at 0.533 (See Table 20). As both organising processes were elicited in the real-life academic conditions, it is important for EAP tests to sample both processes from the test takers. The process of organising in relation to textual and intertextual representation of the

input texts has been neglected in most standardised academic writing tests which use the impromptu writing task type.

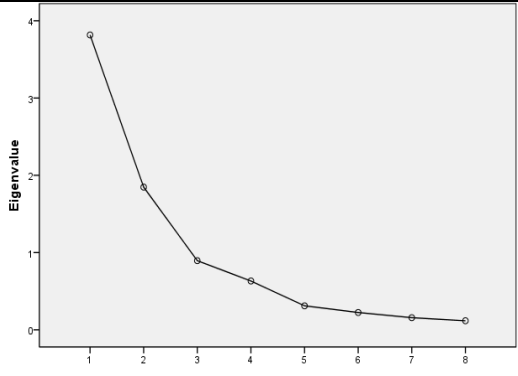
The underlying structure of the low-level monitoring and revising phase (real-life)

Low-level monitoring involves primarily checking the linguistic accuracy, e.g. spelling, grammar and sentence structure of the text. After monitoring, a writer usually revises the unsatisfactory parts of the text (Field, 2004: 330). The low-level monitoring and revising processes can be done at any time during writing at the word, sentence or paragraph level, or after the whole draft has been completed. The low-level monitoring and revision construct was measured by 8 questionnaire items in this study.

The initial factor extraction for the low-level monitoring and revision construct produced two factors with eigenvalues greater than 1.0. The scree plot suggested two- or four- factor solutions (see Table 21).

Table 21 Eigenvalues and scree plot for the low-level monitoring and revision phase (real-life)

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3.816	47.703	47.703
2	1.847	23.087	70.790
3	.895	11.190	81.979
4	.632	7.899	89.879
5	.311	3.884	93.762
6	.225	2.808	96.571
7	.157	1.965	98.535
8	.117	1.465	100.000



The rotated two- and four- factor solutions were compared. The four-factor solution reflected to some extent the categorisation of different types of low-level monitoring and revising processes, e.g. grammar vs. use of own words. However, Factor 3 and Factor 4 consisted of one item only. In addition, two items loaded on two factors. Therefore, the four-factor solution (provided in Appendix 5 Table 3) was rejected.

On the other hand, the two-factor solution reflected clearly the distinction between the low-level monitoring and revising processes employed during the writing process and those employed after the whole draft has been completed (see Table 22).

Factor 1 contains four low-level monitoring and revising processes employed after the whole draft has been completed. The factor was named *low-level editing after writing*. Factor 2 contains four low-level monitoring and revising processes employed during the writing process. The factor was named *low-level editing during writing*.

Table 22 Pattern and interfactor correlations matrix for the low-level monitoring and revising phase (real-life)

		F1 Low-level editing after writing (47.70%)	F2 Low-level editing during writing (23.9%)
5.12	After I had finished the first draft, I checked that the quotations were properly made.	.898	
5.13	After I had finished the first draft, I checked that I had put the ideas of the source texts into my own words.	.872	
5.15	After I had finished the first draft, I checked the grammatical accuracy and range of the sentence structures.	.848	
5.16	After I had finished the first draft, I checked the spelling, usage and range of the vocabulary	.847	
4.16	I checked the spelling, usage and range of the vocabulary while I was writing.		.908
4.15	I checked the grammatical accuracy and range of the sentence structures while I was writing.		.783
4.12	I checked that the quotations were properly made while I was writing.		.515
4.13	I checked that I had put the ideas of the source texts into my own words while I was writing.		.488
Interfactor correlations matrix			
Factor 1 (Low-level editing after writing)		1.000	
Factor 2 (Low-level editing during writing)		.347	1.000

In this study, the participants were asked to report the extent to which they revised different aspects of linguistic accuracy of their text while they were writing their text and after they had completed their first draft. Researchers (Field, 2004; Kellogg, 1996; Shaw & Weir, 2007) argued that monitoring and revising are highly demanding in terms of cognitive effort. Writers, especially L2 writers, tend to focus on one aspect of the text at a time due to short-term memory constraints. With attentional constraints, many writers would set aside the revising process to a later stage of the production. The results here confirm this notion. The results showed that the low-level editing processes employed by the participants in the real-life conditions clustered towards two stages, i.e. *during the writing process* and *at the final stage after drafting the text*.

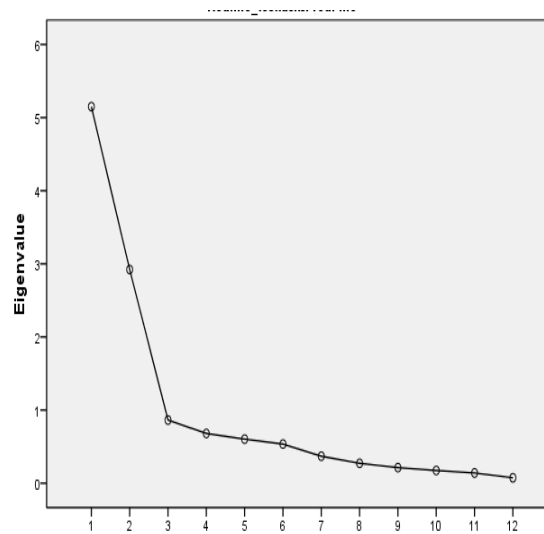
The underlying structure of the high-level monitoring and revising phase (real-life)

High-level monitoring primarily involves checking the effect of the text, in terms of clarity and appropriateness of ideas, coherence of arguments, style, and possible effect on reader. Similar to the low-level revising and monitoring construct, after high-level monitoring, a writer will usually revise the unsatisfactory parts of the text (Field, 2004, 330).

The high-level monitoring and revision phase elicited on the real-life tasks was measured by 12 questionnaire items in this study. The initial factor extraction for the high-level monitoring and revising construct produced two factors with eigenvalues greater than 1.0. The scree plot also suggested a two-factor solution (see Table 23).

Table 23 Eigenvalues and scree plot for the high-level monitoring and revising phase (real-life)

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5.150	42.920	42.920
2	2.922	24.351	67.271
3	.864	7.198	74.469
4	.681	5.673	80.141
5	.603	5.023	85.164
6	.536	4.464	89.628
7	.370	3.081	92.709
8	.272	2.271	94.980
9	.213	1.777	96.757
10	.175	1.454	98.211
11	.140	1.170	99.381
12	.074	.619	100.000



Similar to the low-level monitoring and revising phase, the rotated two-factor solution (see Table 24) produces simple and interpretable factor loadings which reflect the underlying distinction between the high-level editing processes employed while the participants were writing and those employed after they had completed the first draft. Factor 1 contains six high-level monitoring and editing processes employed after the whole draft has been completed. The factor was named *high-level editing after writing*. Factor 2 contains six high-level monitoring and revising processes employed during the writing process. The factor was named *high-level editing during writing*.

Table 24 Pattern and interfactor correlations matrix for the high-level monitoring and revising phase (real-life)

		F1 High-level editing after writing (42.92%)	F2 High-level editing during writing (24.35%)
5.9	After I had finished the first draft, I checked that my text was coherent.	.903	
5.8	After I had finished the first draft, I checked that my text was well-organised.	.880	
5.11	After I had finished the first draft, I checked that I included my own viewpoint on the topic.	.877	
5.7	After I had finished the first draft, I checked that the content was relevant.	.865	
5.10	After I had finished the first draft, I checked that I included all appropriate main ideas from all the source texts.	.797	
5.14	After I had finished the first draft, I checked the possible effect of my writing on the intended reader.	.748	
4.9	I checked that my text was coherent while I was writing.		.747
4.7	I checked that the content was relevant while I was writing.		.618
4.8	I checked that my text was well-organised while I was writing.		.613
4.10	I checked that I included all appropriate main ideas from all the source texts while I was writing.		.581
4.11	I checked that I included my own viewpoint on the topic while I was writing.		.575
4.14	I checked the possible effect of my writing on the intended reader while I was writing.		.446
Interfactor correlations matrix			
Factor 1 (High-level editing after writing)		1.000	
Factor 2 (High-level editing during writing)		.208	1.000

5.2.1.3 Can these cognitive parameters distinguish how well high-achieving and low-achieving students completed the real-life tasks?

The purpose of RQ2 was to define the cognitive parameters which are appropriate for reading-into-writing tests for academic purposes. The above results of exploratory factor analysis have defined eleven cognitive processes which participants in this study employed to complete the real-life writing tasks. A further step to investigate cognitive validity was whether these defined cognitive parameters can effectively distinguish the writing processes of more proficient writers from those of less proficient writers.

The 143 participants who completed either one of the real-life tasks were ranked according to their scores. Each performance can be scored from 0-16. The participants were divided into

four groups (i.e. high, higher middle, lower middle and low), representing roughly 25% of the population each. 40 participants were classified as the high-achieving group (a score of 12 or above) and 39 participants were classified as the low-achieving group (a score of 7.5 or below). The high-achieving group scored significantly higher than the low-achieving group ($t=10.398(77)$, $p<0.001$).

The *mean rank* and *sum of ranks* of the average rating ($4=$ definitely agree; $3=$ mostly agree; $2=$ mostly disagree; $1=$ definitely disagree) of the eleven underlying cognitive process reported by the high-achieving and low-achieving groups are presented in Table 25. The means reported by the two groups were analysed by Mann-Whitney U tests (see Table 25).

Table 25 Comparisons between high-achieving and low-achieving participants

	High-achieving (n=40)		Low-achieving (n=39)		Mann- Whitney U	Wilcoxon W	Z	Asymp. Sig. (2- tailed)
	Mean Rank	Sum of Ranks	Mean Rank	Sum of Ranks				
Conceptualisation								
Task representation and macro-planning	44.86	1794.50	35.01	1365.50	585.500	1365.500	-1.917	.055
Revising macro plan	39.17	1567.00	40.85	1593.00	747.000	1567.000	-.330	.741
Meaning and discourse construction								
Careful global reading	43.51	1740.50	36.40	1419.50	639.500	1419.500	-1.412	.158
Selecting relevant ideas	47.64	1905.50	32.17	1254.50	474.500	1254.500	-3.058	.002
Connecting and generating	43.80	1752.00	36.10	1408.00	553.500	1333.500	-2.243	.025
Organising								
Organising ideas in relation to input texts	46.46	1858.50	33.37	1301.50	521.500	1301.500	-2.552	.011
Organising ideas in relation to own text	45.95	1838.00	33.90	1322.00	542.000	1322.000	-2.360	.018
Low-level monitoring and revising								
Low-level editing while writing	46.44	1857.50	33.40	1302.50	522.500	1302.500	-2.578	.010
Low-level editing after writing	39.19	1567.50	40.83	1592.50	747.500	1567.500	-.322	.747
High-level monitoring and revising								
High-level editing while writing	45.66	1826.50	34.19	1333.50	553.500	1333.500	-2.233	.026
High-level editing after writing	39.23	1569.00	40.79	1591.00	749.000	1569.000	-.306	.760

As shown in Table 25, the high achieving participants reported employing eight of the eleven cognitive processes (i.e. *task representation and macro-planning*, *careful global reading*, *selecting relevant ideas*, *connecting and generating*, *organising ideas in relation to input texts*, *organising ideas in relation to own text*, *low-level editing while writing* and *high-level editing while writing*) more than the low achieving groups. Apart from *task representation and macro-planning* and *careful global reading*, all differences were significant. The mean ranks of the processes of *revising macro plan*, *low-level editing after writing* and *high-level editing after writing* between the high-achieving and low-achieving groups were very close (see Table 25).

For the conceptualisation phase, the high-achieving participants employed the processes of *task representation and macro-planning* more than the low-achieving participants, though the

difference was not significant. On the other hand, there was not much difference in the extent to which they employed the processes of *revising macro plan* at later stages of the writing process. According to Scardamalia & Bereiter's (1987) continuum of writing expertise, mature writers are cautious about the rhetorical situation of the writing task and therefore actively establish writing goals to fulfil the communicative purpose of the task. In contrast, most immature writers might not be able to establish a complete representation of the writing task (e.g. communicative purpose, intended readership, genre, content and style) and therefore would not establish comprehensive macro plans before they start to write. The results here seem to support this notion. Researchers (Field, 2004; Kellogg, 2001; Shaw & Weir, 2007) further argued that due to attentional constraints, L2 writers may have extra difficulty in building a comprehensive task representation and establishing macro plans because most of their of attention may be devoted to lower-level reading processes, e.g. lexical decoding and parsing (i.e. connecting words to meaning). Another challenge for weaker L2 writers is that they may not be able to hold their task representation and macro-plan in working memory while they are executing their plans (Field, 2013).

With regards to the meaning and discourse construction phase, the high-achieving participants reported employing the processes of *careful global reading*, *selecting relevant ideas* and *connecting and generating* more than the low-achieving group (See Table 25). Field's (2004, 2011, 2013) model of receptive skills illustrated that meaning construction and discourse construction are more cognitively demanding than decoding, lexical searching and parsing. He argued that less proficient language users would focus on the lower processes while the proficient language users would have high automaticity in executing these lower processes. The proficient language users would then be able to focus on meaning construction and discourse construction when they read. Khalifa & Weir's (2007) model of reading skills similarly argued that lower proficiency L2 readers would focus on establishing understanding at local level (words, phrases, sentences) while high proficiency readers would be able to establish understanding at global textual and intertextual level. In addition, researchers (e.g. Spivey 1991) argued that mature writers are able to connect ideas from different internal sources (e.g. topical knowledge and discourse knowledge) and external sources (e.g. input texts) to the context of writing task and generate new understanding of the rhetorical challenge presented by the writing task. Scardamalia & Bereiter (1987) argued that this is why mature writers are able to transform knowledge as they write. The results of this study showed that in the real-life academic context, these high-level processes of careful and selective reading skills at global textual and intertextual level and the processes of connecting and generating distinguish the high-achieving the low-achieving participants. In other words, these high level processes are important for success in academic writing. Therefore, it is important that these processes are targeted in academic writing tests.

With regards to the organising phase, the high-achieving participants reported organising ideas both in relation to input texts and the writer's own text significantly more than the low-achieving participants.

Regarding the low-level and high level monitoring and revising phases, the high-achieving participants reported employing the processes of *while writing* low-level editing and high-level editing significantly more than the low-achieving participants did. However, there was not much difference in the extent to which they reported the use of *after writing* low-level and high-level editing. Field (2004) argued that monitoring and revising processes are highly cognitively demanding, particularly for L2 writers. While monitoring and revising can be employed at any time during the writing process, most writers can only focus on one aspect

of the editing at one time (e.g. grammatical accuracy at low level, or argument coherence at high level). Therefore, weaker writers who have not acquired high automaticity in the translating process tend not to be able spare attention on monitoring and revising. The results reported here confirm this notion. The high-achieving participants in this study seemed to be more capable of performing editing at both low- and high- levels *while they were writing* than the low-achieving participants.

5.2.1.4 Summary of the cognitive processes elicited by the real-life tasks

The results of the EFA showed that the conceptualisation phase in the real-life academic context involved *task representation* and *macro-planning* processes to conceptualise an understanding of the writing task and establish their macro plans. The process of *revising macro plan* is particularly important in real-life academic writing. The two processes explained 53.9% of the variance of the conceptualisation phase. The meaning and discourse construction phase involved three underlying processes: *connecting and generating*, *selecting relevant ideas* and *global careful reading*. The three processes explained 58.58% of the variance of the meaning and discourse construction phase. The organising phase involved the processes of *organising ideas in relation to input texts* as well as *organising ideas in relation to writer's own text*. The two processes explained 51.33% of the variance of the organising phase. For both low-level monitoring and revising and high-level monitoring and revising phases, there was a clear distinction between the *editing processes employed while writing* and those employed *after the first draft has been completed*. The two low-level monitoring and revising processes explained 71.6% of the variance of the low-level monitoring and revising phase whereas the two high-level monitoring and revising processes explained 67.27% of the variance of the low-level monitoring and revising phase. While the results of this study identified eleven processes involved in the five cognitive phases elicited by the real-life academic writing tasks, the results indicated that the variance of each cognitive phase was not fully accounted. Future study should explore the additional cognitive processes involved in each phase, perhaps with a different research method, such as keystroke logging and stimulated recall. In addition, the results showed that the high achieving participants reported employing eight of the eleven cognitive processes more than low achieving participants, and the differences in six processes were significant. There seems to be a good case for considering these processes as the target cognitive validity parameters for a valid writing test for academic purposes.

5.2.2 Investigating the cognitive validity of GEPT Advanced Writing Task 1

Having defined the target cognitive validity parameters by the real-life data, we now move on to reporting the results that compared the extent which these parameters were elicited between GEPT Advanced Writing Task 1 and the real-life tasks (whole group and in groups of high-, medium- and low-achievement), followed by the results from the exploratory factor analysis (EFA) of the underlying structure of the cognitive processes elicited by GEPT Advanced Writing Task 1.

5.2.2.1 Comparisons of the cognitive processes elicited between the real-life and test conditions

This sub-section presents the results comparing how the eleven cognitive parameters were employed by the participants under the test and real-life conditions. First of all, the 160 participants in the UK context were ranked according to their scores on the GEPT task. They were divided into four groups (i.e. high, higher-medium, lower-medium and low achieving),

representing roughly 25% of the population each. 36 participants were classified as high-achieving (with a score of 10.5 or above out of 20) and 34 participants were classified as low-achieving (with a score of 7.5 or below). The high-achieving group scored significantly higher than the low-achieving group ($t=17.148(124)$, $p<0.001$).

After that, a total of 109 out of 160 participants were then selected into three achievement groups based on their GEPT Advanced Writing Task 1 and real-life scores:

- (1) High-achieving group – 14 participants who were identified as high-achieving on both the real-life and test tasks
- (2) Middle-achieving group – 85 participants who were identified as either higher-medium or lower-medium on both the real-life and test tasks
- (3) Low-achieving group – 13 participants who were identified as low-achieving on both the real-life and test tasks

The comparisons of cognitive processes elicited under test and real-life conditions by the whole population and in groups of high-, medium- and low-achievement are presented in Table 26. The results reported here will indicate if these eleven cognitive parameters can sufficiently resemble the cognitive processes that the test takers (as a whole group as well as in groups of high-, medium, or low- achievement) would normally employ in non-test conditions. The comparisons were analysed by the Wilcoxon signed ranks test (non-parametric related sample).

Table 26 Comparisons between GEPT and real-life cognitive processing data

	High-achieving group						Middle-achieving group					
	GEPT		Real-life tasks		Z	Sig.	GEPT		Real-life tasks		Z	Sig.
	Mean	Std Dev	Mean	Std Dev			Mean	Std Dev	Mean	Std Dev		
Conceptualisation												
Task representation and macro-planning	3.21	.58	3.14	.47	-.566	.572	2.94	.64	3.15	.525	-2.005	.045
Revising macro plan	2.39	.59	2.75	.80	-1.650	.099	2.3	.75	2.69	.90	-2.248	.025
Meaning and discourse construction												
Careful global reading	2.42	1.17	2.89	1.07	-1.767	.077	2.93	.65	2.83	.65	-.947	.344
Selecting relevant ideas	3.66	.36	3.50	.53	-.654	.513	3.21	.82	3.17	.55	-1.031	.303
Connecting and generating	2.92	.74	3.28	.44	-1.446	.148	3.00	.68	3.15	.53	-1.261	.207
Organising												
Organising ideas in relation to source texts	3.10	.62	3.00	.46	-.594	.552	2.92	.60	2.89	.49	-.705	.481
Organising ideas in relation to own text	3.11	.62	3.41	.51	-1.259	.208	2.97	.65	3.10	.53	-1.732	.083
Low-level monitoring and revising												
Low-level editing while writing	3.10	.82	3.35	.55	-.595	.552	2.75	.72	3.08	.59	-3.262	.001
Low-level editing after writing	2.37	1.01	2.66	.95	-.971	.331	2.37	1.01	2.81	.90	-3.140	.002
High-level monitoring and revising												
High-level editing while writing	3.15	.79	3.23	.51	-.118	.906	2.81	.65	3.11	.54	-3.904	.000
High-level editing after writing	2.61	1.25	2.76	1.00	-.070	.940	2.36	1.02	2.76	.92	-2.69	.010

Table 26 Comparisons between GEPT and real-life cognitive processing data (continue)

	Low-achieving group						Whole group					
	GEPT		Real-life task		Z	Sig.	GEPT		Real-life tasks		Z	Sig.
	Mean	Std Dev	Mean	Std Dev			Mean	Std Dev	Mean	Std Dev		
Conceptualisation												
Task representation and macro-planning	2.98	.56	2.98	.28	-.039	.969	2.92	.66	3.17	.49	-1.814	.070
Revising macro plan	2.53	.74	2.76	.64	-1.08	.280	2.56	.71	2.78	.71	-2.890	.004
Meaning and discourse construction												
Careful global reading	3.15	.51	2.81	.63	-1.218	.223	2.86	.74	2.94	.69	-.577	.564
Selecting relevant ideas	3.43	.58	3.05	.67	-1.620	.105	3.22	.82	3.24	.56	-1.403	.161
Connecting and generating	3.00	.47	2.96	.46	-.356	.722	2.89	.70	3.20	.51	-1.861	.063
Organising												
Organising ideas in relation to source texts	3.06	.39	2.78	.46	-1.505	.132	2.91	.62	2.95	.49	-1.069	.285
Organising ideas in relation to own text	2.76	.52	2.86	.33	-.768	.443	3.01	.72	3.13	.52	-2.355	.019
Low-level monitoring and revising												
Low-level editing while writing	2.59	.62	2.88	.67	-.994	.320	2.81	.74	3.12	.61	-3.562	.000
Low-level editing after writing	1.61	.92	2.63	1.06	-2.558	.011	2.43	.99	2.80	.90	-3.595	.000
High-level monitoring and revising												
High-level editing while writing	2.98	.53	2.93	.38	-.223	.823	2.86	.72	3.16	.54	-3.889	.000
High-level editing after writing	1.74	1.04	2.61	1.02	-2.572	.010	2.44	1.04	2.79	.92	-2.808	.005

As shown in Table 26, the high-achieving participants, those whose performances were ranked in the top 25% on both GEPT and real-life tasks, reported employing all cognitive processes similarly in both conditions. No differences obtained were statistically significant. The low-achieving group (those whose performances were ranked in the bottom 25% on both The GEPT and real-life tasks) reported employing most of the cognitive processes similarly in both conditions. However, they employed the low-level and high-level editing processes after writing significantly more on the real-life tasks than on GEPT. The low-achieving participants reported an average rating of 1.61 (4=definitely agree; 3=mostly agree; 2=mostly disagree; 1=definitely disagree) for low-level editing after writing and 1.74 for high-level editing after writing on GEPT. This implies that they did not employ these after writing editing processes on GEPT, probably because the low-achieving test takers did not have the processing capacity to deal with these editing processes. Besides, they were unlikely to be less aware of the importance of after-writing editing than the high-achieving writers (see Table 26). Compared to the other two achievement groups, the middle-achieving group showed greater discrepancy in how the participants employed the processes when carrying out GEPT and real-life tasks (see Table 26). There was no significant difference in the extent to which they reported employing the processes of the meaning and discourse construction and organising phases (i.e. *careful global reading*, *selecting relevant ideas*, *connecting and generating*, *organising ideas in relation to input texts* and *organising ideas in relation to own text*) between the two conditions. However, they reported employing the processes of the conceptualisation phase (i.e. *task representation and macro-planning* and *revising macro plan*) and of the monitoring and revising phases (i.e. *low-level editing while writing*, *low-level editing after writing*, *high-level editing while writing*, and *high-level editing after writing*) significantly more in the real-life conditions than under the test conditions (see Table 26).

Regarding the whole population, the participants reported employing all the eleven cognitive processes more on a scale from 1 (*definitely disagree*) to 4 (*definitely agree*) in the real-life conditions than under the test condition. The differences reported in six processes (which include *revising macro plan*, *organising ideas in relation to own text*, *low-level editing during writing*, *low-level editing after writing*, *high-level editing during writing*, and *high-level editing after writing*) were significant ($p < 0.05$) (see Table 26). It must be noted that in real life students have considerably more time to monitor and revise and iteratively revisit their work. Few tests give dedicated time for either planning or monitoring, and exam tasks are performed under serious time pressure. Time for monitoring and revising is more limited in the exam situation (see Table 26).

In short, the results provided empirical evidence supporting the cognitive validity of this reading-into-writing task type (essay task with multiple verbal inputs). The eleven cognitive processes were elicited more in the real-life conditions than under the test condition. However, there was no significant difference in the extent to which the high-achieving participants reported employing all processes on GEPT Advanced Writing Task 1 and the real-life tasks. The high-achieving group seemed to be able to utilise the processes under both conditions. This suggests that provided that the test takers were proficient in academic writing, the GEPT task was able to elicit the same processes from test takers to a similar extent as they employed them on the real-life tasks. Apart from the processes of after writing low-level and high-level editing, the low-achieving participants reported employing all cognitive processes on the GEPT task in a similar manner as they employed them in the real-life academic contexts. The middle group reported employing five of the processes on the GEPT task in a similar manner as they did on the real-life tasks. The middle group seemed to show greater discrepancy in how they employed the processes under the test and real-life

conditions. This is probably because they were in transitional state of developing their academic writing skills and perhaps more affected by the performance conditions, e.g. stricter time allowance. In addition, it is encouraging that all three proficiency groups reported employing all processes of the meaning and discourse construction phase and the organising phase (*i.e. careful global reading, selecting relevant ideas, connecting and generating, organising ideas in relation to input texts, and organising ideas in relation to own text*) in the test and real-life conditions without any significant differences in their mean rating. This supports the literature that the reading-into-writing task type is a valid task type to test the process of discourse synthesis (Spivey, 1984, 2001; Spivey & King, 1989), which is a core set of academic writing skills. These processes are also believed to play an important role in critical academic literacy (Flower et al., 1990).

5.2.2.2 Comparisons of the underlying constructs elicited by GEPT Advanced Writing Task 1 and those by real-life tasks

The last step of Phase 1 was to examine the underlying structure of the cognitive processes elicited by GEPT Advanced Writing Task 1, to investigate to what extent the GEPT task elicited from the participants the same underlying cognitive constructs as the real-life tasks did. Here, exploratory factor analysis was used again to examine the underlying structure of the five hypothesised writing phases elicited by GEPT Advanced Writing Task 1. It was thought that exploratory factor analysis would be more appropriate for the analysis here rather than confirmatory factor analysis. Confirmatory factor analysis is usually used to test whether a particular data set fit a particular measurement model, *i.e.* a model of cognitive processes in this case. Since there is apparently insufficient empirical evidence of the cognitive processes elicited by reading-into-writing tests in the literature, exploratory factor analysis seems most appropriate to examine the underlying structure of the cognitive processes elicited by GEPT Advanced Writing Task 1.

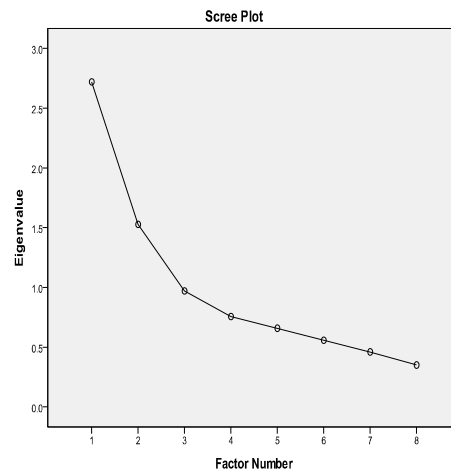
Kaier-Meyer-Olkin (KMO) test and Bartlett's test of sphericity were performed for each of the five academic writing cognitive constructs. The results showed that the data collected on the GEPT was appropriate for factor analyses. As the data was not normally distributed (K-S test: $p < 0.01$), the principal axis factor method with the promax rotation procedure was performed. The factor solutions indicated by the eigenvalues and the scree plot were evaluated to determine the ultimate number of underlying factors to be extracted. The EFA results on each hypothesised academic writing cognitive construct elicited by the real-life tasks and GEPT Advanced Writing Task 1 are presented one by one below.

The underlying structure of the conceptualisation phase (GEPT)

Conceptualisation is the first phase of the writing process where writers conceptualise the writing task and set macro plans. As presented in Section 5.2.1.2, the conceptualisation phase elicited by the real-life tasks in this study involved the processes of *task representation and macro-planning* (which was measured by six questionnaire items) and *revising macro plan* (which was measured by two items). On the GEPT data, the initial factor extractions yielded two factors with eigenvalues greater than 1.0 and the scree plot also suggested a two-factor solution (see Table 27). The two-factor suggestion was accepted.

Table 27 Eigenvalues and scree plot for the conceptualisation phase (GEPT)

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.720	33.996	33.996
2	1.527	19.090	53.086
3	.970	12.123	65.210
4	.757	9.457	74.667
5	.658	8.228	82.895
6	.557	6.968	89.863
7	.460	5.751	95.614
8	.351	4.386	100.000



The pattern matrix and the interfactor correlations for the conceptualisation construct on The GEPT task are presented in Table 28. The percentage in brackets indicates the extent to which each factor accounts for the variance.

Table 28 Pattern and interfactor correlations matrix for the conceptualisation phase (GEPT)

		F1 Task representation and macro-planning (33.98%)	F2 Revising macro plan (19.04%)
1.2	I thought of what I might need to write to make my text relevant and adequate to the task.	.805	
1.5	I thought about the purpose of the task.	.663	
1.4	I was able to understand the instructions for this writing task very well.	.584	
1.3	I thought about how my text would suit the expectations of the intended reader.	.510	
2.6	I read the task prompt again while I was reading the source texts.	.465	
4.4	I re-read the task prompt while I was writing.	.445	
2.13	I changed my writing plan while reading the source texts.		.799
4.6	I changed my writing plan while I was writing.		.584
Interfactor correlations			
Factor 1 (Task representation and macro-planning)		1.000	
Factor 2 (Revising macro plan)		.042	1.000

The two-factor solution presented in Table 28 resembles the two factors generated by the real-life data. Like the conceptualisation phase elicited by the real-life tasks, the same six items loaded on Factor 1 (which was named *task representation and macro-planning*) and the same two items loaded on Factor 2 (which was named *revising macro plan*).

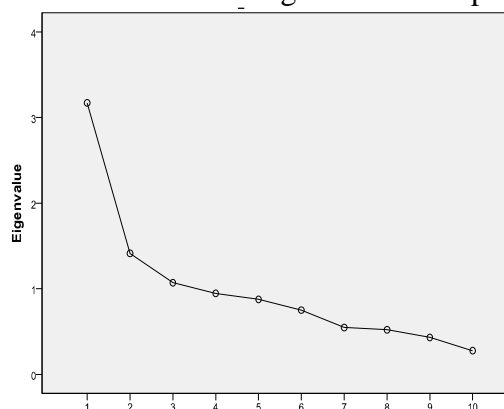
The underlying structure of the meaning and discourse construction phase (GPET)

Meaning and discourse construction is a higher-level phase where students contextualise meaning and establish discourse representations from different sources. As presented in Section 5.2.1.2, the meaning and discourse construction phase elicited on the real-life tasks involved three distinct underlying cognitive processes. The first process was careful reading

at global level (*careful global reading*). The second one was to select ideas which are relevant to the writing task (*selecting relevant ideas*). The last one was to generate links between ideas or new meaning by connecting ideas/discourse features provided in the source texts (*connecting and generating*). On GEPT, the initial factor extraction for the meaning and discourse construction construct produced three factors with eigenvalues greater than 1.0. The scree plot suggested two- or three- factor solutions (see Table 29). The rotated two- and three- factor solutions were compared. The three-factor solution (provided in Appendix 6 Table 1) was rejected because Factor 3 includes one item only. Besides, Item 4.2 did not load on any factors at the level of 0.3 or above.

Table 29 Eigenvalues and scree plot for the discourse and meaning construction phase (GEPT)

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3.171	31.712	31.712
2	1.413	14.128	45.840
3	1.070	10.702	56.542
4	.945	9.455	65.996
5	.876	8.757	74.753
6	.750	7.502	82.256
7	.547	5.472	87.728
8	.521	5.207	92.935
9	.431	4.307	97.242
10	.276	2.758	100.000



The two-factor solution was accepted (see Table 30). Like the meaning and discourse construction phase elicited by real-life tasks, Factor 1 (*selecting relevant ideas*) includes the same three items of search reading for ideas which are relevant to the task. However, Factor 2 on the GEPT task includes four items, which were loaded as separate factors in real-life conditions. Factor 2 on The GEPT task was named *connecting and generating with careful global reading*.

Table 30 Pattern matrix and interfactor correlations for the discourse and meaning construction phase (GEPT)

		F1 Selecting relevant ideas (33.38%)	F2 Connecting and generating with careful global reading (14.09%)
2.5	I read some relevant part(s) of the texts carefully.	.892	
2.4	I searched quickly for part(s) of the texts which might answer the question.	.807	
2.7	I took notes on or underlined the important ideas in the source texts.	.607	
2.9	I linked the important ideas in the source texts to what I know already.		.649
4.3	I made further connections across the source texts while I was writing.		.531
2.1	I read through the whole of each source text carefully.		.489
4.2	I developed new ideas while I was writing.		.476
2.2	I read the whole of each source text more than once.		.447
2.12	I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.		.402
Interfactor correlations matrix			
Factor 1 (Selecting relevant ideas)		1.000	
Factor 2 (Connecting and generating with careful global reading)		.175	1.000

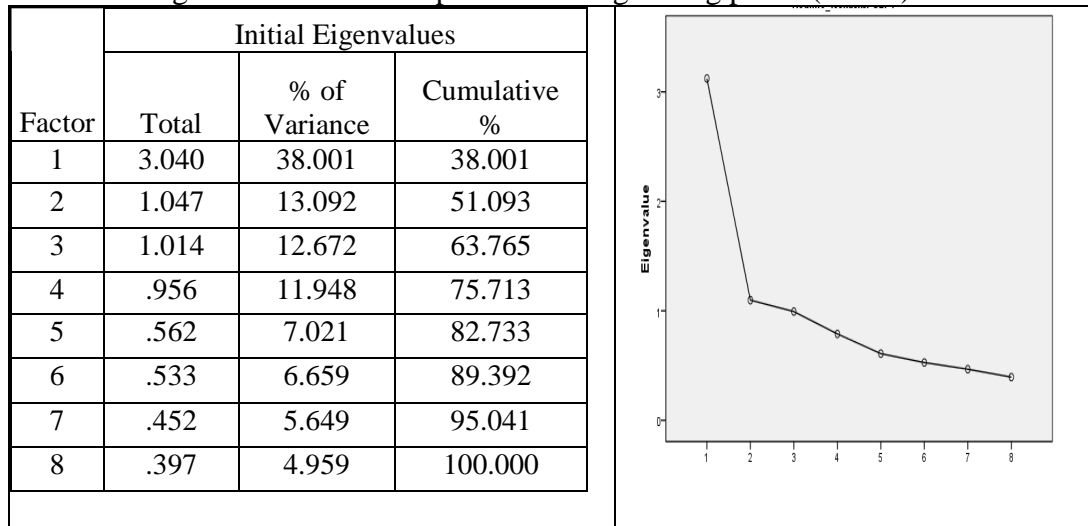
It is interesting to explore why the processes of careful global reading employed on the GEPT task did not represent a stand-alone factor. Descriptive results reported earlier showed that the mean rating of the careful reading was 2.86 (on a scale of 1 to 4) whereas the use of selecting relevant ideas was 3.22 on the GEPT task. When compared to real-life conditions, the GEPT task imposes a tighter time constraint on test takers. The contextual analysis reported in Section 5.1 may also offer an explanation why the participants employed less global careful reading on the GEPT task than real-life tasks. When compared to real-life tasks, according to the judges, the GEPT task was placed more towards the lower end of the cognitive demands of transforming content from source texts to writer's own text. In addition, the content of the GEPT task source texts was more concrete, and the textual organisation of the GEPT task source texts was clearer than real-life source texts.

The underlying structure of the organising phase (GEPT)

The use of organising processes is an important academic writing construct which provides evidence of distinguishing different levels of writing expertise. Scardamalia & Bereiter (1987) argued that immature writers tend to translate the ideas from their long-term memory to their text in the same order as the idea retrieval. In contrast, mature writers would explicitly organise the ideas which have been retrieved from long-term memory according to their macro writing plans. Section 5.2.1.2 showed the organising phase elicited by the real-life tasks involved two distinct cognitive processes. The first one (*organising ideas in relation to input texts*) was related to the processes of organising the textual and intertextual representations of the input texts. The second one (*organising ideas in relation to new text*)

was related to the processes of organising the ideas to be put in the writer's own text. On the GEPT task, the initial factor extraction for the organisation construct produced three factors with eigenvalues greater than 1.0. The scree plot, however, suggested one or two-factor solutions (see Table 31). The one-factor solution was not investigated. Rotated solutions with two or three factors were compared.

Table 31 Eigenvalues and scree plot for the organising phase (GEPT)



The three-factor solution (provided in Appendix 6 Table 2) was rejected because Factor 2 and Factor 3 included one primary item only. Besides, Item 3.1 loaded on two factors. The two-factor solution was accepted. The initial results showed that Item 3.2 loaded on both factors while Item 3.3 loaded on neither at a level of 0.3 or above (see Table 32). They were dropped from the analysis. The process of reordering and recombining ideas (Item 3.2) loaded on both factors. This implies that while the participants reordered and recombined ideas, they might have focused on both organising their representation of the input texts as well as their own text. Further evidence is needed to confirm this. Besides, the process of removing ideas (Item 3.3) did not load on either of the factors at the level of 0.3 or above. This suggests that the participants did not employ the process of removing ideas from their plans in the same way as they employed other organising processes.

Table 32 Pattern matrix for the organising phase (GEPT): initial two-factor solution

Items		F1	F2
2.11	I worked out how the main ideas across the source texts relate to each other.	.858	
2.10	I worked out how the main ideas in each source text relate to each other.	.614	
2.8	I prioritised important ideas in the source texts in my mind.	.587	
3.2	I recombined or reordered the ideas to fit the structure of my essay.	.494	.309
2.3	I used my knowledge of how texts like these are organised to find parts to focus on.	.493	
3.3	I removed some ideas I planned to write.	<.3	<.3
4.1	While I was writing I sometimes paused to organise my ideas.		.633
3.1	I organised the ideas I planned to include in my essay.		.426

After the removal, the two-factor solution was extracted again (see Table 33). Similar to the real-life data, the underlying structure of the organising processes elicited on the GEPT task yields two distinct cognitive processes. Factor 1 (*organising ideas in relation to input texts*) consisted of the same four items of organising the textual or intertextual representations of the input texts, as identified by the real-life data. However, Factor 2 (*organising ideas in relation to own texts*) elicited on the GEPT task included fewer items than the same factor elicited by the real-life tasks. The two factors correlated at a level of 0.61.

Table 33 Pattern and interfactor correlations matrix for the organising phase (GEPT)

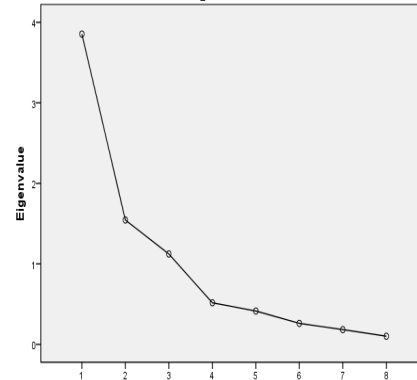
		F1 Organising ideas in relation to input texts (41.70%)	F2 Organising ideas in relation to own text (17.65%)
2.11	I worked out how the main ideas across the source texts relate to each other.	.842	
2.8	I prioritised important ideas in the source texts in my mind.	.604	
2.10	I worked out how the main ideas in each source text relate to each other.	.545	
2.3	I used my knowledge of how texts like these are organised to find parts to focus on.	.447	
3.1	I organised the ideas I planned to include in my essay.		.637
4.1	While I was writing, I sometimes paused to organise my ideas.		.505
Interfactor correlations matrix			
Factor 1 (Organising ideas in relation to input texts)		1.000	
Factor 2 (Organising ideas in relation to own text)		.612	1.000

The underlying structure of the low-level monitoring and revising phase (GEPT)

Low-level monitoring and revising is a phase where the writer monitors the quality of their own text (mainly in terms of grammatically accuracy) and revises the unsatisfactory parts of the text (Field, 2004). Section 5.2.1.2 showed that the low-level monitoring and revising phase elicited by the real-life tasks consisted of two distinct processes. The first one was low-level editing employed during writing, and the other one was low-level editing employed after the first draft has been produced. On the GEPT task, the initial factor extraction for the low-level monitoring and revising processes produced three factors with eigenvalues greater than 1.0. The scree plot suggested two- or three- factor solutions (see Table 34).

Table 34 Eigenvalues and scree plot for the low-level revising phase (GEPT)

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3.855	48.187	48.187
2	1.545	19.308	67.495
3	1.123	14.032	81.527
4	.517	6.460	87.987
5	.415	5.181	93.168
6	.261	3.267	96.436
7	.183	2.291	98.727
8	.102	1.273	100.000



Rotated solutions with two- and three- factor solutions were compared. The first factors extracted by both solutions were the same. According to the three-factor solution (provided in Appendix 6 Table 3), Factor 2 focused on the linguistic accuracy whereas Factor 3 focused on the appropriate use of source texts. The two-factor solution (see Table 35) was taken because it resembled more closely the underlying structure extracted from the real-life data than the three-factor solution. As per the real-life data, the low-level monitoring and revising phase elicited by the GEPT task involved two distinctive cognitive processes: *low-level editing after writing* and *low-level editing during writing*. The items loaded on each factor were the same as identified by the real-life tasks. The two processes correlated moderately at a level of 0.45.

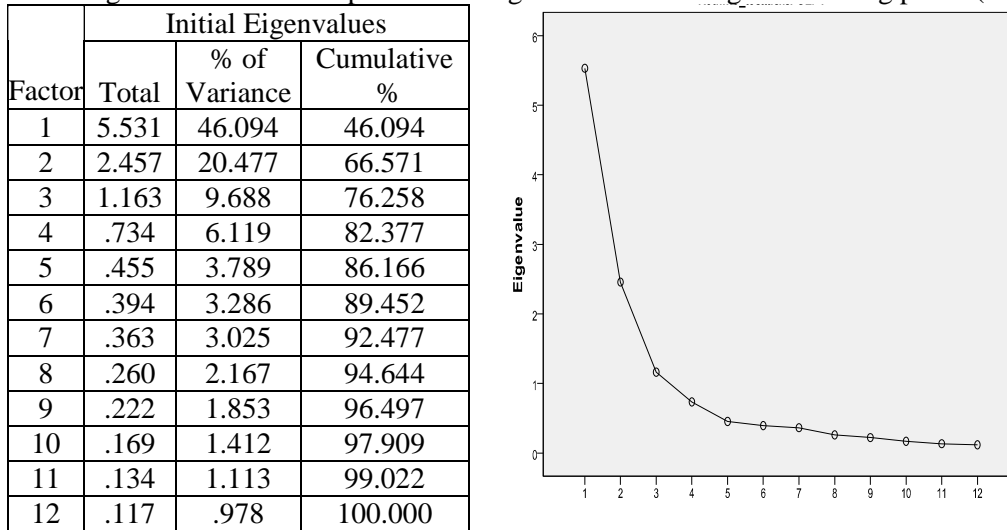
Table 35 Pattern and interfactor correlations matrix for the low-level revising phase (GEPT)

		F1 Low-level editing after writing (48.18%)	F2 Low-level editing while writing (19.31%)
5.12	After I had finished the first draft, I checked that the quotations were properly made.	.857	
5.15	After I had finished the first draft, I checked the grammatical accuracy and range of the sentence structures.	.832	
5.16	After I had finished the first draft, I checked the spelling, usage and range of vocabulary.	.809	
5.13	After I had finished the first draft, I checked that I had put the ideas of the source texts into my own words.	.801	
4.15	I checked the grammatical accuracy and range of the sentence structures while I was writing.		.930
4.16	I checked the spelling, usage and range of vocabulary while I was writing.		.724
4.12	I checked that the quotations were properly made while I was writing.		.501
4.13	I checked that I had put the ideas of the source texts into my own words while I was writing.		.421
Interfactor correlations matrix			
Factor 2 (Low-level editing after writing)		1.000	
Factor 1 (Low-level editing while writing)		.450	1.000

The underlying structure of the high-level monitoring and revising phase (GEPT)

Similar to the low-level monitoring and revising phase, Section 5.2.1.2 showed that the high-level monitoring and revising phase elicited by the real-life tasks consisted of two distinct processes. Factor 1 included the processes of high-level editing employed during writing, and Factor 2 included the processes of high-level editing employed after the first draft has been produced. On the GEPT task, the initial factor extraction for the high-level monitoring and revising phase produced three factors with eigenvalues greater than 1.0. The scree plot suggested a two-factor or four-factor solutions (see Table 36). Rotated two-, three- and four-factor solutions were compared. The three-factor solution (provided in Appendix 6 Table 4) was rejected because Factor 3 included only one primary item. The four-factor solution (provided in Appendix 6 Table 5) was rejected because Factor 3 included only one item and Factor 4 had no primary factor (i.e. all its items loaded more heavily on another factor).

Table 36 Eigenvalues and scree plot for the high-level monitoring and revising phase (GEPT)



The two-factor solution was accepted. The initial two-factor solution (see Table 37) showed that Item 4.14 did not load on any factor at the level of 0.3 or above. The item was dropped from the analysis.

Table 37 Pattern matrix for the high-level monitoring and revising phase (GEPT): initial two-factor solution

Items		F1	F2
5.7	After I had finished the first draft, I checked that the content was relevant.	.914	
5.10	After I had finished the first draft, I checked that I included all appropriate main ideas from all the source texts.	.878	
5.8	After I had finished the first draft, I checked that my text was well-organised.	.861	
5.9	After I had finished the first draft, I checked that my text was coherent.	.861	
5.11	After I had finished the first draft, I checked that I included my own viewpoint on the topic.	.830	
5.14	After I had finished the first draft, I checked the possible effect of my writing on the intended reader.	.806	
4.10	I checked that I included all appropriate main ideas from all the source texts while I was writing.		.789
4.7	I checked that the content was relevant while I was writing.		.780
4.8	I checked that my text was well-organised while I was writing.		.704
4.11	I checked that I included my own viewpoint on the topic while I was writing.		.701
4.9	I checked that my text was coherent while I was writing.		.686
4.14	I checked the possible effect of my writing on the intended reader.	<.3	<.3

After the removal, the rotated two-factor solution was extracted again (see Table 38). Similar to the real-life data, the high-level monitoring and revising phase elicited on the GEPT task consisted of two distinctive processes: *high-level editing after writing* (F1) and *high-level*

editing while writing (F2). However, Factor 2 elicited by the GEPT task did not include the process of checking the possible effect on the intended reader (Item 4.14).

Table 38 Pattern and interfactor correlations matrix for the high-level monitoring and revising phase (GEPT)

		F1 High-level editing after writing (47.42%)	F2 High-level editing while writing (19.46%)
5.7	After I had finished the first draft, I checked that the content was relevant.	.914	
5.10	After I had finished the first draft, I checked that I included all appropriate main ideas from all the source texts.	.879	
5.8	After I had finished the first draft, I checked that my text was well-organised.	.866	
5.9	After I had finished the first draft, I checked that my text was coherent.	.865	
5.11	After I had finished the first draft, I checked that I included my own viewpoint on the topic.	.832	
5.14	After I had finished the first draft, I checked the possible effect of my writing on the intended reader.	.800	
4.10	I checked that I included all appropriate main ideas from all the source texts while I was writing.		.802
4.7	I checked that the content was relevant while I was writing.		.783
4.11	I checked that I included my own viewpoint on the topic while I was writing.		.707
4.8	I checked that my text was well-organised while I was writing.		.692
4.9	I checked that my text was coherent while I was writing.		.669
Interfactor correlations matrix			
Factor 2 (High-level editing after writing)		1.000	
Factor 1 (High-level editing while writing)		.373	1.000

5.2.2.3 Summary

Table 39 summarises the findings of the underlying structure of the cognitive processes employed in the five writing phases elicited by the real-life tasks and the GEPT task (the order of the factor follows the results of the explanatory factor analyses). As described in detail above, 5 out of 48 items were dropped from the analysis. 2 individual items which yielded complex loadings on the *meaning and discourse construction* construct were dropped from the EFA analyses of the GEPT data. 2 items were dropped from the *organisation* construct, and 1 item from the *high-level monitoring and revising* construct. Apart from that, the GEPT task was largely able to elicit from the participants the same underlying cognitive processes as the real-life tasks did. The number of factors involved in 4 out of 5 phases of academic writing, which included **conceptualisation**, **organising**, **low-level organising and revising**, and **high-level monitoring and revising**, elicited on the GEPT task and the real-life tasks were identical. Seven factors within these phases, which included *task representation and macro-planning*, *revising macro plan*, *selecting relevant ideas*, *organising ideas in relation to input texts*, *low-level editing after writing*, *low-level editing while writing* and *high-level editing after writing*, elicited by the GEPT task contained the same individual questionnaire items as the corresponding factors identified by the real-life tasks.

Table 39 Comparisons of the underlying structure of the cognitive processes (real-life vs GEPT)

Underlying cognitive processes elicited on real-life tasks	Underlying cognitive processes elicited on the GEPT task
Conceptualisation	
F1: Task representation and macro-planning (34%)	F1: Task representation and macro-planning (33.98%)
F2: Revising macro plan (19.9%)	F2: Revising macro plan (19.04%)
Meaning and discourse construction	
F1: Connecting and generating (34.54%)	F1: Selecting relevant ideas (33.38%)
F2: Selecting relevant ideas (13.88%)	F2: Connecting and generating with careful global reading (14.09%)
F3: Careful global reading (10.16%)	-
Organisation	
F1: Organising ideas in relation to input texts (34.73%)	F1: Organising ideas in relation to input texts (41.70%)
F2: Organising ideas in relation to own text (16.60%)	F2: Organising ideas in relation to own text (17.65%)
Low-level monitoring and revising	
F1: Low-level editing after writing (47.70%)	F1: Low-level editing after writing (47.70%)
F2: Low-level editing during writing (23.9%)	F2: Low-level editing during writing (19.31%)
High-level monitoring and revising	
F1: High-level editing after writing (42.92%)	F1: High-level editing after writing (47.42%)
F2: High-level editing during writing (24.35%)	F2: High-level editing during writing (19.46%)

The results of the explanatory factor analysis revealed that GEPT Advanced Writing Task 1 was able to elicit most of the academic writing processes in the same manner as the real-life tasks. Nevertheless, some discrepancies were shown on the underlying structure of the cognitive processes of the meaning and discourse construction phase between the GEPT and

the real-life tasks. The process of *connecting and generating* was identified as the first factor on the real-life tasks, *selecting relevant ideas* the second, and *careful global reading* the third. The order of the factors indicates the percentage of the variance explained by each factor. In other words, the process of *connecting and generating* accounted for the largest percentage (i.e. 34.54%) of the variance of all cognitive processes within the meaning and discourse construction phase elicited by the real-life tasks. In contrast, the GEPT task had the process of *selecting relevant ideas* (search reading) as the first factor of the meaning and discourse construction phase. This indicates that the process of *selecting relevant ideas* was most important within the meaning and discourse construction phase elicited by the test tasks. In addition, the process of careful global reading employed on the GEPT task did not yield a stand-alone factor. The process of careful global reading clustered with the process of connecting and generating as the second factor. Future studies should investigate these issues. Although there were some discrepancies in the underlying structure of the meaning and discourse construction phase activated by the real-life academic writing tasks and GEPT Advanced Writing Task 1, the participants reported employing the process of expeditiously selecting relevant ideas more than careful reading processes on the GEPT task as they did on the real-life tasks.

In addition, the organising phase elicited by the real-life tasks involved two distinct cognitive processes. The findings showed that the participants distinctively employed the processes to organise their representation in relation to the input texts and those to their own text, with a stronger attention on the latter, under real-life conditions. Generally speaking, GEPT Advanced Writing Task 1 was able to elicit these two distinct organising processes from the participants. However, the process of *organising ideas in relation to own text* elicited on the GEPT task involved fewer process items than the corresponding factor identified by the real-life data. It requires further evidence to discuss why the participants did not organise their own text in the same way as they did in real-life conditions. For example, the process of removing ideas (Item 3.3) did not load on any of the factors at the level of 0.3 or above. It might be helpful to provide guidelines on task prompts to encourage test takers to devote equal attention to both organising processes.

GEPT Advanced Writing Task 1 did well in eliciting the same underlying processes of the low-level and high-level monitoring and revising phases. The GEPT task was able to elicit the processes of *while writing* and *after writing* low-level editing and high-level editing from test takers in the same manner as they were employed on the real-life tasks.

5.3 The cognitive processes elicited under the GEPT live test condition

Phase 2 was a replication study on the cognitive processes elicited by the GEPT task carried out under the live test condition in Taiwan.

5.3.1 Comparisons of the cognitive processes elicited by the GEPT task: university context in the UK vs. live test conditions in Taiwan

The processes elicited by the GEPT Advanced Writing Task 1 under live test conditions in Taiwan were compared to the results obtained in the university context in the UK by Mann-Whitney U tests (see Table 40).

Table 40 Comparisons of the processes elicited by the GEPT task university research context and the live test conditions

	GEPT university context in UK		GEPT live test condition in Taiwan		Mann-Whitney U	Wilcoxon W	Z	Sig.
	Mean Rank	Sum of Ranks	Mean Rank	Sum of Ranks				
Task representation and macro-planning	179.57	28731.50	173.94	33396.50	14868.500	33396.500	-.521	.602
Revising macro plan	176.42	28227.50	176.57	33900.50	15347.500	28227.500	-.013	.989
Careful global reading	178.01	28482.00	175.24	33646.00	15118.000	33646.000	-.262	.793
Selecting relevant ideas	177.40	28383.50	175.75	33744.50	15216.500	33744.500	-.156	.876
Connecting and generating	177.15	28344.50	175.96	33783.50	15255.500	33783.500	-.111	.912
Organising ideas in relation to source texts	175.85	28136.00	177.04	33992.00	15256.000	28136.000	-.111	.912
Organising ideas in relation to own text	178.63	28580.50	174.73	33547.50	15019.500	33547.500	-.363	.716
Low-level editing while writing	180.23	28837.00	173.39	33291.00	14763.000	33291.000	-.637	.524
Low-level editing after writing	178.42	28546.50	174.90	33581.50	15053.500	33581.500	-.326	.744
High-level editing while writing	177.53	28404.00	175.65	33724.00	15196.000	33724.000	-.174	.862
High-level editing after writing	178.09	28495.00	175.17	33633.00	15105.000	33633.000	-.270	.787

As presented in Table 38, there was no significant difference in the mean rank of the average rating of all the eleven processes (4=definitely agree; 3=mostly agree; 2=mostly disagree; 1=definitely disagree). This indicates that the processes elicited by GEPT Advanced Writing Task 1 generated under the two conditions, i.e. Phase 1 research condition in the UK and Phase 2 live test condition in Taiwan, were comparable.

5.3.2 The underlying constructs elicited by GEPT Advanced Writing Task 1 under live test conditions in Taiwan

Exploratory factor analysis was used to examine the underlying structure of the five hypothesised writing phases elicited by GEPT Advanced Writing Task 1 under live test conditions. The same procedures used in Phase 1 were followed. Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity were performed for each of the five academic writing cognitive constructs. The results showed that the data collected on the GEPT (live test) was appropriate for factor analyses. The data was not normally distributed (K-S test: $p < 0.01$), therefore, the principal axis factor method with the promax rotation procedure was performed. The factor solutions indicated by the eigenvalues and the scree plot were evaluated to determine the ultimate number of underlying factors to be extracted. The data was analysed by LTTC to ensure the independence of the analysis of the two GEPT data sets. The EFA results on each hypothesised academic writing cognitive construct elicited by the real-life tasks and GEPT Advanced Writing Task 1 are discussed one by one below in line with the results of Phase 1.

The underlying structure of the conceptualisation phase (GEPT live test)

Conceptualisation is the first phase of the writing process where writers conceptualise the writing task and set macro plans. The two-factor solution of the GEPT live test data presented in Table 41 resembles the structure reported on the real-life and GEPT data in Phase 1. However, Factor 1 (*task representation and macro-planning*) on the GEPT live test data involves fewer processes. Item 4.4 and 2.6 loaded on Factor 2 (*revising macro plan*). This indicates that under live test conditions, test takers engaged more in rereading the task prompt when they revised their macro plan at later stage (Factor 2) than when they created their task representation and macro plan at the beginning. In other words, the process of rereading the task prompt associated more with the process of revising macro plan (Factor 2) than the process of task representation and macro-planning (Factor 1).

Table 41 Pattern and interfactor correlations matrix for the conceptualisation phase (GEPT live test)

		F1 Task representation and macro-planning (30.57%)	F2 Revising macro plan (20.93%)
1.3	I thought of how my text would suit the expectations of the intended reader.	.803	
1.2	I thought about what I might need to write to make my essay relevant and adequate to the task.	.799	
1.5	I thought about the purpose of the task.	.667	
1.4	I was able to understand the instructions for this writing test very well.	.666	
4.6	I changed my writing plan I was writing.		.758
2.13	I changed my writing plan while reading the source texts		.697
4.4	I re-read the task prompt while I was writing.		.632
2.6	I read the task prompt again while I was reading the source texts.		.597
Interfactor correlations			
Factor 1 (Task representation and macro-planning)		1.000	
Factor 2 (Revising macro plan)		.165	1.000

The underlying structure of the meaning and discourse construction phase (GEPT live test)

Meaning and discourse construction is a higher-level phase where writers contextualise meaning and establish discourse representations from different sources. The three-factor solution of the GEPT live test data presented in Table 42 resembles the structure more similarly to the structure of real-life rather than GEPT data reported in Phase 1. The only discrepancy between the GEPT live test data and the real-life data was that Item 2.9 loaded on Factor 1 (*Connecting and generating*) rather than Factor 2 (*Selecting relevant ideas*). This suggests that under the live test condition, test takers engaged in linking the important ideas in the source texts to their prior knowledge while they were selecting relevant ideas from the source texts.

Table 42 Pattern and interfactor correlations matrix for the meaning and discourse construction phase (GEPT live test)

		F1 Connecting and generating (29.24%)	F2 Selecting relevant ideas (13.91%)	F3 Careful global reading (12.44%)
4.3	I made further connections across the source texts while I was writing.	.859		
4.2	I developed new ideas while I was writing.	.792		
2.12	I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.	.586		
2.7	I took notes on or underlined the important ideas in the source texts.		.814	
2.4	I searched quickly for part(s) of the texts which might answer the question.		.650	
2.5	I read some relevant part(s) of the texts carefully.		.559	
2.9	I linked the important ideas in the source texts to what I know already.		.472	
2.1	I read through the whole of each source text carefully.			.794
2.2	I read the whole of each source text more than once.			.749
Interfactor correlations				
Factor 1 (Connecting and generating)		1.000		
Factor 2 (Selecting relevant ideas)		.271	1.000	
Factor 3 (Careful global reading)		.326	.231	1.000

The underlying structure of the organising phase (GEPT live test)

The organising phase is an important phase where writers would organise ideas from different sources to fulfil the writing goal. The two-factor solution of the GEPT live test data presented in Table 43 again resembles more closely the structure of real-life rather than GEPT data reported in Phase 1. The only discrepancy between the GEPT live test data and the real-life data here was that Item 3.3 loaded on Factor 1 (*Organising ideas in relation to input texts*) rather than Factor 2 (*Organising ideas in relation to own text*). This indicates that under the GEPT live test condition, test takers tended to 'remove ideas' while they were organising ideas in relation to the input texts whereas under real-life academic context, students tended to do so while they were organising ideas in relation to their own text. Nevertheless, it is encouraging to find that test takers engaged in this process under the live test condition whereas the same process did not load on any factor on the GEPT data under the research condition.

Table 43 Pattern and interfactor correlations matrix for the organising construct (GEPT live test)

		F1 Organising ideas in relation to input texts (38.82%)	F2 Organising ideas in relation to own text (13.61%)
2.11	I worked out how the main ideas across the source texts relate to each other.	.833	
2.8	I prioritised important ideas in the source texts in my mind.	.729	
2.10	I worked out how the main ideas in each source text relate to each other.	.726	
2.3	I used my knowledge of how texts like these are organised to find parts to focus on.	.585	
3.3	I removed some ideas I planned to write.	.327	
4.1	While I was writing, I sometimes paused to organize my ideas.		.849
3.1	I organised the ideas for my text before starting to write.		.772
3.2	I recombined or reordered the ideas to fit the structure of my essay.		.757
Interfactor correlations			
Factor 1 (Organising ideas in relation to input texts)		1.000	
Factor 2 (Organizing ideas in relation to own text)		.461	1 .000

The underlying structure of the low-level monitoring and revising phase (GEPT live test)

Low-level monitoring and revising is a phase where the writer monitors the quality of their own text (mainly in terms of grammatically accuracy) and revises the unsatisfactory parts of the text (Field, 2004). The two-factor solution of the GEPT live test data presented in Table 44 largely resembled the structure elicited by the real-life and GEPT data reported (they were the same) in Phase 1. However, Item 4.13 did not load on either F2 (*low-level editing while writing*) as it did on the real-life data and GEPT data in Phase 1. Item 4.13 concerns about the use of own words. The result indicates that under the live test condition, test takers did not engage in this process (i.e. checking the use of own words) together with the other three low-level editing processes while they were writing the task.

Table 44 Pattern and interfactor correlations matrix for the low-level monitoring and revising phase (After the removal of Item 4.13) (GEPT live test)

		F1 Low-level editing after writing (61.26%)	F2 Low-level editing while writing (14.43%)
5.13	After I had finished the first draft, I checked that I had put the ideas of the source texts into my own words.	.951	
5.12	After I had finished the first draft, I checked that the quotations were properly made.	.848	
5.16	After I had finished the first draft, I checked the spelling, usage and range of vocabulary.	.840	
5.15	After I had finished the first draft, I checked the grammatical accuracy and range of the sentence structures.	.821	
4.15	I checked the grammatical accuracy and range of the sentence structures while I was writing.		.975
4.16	I checked the spelling, usage and range of vocabulary while I was writing.		.886
4.12	I checked that the quotations were properly made while I was writing.		.596
Interfactor correlations			
Factor 1 (Low-level editing after writing)		1.000	
Factor 2 (Low-level editing while writing)		.581	1.000

The underlying structure of the high-level monitoring and revising phase (GEPT live test)

The two-factor structure of the high-level monitoring and revising phase of the GEPT live test data presented in Table 45 is exactly the same as the structure elicited by the real-life data reported in Phase 1.

Table 45 Pattern and interfactor correlations matrix for the high-level monitoring and revising phase (GEPT live test)

		F1 High-level editing after writing (51.36%)	F2 High-level editing while writing (13.32%)
5.11	After I had finished the first draft, I checked that I included my own viewpoint on the topic.	.963	
5.10	After I had finished the first draft, I checked that I included all appropriate main ideas from all the source texts.	.914	
5.9	After I had finished the first draft, I checked that my text was coherent.	.879	
5.7	After I had finished the first draft, I checked that the content was relevant.	.841	
5.8	After I had finished the first draft, I checked that my text was well-organised.	.803	
5.14	After I had finished the first draft, I checked the possible effect of my writing on the intended reader.	.658	
4.10	I checked that I included all appropriate main ideas from all the source texts while I was writing.		.863
4.9	I checked that my text was coherent while I was writing.		.755
4.8	I checked that my text was well-organised while I was writing.		.748
4.7	I checked that the content was relevant while I was writing.		.712
4.11	I checked that I included my own viewpoint on the topic while I was writing.		.661
4.14	I checked the possible effect of my writing on the intended reader while I was writing.		.658
Interfactor correlations			
Factor 1 (High-level editing after writing)		1.000	
Factor 2 (High-level editing while writing)		.562	1.000

5.3.3 Summary

Table 46 summarises the findings of the underlying structure of the cognitive processes employed at the five writing phases elicited by the GEPT live test (the order of the factor follows the results of the explanatory factor analyses).

Table 46 Comparisons of the underlying structure of the cognitive processes (real-life vs GEPT live test)

Underlying cognitive processes elicited on real-life tasks	Underlying cognitive processes elicited on the GEPT task (live test condition)
Conceptualisation	
F1: Task representation and macro-planning (34%)	F1: Task representation and macro-planning (30.57%)
F2: Revising macro plan (19.9%)	F2: Revising macro plan (20.93%)
Meaning and discourse construction	
F1: Connecting and generating (34.54%)	F1: Connecting and generating (29.24%)
F2: Selecting relevant ideas (13.88%)	F2: Selecting relevant ideas (13.91%)
F3: Careful global reading (10.16%)	F3: Careful global reading (12.44%)
Organising	
F1: Organising ideas in relation to input texts (34.73%)	F1: Organising ideas in relation to input texts (38.82%)
F2: Organising ideas in relation to own text (16.60%)	F2: Organising ideas in relation to own text (13.61%)
Low-level monitoring and revising	
F1: Low-level editing after writing (47.70%)	F1: Low-level editing after writing (61.26%)
F2: Low-level editing during writing (23.9%)	F2: Low-level editing during writing (14.43%)
High-level monitoring and revision	
F1: High-level editing after writing (42.92%)	F1: High-level editing after writing (51.36%)
F2: High-level editing during writing (24.35%)	F2: High-level editing during writing (13.32%)

As presented above, the results elicited by the GEPT live test data resembled very closely the results elicited by the real-life data in terms of the number of factors involved in each academic writing phase. However, 4 individual out of 48 items (i.e. Item 2.6, 4.4 of the conceptualisation phase, Item 2.9 of the meaning and discourse construction phase, and Item 3.3 of the organising phase) did not load on the same factor as they did on the real-life data. Only one item (i.e. Item 4.13 of the low-level monitoring and revising phase) was dropped from the analysis because it did not load on any factor at a level of 0.3 or above. It seems that the results elicited by the GEPT live test data resembled the results of the underlying structure of the real-life academic writing constructs even more closely than those elicited by the GEPT data collected under the research context in Phase 1 (e.g. a few more items were dropped in the analysis of the GEPT data in Phase 1).

The results reported here have provided strong evidence for the cognitive validity of the GEPT Advanced Writing Task 1 because the data was collected under live test condition, which is not always achievable in research. The findings are, to some extent, surprising given that Phase 1 was a repeated-measure design (i.e. the same participants to do both the GEPT and the real-life tasks). One possible explanation is that both the academic writing tasks under the real-life condition and the GEPT task under the live test condition were high-stakes

to the participants (either affecting their final GPA or their results of the GEPT) whereas the GEPT task did not affect the Phase 1 participants' course results in any way. This might be a reason why the processes elicited by the GEPT task under the live test condition resembled even more closely the processes that students would employ in a real life context than the same GEPT task carried out under a research condition.

6. Conclusion

6.1 Summary of the main findings

Overall task setting of the GEPT task

Based on the analysis of the expert panel judgement, it was found that the overall task setting of the GEPT test task resembled that of the real-life tasks that students would normally encounter in a UK academic context in a number of important ways. The GEPT task required an output of an essay, which was one of the most common genres required in the real-life academic context identified in this study. Language functions expected on the tasks were similar. Language functions, such as *cite sources*, *describe*, *reason*, *evaluate*, *express personal views*, and *synthesis*, were considered important for both real-life tasks. These core language functions were also expected in the GEPT task, with functions such as *express personal views*, *summarise*, *cite sources*, *reason* and *evaluate* regarded as most important on the GEPT task by the judges. Regarding the cognitive demands imposed on the writer, both real-life tasks were knowledge-transforming tasks which required the highest level of processes, such as planning rhetorical goals, integrating ideas from different sources and transforming ideas. The GEPT task was also mapped towards the higher end of the cognitive demands. It required the test takers to purposely select, organise and summarise relevant ideas from the multiple input sources and connect these ideas to their own text, as well as expressing personal viewpoints on the issues. In addition, the judges considered the clarity of writer-reader relationship expected on the GEPT and the real-life tasks was similar. Nevertheless, while the judges considered the provision of criteria on the GEPT was clearer, the criteria presented on the real-life tasks were more comprehensive. In terms of topic domains, the *academic* and *professional* domains were dominant on the real-life tasks whereas the GEPT task was considered to be in the *social* and, to a lesser extent, *academic* domains.

Features and difficulty level of the GEPT input texts

Based on the analysis of the expert panel judgement, the features of the GEPT input texts were comparable to those of the real-life input texts. Both real-life tasks required students to write based on multiple external reading resources. The GEPT task resembled the performance condition by requiring the test takers to write upon two passages. While the GEPT task studied in this study did not involve any non-verbal inputs, GEPT Advanced Writing Task 2 requires test takers to write upon a diagram or table. While the features of the GEPT input texts were largely comparable to the real-life input texts, the former included a narrower range on some parameters, such as discourse mode and genre. The discourse mode of the real-life input texts was dominantly expository and argumentative. The GEPT task required test takers to process solely argumentative texts. The real-life input texts contained a variety of genres, such as news articles, magazine articles, journal articles and book chapters whereas the GEPT task input texts were dominantly essays.

The difficulty level of the GEPT input texts were analysed in terms of *concreteness of the ideas*, *explicitness of the textual organisation*, *degree of cultural specificity*, *lexical*

complexity, syntactic complexity and degree of cohesion. The difficulty level between the GEPT input texts and the real-life input texts was comparable in terms of most lexical, syntactic and cohesion automated textual analysis indices investigated in the study. However, the GEPT input texts might be slightly easier than the real-life input texts because ideas on the GEPT input texts were considered as more concrete by the judges. The GEPT input texts were also more explicitly organised than the real-life input texts. It is worth noting that the GEPT input texts contained much denser low-frequency words, mostly proper nouns, than the real-life input texts. The GEPT input texts also required test takers to have more cultural specific knowledge than the real-life input texts did.

Cognitive processes elicited by the GEPT task

The results obtained from the two real-life academic writing tasks provided empirical evidence of the target cognitive processes, which we would expect to be measured in an academic writing test. The results generated from the GEPT test (Phase 1: research condition in the UK and Phase 2: live test condition in Taiwan) showed the extent to which GEPT Advanced Writing Task 1 activated these cognitive processes in the same manner as they were activated by the real-life academic writing tasks. The findings show that the hypothesised academic writing cognitive processes derived from the literature were largely supported by the statistical analyses of the real-life data, and that the GEPT task demonstrated good cognitive validity with respect to these processes.

Based upon the literature review, writers are likely to go through several cognitive phases when they write from external sources, though the phases can be overlapping or looping back. This study considered the following five phases to be most relevant to the discussion of integrated reading-into-writing tests for academic purposes: (1) conceptualisation, (2) meaning and discourse construction, (3) organising, (4) low-level monitoring and revising and (5) high-level monitoring and revising (Field, 2004, 2008, 2011, 2013; Kellogg, 1994; Shaw & Weir, 2007). Considering the constraints of the questionnaire, this study focused broadly on the phases which are more metacognitive (i.e. easier to be self-reported) and did not investigate phases such as execution and micro-planning. Exploratory factor analyses on these five hypothesised academic writing constructs on the real-life data showed that each construct involved two to three distinctive cognitive processes. In total, eleven cognitive validity parameters were identified: (1) task representation and macro-planning; (2) revising macro plan; (3) connecting and generating, (4) selecting relevant ideas and (5) global careful reading; (6) organising ideas in relation to the input texts; (7) organising ideas in relation to the writer's own text; (8) while writing low-level editing; (9) after writing low-level editing; (10) while writing high-level editing, and (11) after writing high-level editing.

Any writing test tasks which are cognitively valid should elicit from test takers the cognitive processes they would normally employ in non-test conditions. Having identified the target cognitive processes, the study investigated whether the high-achieving and low-achieving participants employed these processes differently on the real-life academic writing tasks. This analysis was to find out if these cognitive parameters could potentially distinguish the performances of stronger writers from those of weaker writers. Shaw & Weir (2007) argued that when identifying the cognitive parameters to be examined in a test, it is important to demonstrate 'how writers at different levels would employ these cognitive processes with 'educationally significant differences' (p.142). The results showed that the high achieving participants reported employing eight of the eleven cognitive processes (i.e. *task representation and macro-planning, careful global reading, selecting relevant ideas, connecting and generating, organising ideas in relation to source texts, organising in*

relation to new text, while writing low-level editing and while writing high-level editing) more than the low achieving group. Apart from *task representation and macro-planning* and *careful global reading*, all differences were statistically significant. Three process parameters including *revising macro plan*, *after writing low-level editing* and *after writing high-level editing* did not distinguish the high-achieving and low-achieving groups. The results further support the case for considering these eleven cognitive processes involved within the five academic writing phases as the target cognitive process parameters for a valid academic writing test.

The participants as a whole group reported employing all the eleven cognitive processes more on a scale from 1 (*definitely disagree*) to 4 (*definitely agree*) on the real-life tasks than on the GEPT task. The differences reported in six processes were significant ($p < 0.05$). The high-achieving and low-achieving participants in this study employed most of these cognitive processes similarly between the test and the real-life conditions whereas the medium group tended to employ some processes, such as *task representation and macro-planning*, *revising macro plan*, *low-level editing* and *high-level editing* processes, significantly more in the real-life than the test conditions. When compared to the high-achieving group, the middle-achieving group might not have been able to employ all processes with full automaticity as they were at the transitional stage of developing their academic writing ability. Due to limited cognitive capacity, they may have needed more time to complete the processes. While the medium group showed some differences, it is encouraging that all three groups of participants employed the *careful global reading*, *selecting relevant ideas*, *connecting and generating*, and *organising* processes to a similar extent on the GEPT and real-life tasks.

Exploratory factor analyses on the GEPT data (Phase 1 and Phase 2) show that although a few individual items yielded complex loadings, the GEPT task was largely able to elicit from the participants the same underlying cognitive processes as the real-life tasks did. For the Phase 1 GEPT data, the underlying structures of four out of the five academic writing phases, elicited on the GEPT and the real-life tasks, were identical in terms of the number of factors extracted. For the Phase 2 GEPT live test data, the underlying structures of all five phases were identical in terms of the number of factors extracted. The eleven individual factors yielded by the GEPT live test data resembled closely those factors yielded by the real-life data. Only two items from the conceptualisation phase, one from the meaning and discourse construction phase, and one from the organising phase did not load on the same factor as they did on the real-life data. One item of the low-level monitoring and revising phase was dropped from the analysis because it did not load on any factor at a level of 0.3 or above. It was encouraging to learn that, based on self-report data, the cognitive processes that students would normally employ on real-life academic writing tasks were elicited by the GEPT task. The results that some individual processes did not load on the corresponding factors indicate that the test conditions have elicited a slightly different pattern of interaction among the individual sub-processes. It is clear that the research tool, i.e. questionnaire, used in this research cannot establish the exact nature of such interaction. Future studies should employ other tools such as think-aloud protocols and keystroke logging, to reveal how these processes and sub-processes are employed in finer detail.

In short, the results of this study provided strong evidence of GEPT Advanced Task 1, eliciting important real-life processes such as *task representation*, *selecting relevant ideas*, *careful global reading*, and *organising ideas in relation to source texts* thereby supporting claims for its cognitive validity. Such real life processes might be largely absent from a writing only test.

6.2 Limitations of the study

Participants of the investigation of cognitive validity

A total of 160 participants participated in Phase 1 of the study. Their IELTS reading and writing bands ranged from 5.0 to 7.5. While the proficiency level of the participants was considered appropriate for the context of this study, future studies are advised to include more participants at higher proficiency levels. In addition, test takers' background was homogenous in the sense that all test takers were from the Business School in the UK academic context. A comparatively homogenous profile of background knowledge was suitable for the cognitive investigation in this study. Nevertheless, future studies are advised to investigate the cognitive processes and test results of participants from a variety of disciplinary backgrounds.

Real-life academic tasks and test tasks

Due to the limited scope of the study, the sample size of tasks investigated was small. Two real-life academic writing tasks were selected based on a range of carefully chosen criteria (see Section 4.1). For the investigation of the overall task setting, one version of the GEPT and real-life tasks was analysed. For the input text features, the ten most cited input texts of each real-life academic writing task and twenty input texts from ten testlets of the GEPT task were analysed. Although a larger sample size would have been desirable, the real-life tasks and the corresponding input texts in this study were carefully sampled to ensure the generalizability of the findings.

Writing process questionnaire

Think-aloud protocol has been regarded as the most persuasive way of demonstrating the processes employed (for example see Hayes & Flower, 1983; Spivey, 1997; Plakans, 2010). However, as think-aloud is a very time-consuming method, it is usually used in studies with a small number of participants. Such a method was, therefore, not suitable for the context of this study which involved a large number of L2 participants in both real-life academic and test conditions. This study investigated the writing processes that participants employed on the real-life academic writing tasks and the GEPT tasks by a Writing Process Questionnaire (adapted from Chan, 2013). Following the recommendations of Purpura (1998) on the use of cognitive process questionnaires, the construct of the questionnaire was carefully developed based upon human information processing theory. In addition, the psychometric characteristics of the questionnaire and the underlying construct validity of the questionnaire were verified by a series of statistical analyses (for details of the development of the questionnaire, see Chan 2013, Chapter 3). The questionnaire in this study was constructed while paying particular attention to all the processes which seemed to be most relevant to our discussion of the cognitive validity of writing tests for academic purposes. The questionnaire, however well developed, can only seek evidence of the participants' perceptions of what they did. One should not rely upon these perceptions as evidence of actual performance. Future studies should attempt to triangulate the questionnaire data by other instruments, such as think-aloud protocol, post-test interview and/or keystroke logging.

6.3 Implications of the findings

A more complete construct definition of GEPT Advanced Writing Task 1

This study was a follow-up study of Weir et al (2013) which demonstrated the a posteriori criterion-related validity of the GEPT Advanced Reading and Writing. Their findings showed moderate to strong positive correlations between the GEPT Advanced Reading and Writing Test and IELTS. In addition, the GEPT Advanced Reading and Writing Test scores accounted for 27.89% variance of real-life academic performance at a correlation of 0.529

($p < 0.01$). This study investigated the a priori context validity and cognitive validity of GEPT Advanced Writing Task 1. The a priori context and cognitive validity are arguably the most important components of a test to establish during the test development and validation phases. Khalifa & Weir (2009: 81) argued that 'the contextual parameters operationalised in a test should mirror the criterial features of the target situation activity as far as possible'. In addition, any valid tests have to demonstrate the extent to which they elicit test takers' cognitive processes that correspond to the processes that are elicited by real-life tasks in the target language context (Glaser, 1991; Shaw & Weir, 2007). A major threat to the cognitive validity is that the tasks might tap into a skill which is solely used under test conditions and demonstrate little relation to the real-life processes (Field, 2013; Shaw & Weir, 2007, Chapter Three; Weir et al, 2013, Chapter Three).

As summarised above, GEPT Advanced Writing Task 1, as a type of integrated reading-into-writing test, demonstrates both context and cognitive validity. The results of this study have an important implication for university admissions officers and other stakeholders that GEPT Advanced is a valid option of writing tests for academic purposes and they could be more certain of their students' ability to cope with academic writing. In addition, integrated reading-into-writing tasks as against independent reading tasks and independent writing tasks should play a more important role in academic language assessments than they do at the time being.

Implications for test writers to develop more valid reading-into-writing test tasks for academic purposes

Through a carefully demonstrated link between the test and the real-life conditions, the results of this study strongly suggested that GEPT Advanced Writing Task 1 is a valid tool to assess academic writing ability in terms of both context validity and cognitive validity. To assist test writers to develop more valid reading-into-writing test tasks for academic purposes, recommendations for overall test setting and input text manipulation are provided below.

Overall task setting

- (1) *Incorporating other common academic writing genres, such as report.* Essay is only one of the common genres that students are most frequently expected to produce in the real-life academic context. The task survey in this study showed that report is another most frequently assigned genre.
- (2) *Revisiting topics in the social domain.* Based on the judges' response, topic domains identified in the real-life tasks were academic and professional. However, judges considered that GEPT task sampled in this study was in the *social*, and to a lesser extent, *academic*. This is, however, not a straightforward issue. Item writers would have to consider the facts that 1) topics in the social domain are not entirely appropriate for the context of GEPT Advanced, which is used mainly for academic purposes, but at the same time 2) topics used should not include topics which involve content at a high level of specific (academic) knowledge.
- (3) *Incorporating a wider range of language functions* GEPT Advanced Task 1 sampled in this study required fewer language functions than the real-life tasks did. It is important to cover these language functions which have been identified in real-life academic writing in the test specification, even though not all functions need to be tested in every single testlet. (As mentioned earlier, some language functions, such as *describe* and *synthesise*, are targeted by Task 2 of the GEPT Advanced Writing Test).

Input text manipulation

Apart from overall task setting, test writers need to be aware of the possible effects of their manipulating input texts for test purposes.

- (4) *Incorporating more input genres and a combination of argumentative texts and expository texts.* Real-life tasks incorporated a range of input genres and a combination of argumentative texts and expository texts. However, due to the need of standardisation, the range of the input genres and types of texts of test tasks are often limited. All the GEPT input texts from ten testlets sampled in this study were regarded as belonging to a simplified version of the argumentative essay genre. It is important for test developers to monitor if a sufficient range of the input genres and text types, as stated in the test specifications, is represented across the testlets. In addition to this reading-into-writing task, GEPT Advanced Test contains a separate reading paper. A broader range of input genres including expository have been adopted in the reading paper (for details see Wu, 2013).
- (5) *Reducing the lexical complexity.* The lexical complexity of the GEPT input texts was seemingly more demanding than the real-life input texts. This is likely to be the results of input text manipulation of incorporating sufficient 'idea units' into the input texts with a usually tight word limit. The GEPT input texts had a greater density of *low frequency words*, mostly proper nouns, than the real-life input texts. While standardising the text length of the input texts, test developers have to monitor if lexical complexity of the input texts is increased unnecessarily due to the process of text manipulation.
- (6) *Maintaining a lower syntactic complexity.* Generally speaking, the results suggested that it was less demanding to build the textual representation of the GEPT input texts than real-life-input texts because the test task input texts were more explicitly organised than the real-life input texts. It might be helpful to organise the GEPT input texts less explicitly. Nevertheless, it is advisable to maintain a similar level of syntactic complexity and degree of coherence as reported in this study, so that test takers would be able to perform the processes under the test conditions with greater time restrictions than the real-life conditions.

In short, this study demonstrates that GEPT Advanced Task 1 is a suitable instrument for measuring the writing ability of university students in terms of both context and cognitive validity but as always iterative improvement is possible and desirable.

Reference

- Ackerman, J. M. (1990). Reading, writing and knowing: The role of disciplinary knowledge in comprehension and composing. *Technical Report*, 40. Berkley: Center for the Study of Writing at the University of California. Berkerley and Carnegie Mellon University.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bereiter, C., & Scardamalia, M. (1987). *The Psychology of Written Composition*. Hillside, NJ: Lawrence Erlbaum Associates.
- Brewer, W. F. (1980). Literary theory, rhetoric, and stylistics: Implications for psychology. In R. J. Spiro, B. C. Bruce, & Brewer, W. F (Eds.), *Theoretical Models and Processes of Reading Comprehension* (pp. 221–239). Hillsdale, NJ: Erlbaum.
- Bridgeman, B., & Carlson, S. (1983). *Survey of Academic Writing Tasks Required of Graduate and Undergraduate Foreign Students*. Princeton, NJ: Educational Testing Service.
- Brown, G., & Yule, G. (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Carson, J. (2001). A task analysis of reading and writing in academic contexts. In A. Hirvela (Ed.), *Linking Literacies: Perspectives on L2 Reading-Writing Connections* (pp. 48-83). Ann Arbor, MI: University of Michigan Press.
- Chan, S. H. C. (2013). *Establishing the validity of reading-into-writing test tasks for the UK academic context*. Unpublished PhD thesis. University of Bedfordshire.
- Clapham, C. (1996). *The development of IELTS: A study of the effect of background on reading comprehension*, *Studies in Language testing 4. Studies in Language Testing*. Cambridge: UCLES/Cambridge University Press.
- Cobb, T. (2003). VocabProfile, The Compleat Lexical Tutor. Retrieved from <http://www.lextotor.ca>
- Cooper, A., & Bikowski, D. (2007). Writing at the graduate level: What tasks do professors actually require? *Journal of English for Academic Purposes*, 6, 206–221.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)*. Strasbourg: Council of Europe.
- Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 984–989). Austin, TX: Cognitive Science Society.
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The role of cohesion, readability, and lexical difficulty. *Journal of Research in Reading*, 35(2), 115–135.

- Crossley, S. A., Louwerse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, 91(2), 15–30.
- Crossley, S., Greenfield, J., & McNamara, D. (2008). Assessing text readability using cognitively based indices. *Tesol Quarterly*, 42(3), 475–493.
- Douglas, D. (2000). *Assessing Language for Specific purposes: Theory and Practice*. Cambridge: Cambridge University Press.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. and Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3) 272-299.
- Field, J. (2004). *Psycholinguistics: the Key Concepts*. London: Routledge.
- Field, J. (2008). *Listening in the Language Classroom*. Cambridge: Cambridge University Press.
- Field, J. (2011). Cognitive validity. In L. Taylor (Ed.), *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, *Studies in Language Testing* 30 (pp. 65–111). Cambridge: UCLES/Cambridge University Press.
- Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining Listening: Research and Practice in Assessing Second Language Listening*, *Studies in Language Testing* 35 (pp. 77–151). Cambridge: UCLES/Cambridge University Press.
- Flower, L. (1990). Reading-to-write: Understanding the task. In L. Flower, V. Stein, J. Ackerman, M. J. Kantz, K. McCormick, & W. C. Peck (Eds.), *Reading to write. Exploring a cognitive & social process* (pp. 35–73). New York: Oxford University Press.
- Flower, L., Stein, V., Ackerman, J., Kantz, M. J., McCormick, K., & Peck, W. C. (1990). *Reading-to-write: Exploring a Cognitive and Social Process*. New York: Oxford University Press.
- Galbraith, D., & Torrance, M. (1999). *From Problem Solving to Text Production*. *Studies in Writing*. Amsterdam: Amsterdam University Press.
- Geranpayeh, A., & Taylor, L. (Eds.). (2013). *Examining Listening: Research and practice in assessing second language listening*, *Studies in Language Testing* 35. Cambridge: UCLES/Cambridge University Press.
- Glaser, R. (1991). Expertise and assessment. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and Cognition* (pp. 17–30). Englewood Cliffs: Prentice Hall.
- Grabe, W., & Kaplan, F. L. (1996). *Theory and Practice of Writing: An applied linguistic perspective*. London: Longman.
- Graesser, A. C., McNamara, D., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193–202.
- Green, A. (2012). *Theoretical and Empirical Bases for Language Construct Definition across the Ability Range*. Cambridge: Cambridge University Press.
- Green, A., Unaldi, A., & Weir, C. J. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing*, 27(2), 191–211.
- Green, A., Weir, C. J., Chan, S. H. C., Field, J., Taylor, L., Bax, S., & Nakatsuhara, F. (2012). *Report to Cambridge ESOL UK on the Contextual Parameters of CAE Examinations*. Unpublished research report.

- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A Study of Writing Tasks Assigned in Academic Degree Program. TOEFL Research Reports, RR-95-44*, Princeton, NJ: Educational Testing Service.
- Hayes, J. R. (1996). A new Framework for understanding cognition and affect in writing. In S. Ransdell (Ed.), (pp. 1–28). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Hayes, J. R., & Flower, L. (1980). Identifying the organisation of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive Processes in Writing* (pp. 3–30). Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Hayes, J. R., & Flower, L. (1983). Uncovering Cognitive Processes in Writing: an Introduction to Protocol Analysis. In P. Mosenthal, L. Tamor, & S. Walmsley (Eds.), *Research on Writing: Principles and Methods* (pp. 207–220). New York: Longman.
- Horowitz, D. M. (1986a). What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20, 445–460.
- Horowitz, D. M. (1986b). Essay examination prompts and the teaching of academic teaching. *English for Specific Purposes*, 5, 107–120.
- Horowitz, D. M. (1991). ESL writing assessments: Contradictions and resolutions. In Liz Hamp-Lyons (Ed.), *Assessing Second Language Writing in Academic Contexts* (pp. 71–86). Norwood, NJ: Ablex.
- Johns, A. M. (1993). Reading and writing tasks in English for academic purposes classes: Products, processes and resources. In J. G. Carson & I. Leki (Eds.), *Reading in the composition classroom* (pp. 274–289). New York: Newbury House.
- Kellogg, R. T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition*, 15(3), 256–266.
- Kellogg, R. T. (1996). A model working memory in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications*. (pp. 57–71). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Kellogg, R. T. (1999). Components of working memory in writing. In M. Torrance & G. Jeffery (Eds.), *The cognitive demands of writing: Processing capacity and working memory effects in text production* (pp. 43–61). Amsterdam: Amsterdam University Press.
- Kellogg, R. T. (2001). Competition for working memory among writing processes. *American Journal of Psychology*, 114(2), 175–192.
- Kennedy, C., & Thorp, D. (2007). A corpus-based investigation of linguistic responses to an IELTS Academic Writing task. In L. Taylor (Ed.), *IELTS Collected Papers; Research in Speaking and Writing Assessment* (pp. 316–377). Cambridge: Cambridge University Press.
- Khalifa, H., & Weir, C. J. (2009). *Examining Reading: Research and Practice in Assessing Second Language Reading, Studies in Language Testing 29*. Cambridge: UCLES/Cambridge University Press.
- Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28, 81–101.
- Levelt, W. J. M. (1989). *Speaking. From Intention on Articulation*. Cambridge, MA-London: ACL-MIT Press.

- LTTC. (2013). About GEPT. Retrieved from http://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT.htm
- Mathison, M. A., & Spivey, N. N. (1993). Writing from academic sources: Authorship in writing the critique. Berkeley, CA.: National Center for the Study of Writing and Literacy.
- 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*, 2, paper 4. Canberra, IDP: IELTS Australia.
- Moore, T., & Morton, J. (2005). Dimensions of difference: a comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4, 43-66.
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13, 111–129.
- Plakans, L. (2009a). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, 8, 252-266.
- Plakans, L. (2009b). Discourse synthesis in integrated second language writing assessment. *Language Testing*, 26(4), 561–587.
- Plakans, L. (2010). Independent vs . Integrated Writing Tasks: A Comparison of Task Representation. *Tesol Quarterly*, 44(1), 185–194.
- Purpura, J E. (1998). The development and construct validation of an instrument designed to investigate selected cognitive background characteristics of test-takers. In A. J. Kunnan (Ed.), *Validation in Language Assessment* (pp. 111–140). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Purves, A. C., Soter, A., Takala, S., & Vahapassi, A. (1984). Towards a domain-referenced system for classifying assignments. *Research in the Teaching of English*, 18(4), 385–416.
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics, Volume 2: Reading, Writing and Language Learning*. Cambridge: Cambridge University Press.
- Scardamalia, M., & Bereiter, C. (1991). Literate expertise. In K. A. Ericsson & J. Smith (Eds.), *Toward a General Theory of Expertise: Prospects and Limits* (pp. 172–194). Cambridge: Cambridge University Press.
- Shaw, S., & Weir, C. J. (2007). *Examining Writing: Research and Practice in Assessing Second Language Writing*, *Studies in Language Testing* 26. Cambridge: UCLES/Cambridge University Press.
- Spivey, N. N. (1984). *Discourse Synthesis: Constructing Texts in Reading and Writing. Outstanding Dissertation Monograph*. Newark, DE: International Reading Association.
- Spivey, N. N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication*, 7(2), 256–287.
- Spivey, N. N. (1991). The shaping of meaning: options in writing the comparison. *Research in the Teaching of English*, 25, 390–418.
- Spivey, N. N. (1997). *The Constructivist Metaphor: Reading, Writing and the Making of Meaning*. San Diego, CA: Academic Press.

- Spivey, N. N. (2001). Discourse synthesis: Process and product. *Discourse synthesis: Studies in historical and contemporary social epistemology* (pp. 379–396). Westport, CT: Praeger.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7–26.
- Taylor, L. (Ed). (2011). *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, *Studies in Language Testing* 30. Cambridge: UCLES/Cambridge University Press.
- UKCISA. (2012). Annual report 2011-2012. Retrieved from http://www.ukcisa.org.uk/files/pdf/about/annual_review.pdf
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9(9), 27–55.
- Weir, C J, & Wu, J. R. W. (2006). Establishing test form and individual task comparability: a case study of a semi-direct speaking test. *Language Testing*, 23(2), 167–197.
- Weir, C J, Chan, S. H. C., & Nakatsuhara, F. (2013). Examining the Criterion-related Validity of the GEPT Advanced Reading and Writing Tests: Comparing GEPT with IELTS and Real-life Academic Performance. *LTTC-GEPT Research Reports*, RG-01, 1–43.
- Weir, C J. (1983). *Identifying the Language Problems of the Overseas Students in Tertiary Education in the United Kingdom*. Unpublished PhD dissertation: University of London.
- Weir, C J. (1990). *Communicative language testing*. Englewood Cliffs, NJ: Prentice Hall.
- Weir, C J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.
- Weir, C J. (2013). Measured constructs: A history of Cambridge English language examinations 1913-2012. *IELTS Research Report*, 51, 2–10.
- Weir, C. J. (2012). *TEAP Writing Project: A Review of the Washback Concept in Language Education with Particular Implications for Japan and the EIKEN Foundation of Japan Organisation*. Unpublished Research Report.
- Wu, R. Y. F. (2012). *Establishing the validity of the General English Proficiency Test Reading Component through a critical evaluation on alignment with the Common European Framework of Reference*. Unpublished PhD dissertation: University of Bedfordshire.
- Yu, G. and Lin, S.-W. (forthcoming). A comparability study on the cognitive processes of taking GEPT (Advanced) and IELTS (Academic) Writing tasks using graph prompts. *LTTC-GEPT Research Reports*.

Appendix

Appendix 1: Glossary of the Contextual Parameter Proforma

Part 1 – Overall task setting

1. Purpose - Is the communicative purpose for completing the task clearly presented?
2. Topic Domain - What is the domain of the topic?
 - Personal - relates to personal lives (e.g. family, relatives, friends, etc.).
 - Social - relates to issues concerning the members of the public.
 - Professional - relates to expert and specialised knowledge of a profession.
 - Academic - relates to a particular discipline or field of study (which may have no practical purpose or use).
3. Genre - What is the genre of the text to be produced?
 - Essay is a piece of writing which is often written from an author's personal point of view.
 - Report is an informational piece of work made with the specific intention of relaying information or recounting certain events.
 - Case study is an intensive analysis of a person, group, or event in a specific context.
 - Summary is a short document that summarises a longer report or proposal or a group of related reports, in such a way that readers can rapidly become acquainted with a large body of material without having to read it all.
4. Cognitive demands - Which level of cognitive demands does the task impose on the candidates/students? (Think of the minimum requirement to complete the task).
 - Telling/retelling content: the text production is primarily guided by a direct retrieval of content from long-term memory or a direct copy from the input texts.
 - Organising/reorganising content: the text production requires writers to purposefully organise the content they retrieved from long-term memory and/or selected from the input texts in order to solve the rhetorical problems of the writing task.
 - Transforming content: the text production requires writers to establish a high awareness of the rhetorical situation of the writing task. Writers are required to strategically organise as well as transform (e.g. synthesise, interpret, evaluate) the content they retrieved from long-term memory and/or selected from the input texts to fulfil the writing goals.
5. Language functions - What language functions do the candidates/students have to demonstrate?
6. Intended reader - Is the intended reader clearly presented?
7. Knowledge of criteria - Are the marking criteria clearly presented?

Part 2 – Input text features

8. Input format – What is the format of the input?
9. Verbal input genre – What is the genre of the input text?
10. Non-verbal input – What is the non-verbal input provided in the input text?
11. Discourse mode – What is the primary discourse mode of the input text?
 - Narrative texts recount an event or a series of related events.
 - Descriptive texts describe a person, place or thing using sensory details.
 - Expository texts give information about or an explanation of an issue, subject, method or idea.
 - Argumentative texts typically involve a course of reasoning.
12. Concreteness of ideas – How concrete or abstract is the content of the input text?
13. Explicitness of textual organisation – How explicit or implicit is the textual organisation of the input text?
14. Cultural specificity – How culturally neutral or specific is the content of the input text?

Appendix 2: Feedback Evaluation Questionnaire

Based on the experience applying the Contextual Parameter Proforma, how confident do you feel when you choose your response? Please tick 1, 2, 3 or 4 to indicate how confident you were. If your answer is 2 or 1, please specify the reason.

4 = very confident

3 = confident

2 = not confident

1 = not confident at all

	4	3	2	1	Reasons
Part 1 - Overall task setting					
1. Purpose					
2. Topic domain					
3. Genre					
4. Cognitive demands					
5. Language functions					
6. Intended reader					
7. Knowledge of criteria					

	4	3	2	1	Reasons
Part 2 - Input text features					
8. Input format					
9. Verbal input genre					
10. Non-verbal input					
11. Discourse mode					
12. Concreteness of ideas					
13. Explicitness of textual organisation					
14. Cultural specificity					

Feedback from judges

	Mean	Standard Deviation
Part 1 (No. of judges: 16)		
Purpose	3.81	0.40
Topic domain	2.75	0.68
Genre	3.86	0.34
Cognitive demands	3.19	0.54
Language functions	3.25	0.68
Clarity of intended reader	3.44	0.51
Clarity of knowledge of criteria	3.68	0.79
Part 2 (No. of judges: 2)		
Input format	4.00	0.00
Verbal input genre	4.00	0.00
Non-verbal input	4.00	0.00
Discourse mode	3.50	0.71
Concreteness of ideas	3.50	0.71
Explicitness of textual organisation	3.00	0.00
Cultural specificity	3.50	0.71

Appendix 3: Writing Process Questionnaire

全民英語能力分級檢定測驗高級寫作問卷調查 LTTC-CRELLA COLLABORATION PROJECT TEST-TASK WRITING PROCESS QUESTIONNAIRE

複試座位號碼：_____

謝謝您參與今天的測驗，請回答下列問題。本問卷調查結果僅供本中心研究參考，不影響測驗成績。請據實填寫，謝謝配合。為感謝您的填寫，本中心將致贈全民英檢小筆記本一本。

一、個人資料

Section 1: Personal data

1. IELTS results (if any): Overall band: _____ Reading: _____ Writing: _____

二、作答寫作過程

Section 2: Test-task writing processes

請您根據作答寫作第一部份的實際狀況，依以下階段將適合的選項塗黑、塗滿或書寫。

In this section, there are some statements about how you might complete Task 1. Please answer all the questions, thinking about what you did

- 第一階段：閱讀作答說明時 while reading the task prompt
- 第二階段：閱讀試題文章時 while reading the source texts
- 第三階段：開始寫作前 before starting to write
- 第四階段：寫作初稿時 while writing the first draft
- 第五階段：完成初稿後 after writing the first draft

Please choose the number which reflects the extent of your agreement or disagreement to each statement below, using the following 4-point scale:

4分：完全同意 Definitely agree

3分：大致同意 Mostly agree

2分：大致不同意 Mostly disagree

1分：完全不同意 Definitely disagree

第一階段：閱讀作答說明時 1. Reading task prompt for the test (i.e. instructions)		Definitely Agree	Mostly Agree	Mostly Disagree	Definitely Disagree
請根據您在閱讀作答說明時的實際狀況，回答以下問題。 Please think about what you did while you were reading the task prompt.					
1.1	我仔細閱讀作答說明，逐字了解其中的意思。 I read the task prompt (i.e. instructions) carefully to understand each word in it.	4	3	2	1
1.2	我思考要寫什麼內容，以使作文切題，符合試題要求。 I thought about what I might need to write to make my essay relevant and adequate to the task.	4	3	2	1
1.3	我考慮我的作文應如何符合目標讀者的期望。 I thought about how my essay would suit the expectations of the intended reader .	4	3	2	1
1.4	我能夠充分理解作答說明。 I was able to understand the instructions for this writing test very well .	4	3	2	1
1.5	我考慮這篇作文要達到的目的。 I thought about the purpose of the task.	4	3	2	1
除了以上所列，閱讀作答說明時您還做了什麼，請列舉： What else did you do while reading the prompt?					

第二階段：閱讀試題文章時 2. Reading the source texts (i.e. the two articles)		Definitely Agree	Mostly Agree	Mostly Disagree	Definitely Disagree
請根據您在閱讀試題文章時的實際狀況，回答以下問題。 Please think about what you did while you were reading the source texts.					
2.1	我仔細閱讀每篇文章的全文。 I read through the whole of each source text carefully.	4	3	2	1
2.2	我閱讀每篇文章的全文不只一次。 I read the whole of each source text more than once.	4	3	2	1
2.3	我運用對文章結構的認識，去找出文章重點。 I used my knowledge of how texts like these are organized to find parts to focus on.	4	3	2	1
2.4	我快速搜尋文章中答題的線索。 I searched quickly for part(s) of the texts which might answer the question.	4	3	2	1
2.5	我仔細閱讀文章中有助於答題的部份。 I read some relevant part(s) of the texts carefully.	4	3	2	1
2.6	我重新閱讀作答說明。 I read the task prompt again.	4	3	2	1
2.7	我在試題文章的重點處做筆記或畫底線。 I took notes on or underlined the important ideas in the source texts.	4	3	2	1
2.8	我將試題文章的重點，在腦海中依重要性排序。 I prioritised the important ideas in the source texts in my mind.	4	3	2	1
2.9	我將試題文章的重點，與我既有的知識連結。 I linked the important ideas in the source texts to what I know already.	4	3	2	1
2.10	我找出單篇文章內各重點的關聯。 I worked out how the main ideas in each source text relate to each other.	4	3	2	1
2.11	我找出兩篇文章之間重點的關聯。 I worked out how the main ideas across the source texts relate to each other.	4	3	2	1
2.12	我產生新的觀點，或對既有的知識有進一步了解。 I developed new ideas or a better understanding of existing knowledge.	4	3	2	1
2.13	我修改我寫作的規畫（例如內容或結構）。 I changed my writing plan (e.g. structure, content etc)	4	3	2	1
除了以上所列，閱讀試題文章時您還做了什麼，請列舉： What else did you do while reading the source texts?					

第三階段：開始寫作前 3. Before starting to write		Definitely Agree	Mostly Agree	Mostly Disagree	Definitely Disagree
請根據您在開始寫作前的實際狀況，回答以下問題。 Please think about what you did before starting to write your essay.					
3.1	我將要寫的論點加以組織。 I organized the ideas I planned to include in my essay.	4	3	2	1
3.2	我重新結合或排序要寫的論點，以配合作文的結構。 I recombined or reordered the ideas to fit the structure of my essay.	4	3	2	1
3.3	組織論點時，我刪除某些論點。 I removed some ideas I planned to write.	4	3	2	1
3.4	我嘗試採用其中一篇試題文章的結構。 I tried to use the same organizational structure as in one of the source texts.	4	3	2	1
除了以上所列，開始寫作前您還做了什麼，請列舉： What else did you do before writing?					

第四及第五階段：寫作初稿時及完成初稿後 4. While writing the 1 st draft and 5. after writing the 1 st draft		While writing the 1 st draft							
請根據您在寫作初稿時及完成初稿後的實際狀況，回答以下問題。 Please think about what you did while you were writing the first draft of your essay and after you had finished the first draft.		Disagree Disagree	Mostly Disagree	Mostly Agree	Agree Agree				
4.1	我有時暫停寫作，以便組織作文的論點。 While I was writing, I sometimes paused to organize my ideas.	4	3	2	1				
4.2	在寫作過程中，我產生新的觀點。 I developed new ideas while I was writing.	4	3	2	1				
4.3	在寫作過程中，我找出更多兩篇試題文章之間的關聯。 While I was writing, I worked out further connections among the ideas across the source texts (in addition to the connections I made before starting to write).	4	3	2	1				
4.4	我重新閱讀作答說明。 I re-read the task prompt.	4	3	2	1				
4.5	我有選擇地重新閱讀兩篇試題文章。 I selectively reread the source texts.	4	3	2	1				
4.6	在寫作過程中，我修改我寫作的規劃（例如內容或結構）。 I changed my writing plan (e.g. structure, content etc).	4	3	2	1				
4.7	我檢查作文內容是否切題。 I checked that the content was relevant.	4	3	2	1				
4.8	我檢查作文是否結構嚴謹。 I checked that the essay was well-organized	4	3	2	1				
4.9	我檢查作文是否連貫通順（例如主題句和連接詞的運用）。 I checked that the essay was coherent, e.g. appropriate use of topic sentences, connectives etc.	4	3	2	1				
4.10	我檢查作文是否涵蓋兩篇試題文章的主要論點。 I checked that I included all appropriate main ideas from all the source texts.	4	3	2	1				
4.11	我檢查作文是否表達出個人觀點。 I checked that I included my own viewpoint on the topic.	4	3	2	1				
4.12	我檢查作文是否適當引用他人觀點（例如引用是否切題、是否合乎文法）。 I checked that the quotations were properly made, e.g. the quotes were relevant, the quotes were integrated grammatically into the essay, etc.	4	3	2	1				
4.13	我檢查是否用自己的文字表達試題文章的論點。 I checked that I had put the ideas of the source texts into my own words.	4	3	2	1				
4.14	我評估作文對目標讀者可能產生的影響。 I checked the possible effect of my writing on the intended reader.	4	3	2	1				
4.15	我檢查句子的文法是否正確、句構是否富變化。 I checked the grammatical accuracy and range of the sentence structures.	4	3	2	1				
4.16	我檢查詞彙的拼字和用法是否正確、用字是否富變化。 I checked the spelling, usage and range of the vocabulary.	4	3	2	1				
除了以上所列，寫作初稿時或完成初稿後您還做了什麼，請列舉： What else did you do while writing the first draft or after writing the first draft?									

問卷到此結束，謝謝您的協助！

Appendix 4: Comparison of the processes elicited by the two real-life tasks

		Report		Essay		Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
		Mean	Std Dev	Mean	Std Dev				
1.1	I read the whole task prompt carefully.	3.41	.779	3.43	.527	2408.000	4893.000	-.669	.504
1.2	I thought of what I might need to write to make my text relevant and adequate to the task.	3.51	.669	3.34	.634	2158.000	4643.000	-1.796	.072
1.3	I thought of how my text would suit the expectations of the intended reader.	3.18	.770	3.04	.751	2282.000	4767.000	-1.195	.232
1.4	I was able to understand the instructions for this writing task very well.	3.25	.703	3.03	.636	2091.500	4576.500	-2.105	.035
1.5	After reading the prompt, I thought about the purpose of the task.	3.10	.869	3.14	.728	2550.000	5251.000	-.022	.982
2.1	I read through the whole of each source text carefully.	2.95	.797	2.93	.729	2488.000	4973.000	-.299	.765
2.2	I read the whole of each source text more than once.	3.01	.858	2.89	.877	2369.000	4854.000	-.795	.427
2.3	I used my knowledge of how texts like these are organised to find parts to focus on.	3.16	.727	3.01	.648	2228.500	4713.500	-1.459	.145
2.4	I searched quickly for part(s) of the texts which might help complete the task.	3.14	.787	2.97	.680	2195.000	4680.000	-1.581	.114
2.5	I read some relevant part(s) of the texts carefully.	3.45	.688	3.27	.658	2133.500	4618.500	-1.894	.058
2.6	I read the task prompt again.	3.23	.842	3.43	.579	2320.500	5021.500	-1.044	.297
2.7	I took notes on or underlined the important ideas in the source texts.	3.27	.786	3.30	.768	2522.500	5223.500	-.144	.886
2.8	I prioritised important ideas in the source texts in my mind.	2.85	.908	2.96	.669	2445.000	5146.000	-.479	.632
2.9	I linked the important ideas in the source texts to what I know already.	3.18	.887	3.07	.666	2224.500	4709.500	-1.448	.147
2.10	I worked out how the main ideas in each source text relate to each other.	3.07	.805	3.06	.759	2513.500	4998.500	-.180	.857
2.11	I worked out how the main ideas across the source texts relate to each other.	3.05	.880	3.06	.740	2487.000	4972.000	-.293	.769
2.12	I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.	3.19	.811	3.11	.733	2358.500	4843.500	-.861	.389
2.13	I changed my writing plan while reading the source texts.	2.86	.855	2.83	.900	2523.000	5008.000	-.137	.891
3.1	I organised the ideas for my text before starting to write.	3.41	.723	3.30	.709	2312.500	4797.500	-1.079	.281
3.2	I recombined or reordered the ideas to fit the structure of my essay.	3.16	.782	3.17	.659	2518.000	5003.000	-.163	.871
3.3	I removed some ideas I planned to write.	2.99	.858	3.07	.767	2467.500	5168.500	-.382	.702
3.4	I tried to use the same organizational structure as in the source texts.	2.86	.887	2.87	.833	2553.500	5254.500	-.006	.995
4.1	While I was writing I sometimes paused to organize my ideas.	2.86	.751	3.16	.673	2040.500	4741.500	-2.292	.022
4.2	I developed new ideas	3.25	.703	3.19	.804	2483.500	4968.500	-.313	.754
4.3	I made further connections across the source texts.	2.90	.802	2.93	.840	2483.000	5184.000	-.314	.753
4.4	I re-read the task prompt.	3.22	.804	3.23	.820	2528.000	5229.000	-.118	.906
4.5	I selectively re-read the source texts.	3.05	.797	3.20	.773	2307.000	5008.000	-1.070	.285
4.6	I changed my writing plan (e.g. structure, content etc)	2.79	.781	2.60	.907	2271.500	4756.500	-1.222	.222
4.7	I checked that the content was relevant.	3.34	.692	3.30	.622	2419.000	4904.000	-.611	.541
4.8	I checked that my text was well-organised.	3.23	.773	3.09	.737	2248.500	4733.500	-1.344	.179
4.9	I checked that my text was coherent.	3.32	.762	3.13	.721	2169.000	4654.000	-1.686	.092
4.10	I checked that I included all appropriate main ideas from all the source texts.	3.19	.776	3.06	.849	2342.000	4827.000	-.934	.350
4.11	I checked that I included my own viewpoint on the topic.	3.22	.750	3.30	.622	2471.500	5172.500	-.377	.706
4.12	I checked that the quotations were properly made.	3.18	.788	3.24	.859	2385.500	5086.500	-.742	.458
4.13	I checked that I had put the ideas of the source texts into my own words.	3.26	.800	3.40	.646	2368.000	5069.000	-.832	.405
4.14	I checked the possible effect of my writing on the intended reader.	2.75	.846	3.06	.814	2059.500	4760.500	-2.162	.031
4.15	I checked the grammatical accuracy and range of the sentence structures.	2.86	.855	3.06	.740	2263.000	4964.000	-1.290	.197
4.16	I checked the spelling, usage and range of vocabulary.	2.99	.842	3.06	.814	2453.500	5154.500	-.439	.661
5.7	After I had finished the first draft, I checked that the content was relevant.	2.86	.976	2.99	1.136	2274.500	4975.500	-1.195	.232
5.8	After I had finished the first draft, I checked that my text was well-organised.	2.81	1.036	2.86	1.067	2468.000	5169.000	-.371	.711
5.9	After I had finished the first draft, I checked that my text was coherent.	2.86	1.045	2.81	1.094	2512.000	4997.000	-.184	.854
5.10	After I had finished the first draft, I checked that I included all appropriate main ideas from all the source texts.	2.74	.958	2.89	1.015	2304.000	5005.000	-1.065	.287
5.11	After I had finished the first draft, I checked that I included my own viewpoint on the topic.	2.73	1.004	2.90	1.079	2263.000	4964.000	-1.244	.214
5.12	After I had finished the first draft, I checked that the quotations were properly made	2.74	.958	2.87	1.006	2327.000	5028.000	-.968	.333
5.13	After I had finished the first draft, I checked that I had put the ideas of the source texts into my own words.	2.86	1.045	3.00	1.116	2306.000	5007.000	-1.060	.289
5.14	After I had finished the first draft, I checked the possible effect of my writing on the intended reader.	2.48	1.015	2.70	1.026	2246.000	4947.000	-1.320	.187
5.15	After I had finished the first draft, I checked the grammatical accuracy and range of the sentence structures.	2.74	.928	2.79	.991	2450.500	5151.500	-.458	.647
5.16	After I had finished the first draft, I checked the spelling, usage and range of vocabulary.	2.79	.957	2.76	1.042	2530.500	5015.500	-.104	.917

Appendix 5: Rejected factor solutions (real-life)

Table 1 Meaning and discourse construction phase (real-life): two-factor solution (rejected)

		F1	F2
2.5	I read some relevant part(s) of the texts carefully.	.767	
2.4	I searched quickly for part(s) of the texts which might answer the question.	.688	
2.7	I took notes on or underlined the important ideas in the source texts.	.679	
4.2	I developed new ideas while I was writing.	.629	
2.9	I linked the important ideas in the source texts to what I know already.	.565	
4.5	I selectively re-read the source texts while writing.	.456	.397
4.3	I made further connections across the source texts while I was writing.	.343	.338
2.1	I read through the whole of each source text carefully.		.892
2.2	I read the whole of each source text more than once.		.803
1.1	I read the whole task prompt (i.e. instructions) carefully		.518
2.12	I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.	.324	.803

Table 2 Organisation phase (real-life): three-factor solution (rejected)

		F1	F2	F3
2.10	I worked out how the main ideas in each source text relate to each other.	.876		
2.11	I worked out how the main ideas across the source texts relate to each other.	.851		
2.8	I prioritised important ideas in the source texts in my mind.	.599		.312
3.1	I organised the ideas for my text before starting to write.		.859	
3.2	I recombined or reordered the ideas to fit the structure of my text.		.775	
2.3	I used my knowledge of how texts like these are organised to find parts to focus on.	.368	.651	
3.3	I removed some ideas I planned to write.			.788
4.1	While I was writing I sometimes paused to organize my ideas.			.786
3.4	I tried to use the same organizational structure as in the source texts.			.347

Table 3 Low-level monitoring and revising phase (real-life): four-factor solution (rejected)

		F1	F2	F3	F4
5.15	After I had finished the first draft, I checked the grammatical accuracy and range of the sentence structures.	.922			
5.16	After I had finished the first draft, I checked the spelling, usage and range of vocabulary.	.902			
5.12	After I had finished the first draft, I checked that the quotations were properly made.	.860		.364	
5.13	After I had finished the first draft, I checked that I had put the ideas of the source texts into my own words.	.817			.359
4.15	I checked the grammatical accuracy and range of the sentence structures.		.843		
4.16	I checked the spelling, usage and range of vocabulary.		.759		
4.12	I checked that the quotations were properly made.			.632	
4.13	I checked that I had put the ideas of the source texts into my own words.				.563

Appendix 6: Rejected factor solutions (GEPT)

Table 1 Meaning and discourse construction phase (GEPT): three-factor solution (rejected)

		F1	F2	F3
2.5	I read some relevant part(s) of the texts carefully.	.922		
2.4	I searched quickly for part(s) of the texts which might help answer the question.	.752		
2.7	I took notes on or underlined the important ideas in the source texts.	.588		
4.3	I made further connections across the source texts while I was writing.		.621	
2.2	I read the whole of each source text more than once.		.553	
2.9	I linked the important ideas in the source texts to what I know already.		.423	
2.1	I read through the whole of each source text carefully.		.380	
4.2	I developed new ideas while I was writing.			
2.12	I developed new ideas or a better understanding of existing knowledge while I was reading the source texts.			.786

Table 2: Organising phase (GEPT): three-factor solution (rejected)

		1	2	3
2.11	I worked out how the main ideas across the source texts relate to each other	.776		
2.10	I worked out how the main ideas in each source text relate to each other.	.624		
2.3	I used my knowledge of how texts like these are organised to find parts to focus on.	.614		
3.1	I organised the ideas for my text before starting to write.	.542	.422	
2.8	I prioritised important ideas in the source texts in my mind.	.510		
3.2	I recombined or reordered the ideas to fit the structure of my text.	.306		
4.1	While I was writing I sometimes paused to organize my ideas.		.623	
3.3	I removed some ideas I planned to write.			.645

Table 3 High-level monitoring and revising phase (GEPT): three-factor solution (rejected)

		F1	F2	F3
5.15	After I had finished the first draft, I checked the grammatical accuracy and range of the sentence structures.	.868		
	After I had finished the first draft, I checked the spelling, usage and range of	.849		
5.16	vocabulary.			
5.12	After I had finished the first draft, I checked that the quotations were properly made.	.834		
	After I had finished the first draft, I checked that I had put the ideas of the source	.784		
5.13	texts into my own words.			
4.15	I checked the grammatical accuracy and range of the sentence structures.		.872	
4.16	I checked the spelling, usage and range of vocabulary.		.780	
4.12	I checked that the quotations were properly made.			.726
4.13	I checked that I had put the ideas of the source texts into my own words.			.707

Table 4 High-level monitoring and revising phase (GEPT): three-factor solution (rejected)

		1	2	3
5.9	After I had finished the first draft, I checked that my text was coherent.	.908		
5.7	After I had finished the first draft, I checked that the content was relevant.	.899		
5.10	After I had finished the first draft, I checked that I included all appropriate main ideas from all the source texts.	.895		
5.8	After I had finished the first draft, I checked that my text was well-organised.	.887		
5.11	After I had finished the first draft, I checked that I included my own viewpoint on the topic.	.878		
5.14	After I had finished the first draft, I checked the possible effect of my writing on the intended reader.	.670		.440
4.10	I checked that I included all appropriate main ideas from all the source texts.		.798	
4.7	I checked that the content was relevant.		.759	
4.11	I checked that I included my own viewpoint on the topic.		.700	
4.8	I checked that my text was well-organised.		.665	
4.9	I checked that my text was coherent.		.650	
4.14	I checked the possible effect of my writing on the intended reader.			.833

Table 5 High-level monitoring and revising phase (GEPT): four-factor solution (rejected)

		1	2	3	4
5.8	After I had finished the first draft, I checked that my text was well-organised.	.905			
5.7	After I had finished the first draft, I checked that the content was relevant.	.904			
5.9	After I had finished the first draft, I checked that my text was coherent.	.904			
5.10	After I had finished the first draft, I checked that I included all appropriate main ideas from all the source texts.	.882			
5.11	After I had finished the first draft, I checked that I included my own viewpoint on the topic.	.862			
5.14	After I had finished the first draft, I checked the possible effect of my writing on the intended reader.	.707			
4.7	I checked that the content was relevant.		.793		
4.8	I checked that my text was well-organised.		.778		
4.10	I checked that I included all appropriate main ideas from all the source texts.		.734		.331
4.11	I checked that I included my own viewpoint on the topic.		.634		.534
4.9	I checked that my text was coherent.		.629		
4.14	I checked the possible effect of my writing on the intended reader.			.897	



The Language Training and Testing Center (LTTC)
No.170, Sec.2, Xinhai Rd., Daan Dist.,
Taipei City, 10663 Taiwan (R.O.C.)
Tel: +886-2-2735-2565
Email: geptgrants@lttc.ntu.edu.tw
Website: www.lttc.ntu.edu.tw



©LTTC 2014