



*Language unfolds worlds.  
Testing sets standards.*

# **LTTC-GEPT Research Reports RG-02**

A Comparability Study on the Cognitive Processes of  
Taking Graph-based GEPT-Advanced and  
IELTS-Academic Writing Tasks

Guoxing Yu  
Shwu-Wen Lin

**A Comparability Study on the Cognitive Processes of  
Taking Graph-based GEPT-Advanced and  
IELTS-Academic Writing Tasks**

**LTTC-GEPT Research Reports  
RG-02**

**Guoxing Yu  
Shwu-Wen Lin**

This study was funded and supported by the Language Training & Testing Center (LTTC) under the LTTC-GEPT Research Grants Program 2011-2012

**LTTC-GEPT Research Reports RG-02**  
**A Comparability Study on the Cognitive Processes of Taking Graph-based GEPT-Advanced**  
**and IELTS-Academic Writing Tasks**

Published by The Language Training and Testing Center  
No.170, Sec. 2, Xinhai Rd., Daan Dist., Taipei, 10663 Taiwan (R.O.C)

© 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

We have great pleasure in publishing this report: *LTTTC-GEPT Research Reports RG-02*. The study in this report was funded by the 2011-2012 LTTTC-GEPT Research Grants. Headed by Professor Guoxing Yu of University of Bristol, UK, the study investigated test takers' cognitive processes when completing the graph-based writing tasks of the GEPT Advanced Level Writing Test and IELTS Academic Writing Test, with a focus on the cross-test comparability in terms of test-takers' cognitive processes and test performance.

The GEPT, developed more than a decade ago by the LTTTC to serve as a fair and reliable testing system for EFL learners, has gained wide recognition in Taiwan and abroad. It has generated positive washback effects on English education in Taiwan. As the GEPT has successfully reached out to the international academic community with remarkable success over the years, numerous studies and research projects on GEPT-related subjects have been conducted and published as technical monographs, conference papers, and refereed articles in books and journals. In view of the growing scholarly attention on the GEPT, and in order to assist external researchers to conduct quality research on topics related to the test, the LTTTC has set up the LTTTC-GEPT Research Grants Program, which offers funding to outstanding research projects.

The annual call for research proposals is publicized every October, attracting proposals from all over the world. A review board, which comprises scholars and experts in English language teaching and testing from Taiwan and abroad, evaluates the research proposals in terms of the following criteria:

- the relevance to identified areas of research
- the benefit of the research outcomes to the GEPT
- the theoretical framework, aims and objectives, and methodology of the proposed research
- the qualifications and experience of the research team
- the capability of the research outcomes to be presented at international conferences and published in journals
- the timeline and cost effectiveness of the proposed research

Complete and up-to-date information about the GEPT is available at [https://www.lttc.ntu.edu.tw/E\\_LTTTC/E\\_GEPT.htm](https://www.lttc.ntu.edu.tw/E_LTTTC/E_GEPT.htm). Full research reports can be downloaded at <https://www.lttc.ntu.edu.tw/lttc-gept-grants.htm>.

We believe that with the further contributions from the external research community, the GEPT will continue to refine its quality and achieve wider recognition at home and overseas.



Hsien-hao Liao  
Executive Director  
LTTTC

## Author Biodata

**Dr. Guoxing Yu** is currently Senior Lecturer at the Graduate School of Education, University of Bristol. He has directed or co-directed research projects funded by British Council, DFID, Educational Testing Service, GEPT, and Worldwide Universities Network. Dr. Yu has published in international journals including *Applied Linguistics*, *Assessing Writing*, *Assessment in Education*, *Comparative Education*, *Educational Research*, *International Review of Education*, *Language Assessment Quarterly*, *Language Teaching*, and *Language Testing*. Dr. Yu serves as an Executive Editor of *Assessment in Education*, and is also a member of Editorial Boards of *Assessing Writing*, *Language Assessment Quarterly*, *Language Testing*, and *Language Testing in Asia* (visit <http://www.bristol.ac.uk/education/people/guoxing-yu/> for more information).

**Dr. Shwu-Wen Lin** is currently an Assistant Professor at the General Education Center, National Taipei University of Nursing and Health Sciences, Taiwan. She holds a PhD in Language Testing from the University of Bristol. She has carried out projects in language testing.

## 摘要

### ◆ 研究團隊與研究目的

本研究由英國布里斯托大學 Professor Yu Guoxing 主持，比較全民英檢高級寫作測驗第二部分（簡稱 GEPT-AWT2）與 IELTS（學術）寫作測驗第一部分（簡稱 IELTS-AWT1）的寫作認知歷程(cognitive processes)，並探討受試者的圖像理解能力(graphicacy)、試題所選用的圖表類型（如：長條圖、曲線圖、表格、或圓餅圖等），以及寫作能力影響寫作認知歷程和寫作表現的程度，為全民英檢高級寫作測驗提供更多效度證據。

### ◆ 研究問題

- 1) 不同的圖表類型對考生寫作認知歷程的影響為何？
- 2) 考生的圖像理解能力(graphicacy)對其寫作認知歷程的影響為何？
- 3) 考生的英語寫作能力對其寫作認知歷程的關連性為何？
- 4) 考生在 GEPT-AWT2 和 IELTS-AWT1 的認知歷程可比性為何？

### ◆ 受試者

32 位來自數所臺灣的大專院校的大學生與研究生，皆為全民英檢高級測驗的目標考生；其中半數為英語主修；其他考生的主修領域包括護理、管理、設計、神經科學、公共衛生、俄語等。本研究考生的背景和全民英檢高級正式測驗的考生大致相當，具代表性。

### ◆ 資料收集方法摘要

分為三階段：第一階段，收集考生的圖像理解能力(graphicacy)以及考生一般情況下的寫作能力。第二階段，考生分別完成 GEPT-AWT2 和 IELTS-AWT1，且在寫作時同步說出他們的認知歷程。考生的報告全程錄音，並製作逐字稿供研究者後續分析。第三階段，研究者與考生面談，利用第二階段的錄音，鼓勵考生回想他們的寫作認知歷程並提供解釋或評論。面談亦全程錄音並製作逐字稿以供後續分析。

### ◆ 研究結果摘要

- ✓ 試題使用的圖表類型對考生的寫作表現沒有明顯的影響，但和寫作認知歷程存在明顯的關連性。
- ✓ 考生的圖像理解能力對寫作表現沒有明顯的影響，但分析結果顯示完成 GEPT-AWT2 和 IELTS-AWT1 需要一定程度的圖像理解能力。
- ✓ 考生在 GEPT-AWT2 和 IELTS-AWT1 的表現呈高度相關（相關性介於 0.634 至 0.864 之間），但他們在 GEPT-AWT2 和 IELTS-AWT1 的認知歷程存在部分差異。主要原因在於 GEPT 要求考生依據圖表所提供的資訊提供建議或個人意見，但 IELTS 不鼓勵考生在作文中提供個人意見。

## Abstract

GEPT-Advanced Writing Task Two (GEPT-AWT2) and IELTS Academic Writing Task One (IELTS-AWT1) use similar graph prompts, but differ in the amount of verbal and non-verbal information provided in the task directions. Both tasks require test takers to summarize the information contained in the graph prompts, but GEPT additionally requires personal interpretations of the phenomenon observed, which IELTS discourages or even penalises test takers if they do interpret, in their writings. This research investigated the comparability in test takers' cognitive processes when completing the two types of writing tasks, the extent to which test takers' performance and cognitive processes are affected by their graphicacy, English writing ability, and the use of different graph prompts (bar, line graph, pie chart, statistical tables, etc.). Thirty-two potential GEPT-Advanced test takers participated in this study. They completed four writing tasks (2 IELTS-AWT1 and 2 GEPT-AWT2) in randomized order, while thinking-aloud their writing processes. After the tests, all participants were interviewed. Baseline data on the participants' graphicacy and their writing ability under normal examination condition were also collected. The think-aloud and interview data were analysed to identify patterns of cognitive processes, with close reference to the participants' graphicacy, writing ability and features of graph prompts. It was found graphicacy and types of graphs had only negligible impacts on participants' test performance (but cf. Yu, Rea-Dickins and Kiely 2011). The participants' performance in GEPT-AWT2 and IELTS-AWT1 tasks were found highly correlated. However, differences in the participants' cognitive processes when completing IELTS-AWT1 and GEPT-AWT2 were clearly evidenced, in particular, towards the second part of the GEPT-AWT2 tasks which required test takers to make personal interpretations of the data presented in the graphs. Furthermore, the think-aloud and interview data provide ample evidences of the differential impacts of graph prompts, test takers' graphicacy and writing ability on test takers' cognitive processes. Methodologically such findings highlight the importance and usefulness of examining test takers' cognitive processes, in addition to test scores, when conducting test comparability or alignment studies.

## Table of Contents

<b>Introduction</b> .....	1
<b>Literature Review</b> .....	2
<b>Research Questions and Methodology</b> .....	4
Research questions .....	4
Participants .....	5
Procedures .....	5
<i>STAGE 1: Collection of baseline data</i> .....	5
<i>STAGE 2: Collection of cognitive processes</i> .....	6
<i>STAGE 3: Interviews after tests</i> .....	6
Data preparation .....	7
<b>Findings and Discussions</b> .....	8
Participants’ graphicacy.....	8
Participants’ writing performance under normal examination condition.....	10
Participants’ writing performance under think-aloud condition.....	13
Addressing the research questions .....	17
<i>Research question one</i> .....	17
<i>Research question two</i> .....	23
<i>Research question three</i> .....	25
<i>Research question four</i> .....	29
<b>Conclusion</b> .....	35
<b>References</b> .....	37
<b>Appendices</b> .....	38





## Introduction

This study addresses two topics identified by GEPT Call for Proposals 2010-11, namely, *Comparability studies of the GEPT and IELTS*, and *Investigation of cognitive processes of GEPT test takers*. GEPT Advanced Writing Task Two (GEPT-AWT2) and IELTS Academic Writing Task One (IELTS-AWT1) both use graphs as prompts to measure test takers' writing abilities, requiring test takers not only to comprehend graph input, but also to re-present in written continuous discourse in English the information accessible to them. However, there are at least two main differences between the two tasks. Firstly, IELTS-AWT1 tasks require test takers to “describe some information (graph/chart /table/diagram), and to present the description in their own words”. (Note: We use “graph” in this report to refer to any kind of graph input, unless otherwise specified, see Yu et al. 2011 for the rationale for doing this). IELTS test takers are assessed on their ability to organise, present and possibly compare data, describe the stages of a process or procedure, describe an object or event or sequence of events, or explain how something works (IELTS Handbook 2006, p.8). GEPT-AWT2, however, requires test takers to not only summarise the main information as in IELTS-AWT1 but also discuss the underlying reasons/causes for the data and make reasonable personal comments, predictions, interpretations and suggestions. Secondly, in IELTS-AWT1, test takers are recommended to spend about 20 minutes on the task; while GEPT-AWT2 allows 45 minutes. Furthermore, there is more verbal input in GEPT-AWT2 than IELTS-AWT1 (See Appendices 1 and 2).

To what extent these two types of tasks are comparable is both a theoretical and a practical question, in research studies on test comparability, test mapping or alignment as well as in test marketing and test use. GEPT-Advanced is claimed to be designed at the C1 level of CEFR, and passing GEPT-Advanced Speaking and Writing tests is considered as equivalent to achieving IELTS 7.5 or TOEFL iBT 110 (see [https://www.lttc.ntu.edu.tw/e\\_lttc/E\\_GEPT/alignment.htm](https://www.lttc.ntu.edu.tw/e_lttc/E_GEPT/alignment.htm)). Similarly, IELTS Partners claim that those who achieve at certain IELTS grade are considered at C1 level (see Figure 1).

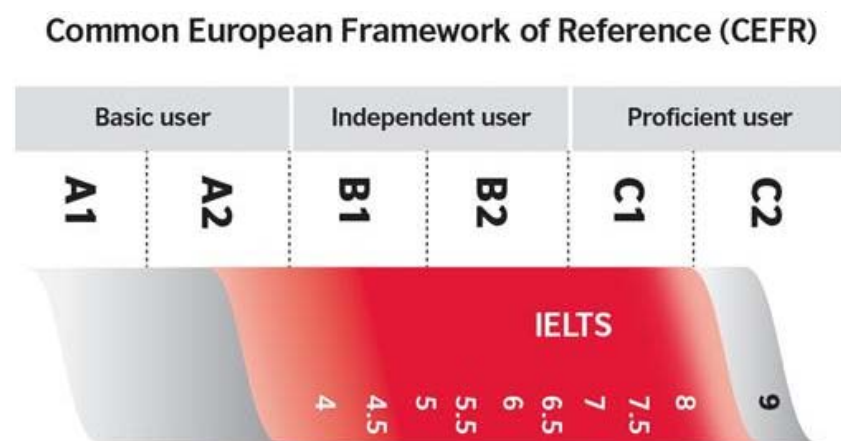


Figure 1: Indicative IELTS band scores at CEFR levels

The present research is neither to endorse nor to challenge such claims, instead it aims to examine the complexity of task comparability between tests, from the perspectives of test takers' cognitive processes in addition to their test performance. The majority of language test comparability studies have heavily relied on test performance data to make their comparability conclusions. It is our hope that by investigating the micro-ecology of test taking cognitive processes we can contribute to answering the question of test comparability.

Furthermore, this present research aims to examine the extent to which the variability in the amount of graph and non-graph input in the task directions, the features of graphs as well as test takers' familiarities and proficiencies in comprehending graphs may affect test takers' cognitive processes as well as the quality of their test performance, because such variability poses a threat to the validity of graph-based tasks as a measure of writing ability.

Findings from this research contribute to our understanding about the comparability of GEPT-AWT2 and IELTS-AWT1, the ongoing validation and development of the two tests as well as other language test tasks that use graphs as prompts (e.g., the oral description of a graph as a Pearson PTE-Academic speaking task). Methodologically, this research highlights the importance of investigating test takers' cognitive processes when we conduct test comparability or test mapping or alignment exercise.

## **Literature Review**

In contrast to the rich knowledgebase in applied cognitive and educational psychology on graph comprehension (for a detailed literature review, see Yu et al. 2011), there is only a small number of studies on the use of graphs or similar non-verbal prompts in language tests, e.g., O'Loughlin & Wigglesworth (2003), Yang (2012) and Yu et al. (2011) on writing, Ginther (2002) on listening, Katz, Xi, Kim & Cheng (2004) and Xi (2005, 2010) on speaking. These studies shed some lights on how the different features of graph prompts, test takers' language proficiency and graphicacy – defined as “proficiency in understanding quantitative phenomena that are presented in a graphical way” (Wainer, 1992, p. 16) – may affect their test performance. Below we review the research studies on the use of graphs in writing tasks, and Weir, Chan and Nakatsuhara (2013) which examined the comparability between GEPT-Advanced and IELTS in their reading and writing tasks.

O'Loughlin and Wigglesworth (2003) examined how the quantity and the manner of presentation of information in graphs (statistical table, bar chart, line graph) might affect the difficulty of the graph-based writing tasks. By primarily analysing the features of the written scripts, they found that test takers produced more linguistically complex writings for tasks with less information contained in the graphs than for those tasks with more information in graphs, regardless of their language proficiency level. The manipulations in the quantity of information in the graphs and the manner of presenting the information did not make substantial change in terms of the difficulty level of the graph-based writing tasks.

Yang (2012) examined Taiwanese medical students' use of test-taking strategies when completing the graph-based writing task of GEPT-Advanced. Based on a questionnaire survey conducted after the students completed the writing task, she found that the students were engaged in three processes, namely, graph comprehension, graph interpretation and graph translation and that test takers' performance was generally positively, but indirectly, related to their engagement with the three key processes, as well as with their graph familiarity, topic and lexical knowledge and test-wiseness. By "indirect", she meant that the impacts of graph comprehension on test performance were via means of graph interpretation and graph translation, and similarly, the impacts of graph interpretation via means of graph translation, due to the significant correlations among the three processes. She therefore suggested that the three processes were "a set of inseparable strategy repertoires" throughout the writing process. A similar finding was noted in Yu et al. (2011) who viewed the IELTS graph-writing process iterative rather than linear; test takers were constantly monitoring and regulating simultaneously their graph comprehension and graph re-production (similar to Yang's "graph translation") in a continuous written discourse. Given that she used GEPT-type graph-based writing task, Yang's (2012) findings are particularly pertinent to the present study, however, it is worth pointing out a number of caveats in its research design before any further extrapolations to the present study can be made. Firstly, she only used one graph-based writing task involving two types of graphs (pie chart and line graph); secondly, the time allowed for completing the task was 30 minutes instead of 45 minutes as in real GEPT tests; thirdly, she used a rating scale developed by herself rather than the official GEPT scoring rubrics; fourthly, she did not actually measure the participants' graphicacy or graph familiarity and knowledge; and fifthly, as she acknowledged this was "essentially a questionnaire study" (p.184) which was not possible to capture fully what was really going on during the test. "More qualitative analyses, such as verbal protocols and eye-tracking data offer more insights into writers' mental operations in responding to the task." (p.184)

Think-aloud was used as the main data collection tool to investigate IELTS AWT1 test takers' cognitive processes in Yu et al. (2011). Although Yu et al. (2011) was mainly a qualitative research study, it also analysed a substantial amount of quantitative data. Based on both test takers' cognitive processes and the quality of their test performance (including test scores as well as discourse features of the written scripts), they reported four main findings:

- Different types of graphs steered the test takers towards certain conventions associated with the graphs, at the information-processing stage as well as at the stage when they re-produced their comprehension in written continuous discourse in English. Such effects of different graph prompts on the cognitive processes were clearly evidenced in the mean scores of the writings, lexical choice, and in whether and how they would make comparisons or assess trends according to the conventions of graphs.
- At the product-level, test takers' graphicacy did not affect their task performance in terms of the marks that their writings were awarded. However, at the process-level, there was some clear evidence of the potential psychological impact of graph familiarity on task performance, although these participants had high graphicacy.

- A strong correlation was observed in test takers' performance between the graph-based integrated writings and the topic-based argumentative writing (as measured in IELTS Academic writing task two).
- Test takers had a natural and strong inclination to interpret, predict and comment by trying to link the graph information with their previous knowledge about the graphs and graph conventions, although they were told explicitly not to do so according to the task directions (see Appendix 1).

In contrast, such interpretations, predictions, suggestions and comments in writing are the essential requirement of GEPT-AWT2 tasks (see Appendix 2, and also Weir, Chan and Nakatsuhara 2013, p.5). It would be interesting to know to what extent GEPT-AWT2 and IELTS-AWT1 tasks may share some similarities or diverge in this particular aspect of cognitive processes. From a wider perspective on the criterion-related comparability between GEPT and IELTS, in terms of test takers' performance, Weir et al. (2013) found a positive and moderate correlation ( $r=0.432$ ) between GEPT total writing scores (including GEPT task 1 and GEPT-AWT2) and IELTS writing bands; furthermore, they found that the correlation was slightly higher between GEPT-AWT2 total scores (i.e., the sum of the four analytical scores, see Appendix 3 for the rating scale of GEPT-Advanced writings) and IELTS writing bands ( $r=0.467$ ) than between GEPT task 1 total scores and IELTS writing bands ( $r=0.332$ ). The slightly higher correlations, as they explained, "is most likely because the task setting of GEPT Writing Task 2 (summary from non-verbal input) is comparable to IELTS Task 1 (Describing visual input)" (Weir et al. 2013, p.15). It is true that GEPT-AWT2 is more comparable than GEPT Task 1 to IELTS-AWT1, at least in terms of their task directions; and their participants' performance data might have lent further support to Weir et al's comparability claim; however, it remains a question as to what extent this level of comparability in test scores can be extrapolated to and evidenced in test taking cognitive processes – the focus of the present study.

## **Research Questions and Methodology**

### **Research questions**

In order to examine the comparability of GEPT-AWT2 and IELTS-AWT1, this present study investigated not only test performance but also cognitive processes of potential GEPT-Advanced test takers when completing both types of tasks using different graph prompts. We aim to address four research questions. The first three research questions examine to what extent the use of different graph prompts, test takers' graphicacy and writing ability affect their cognitive processes of completing the tasks. The fourth research question aims to explore to what extent the similarities and differences in test takers' cognitive processes between the two types of tasks are attributable to the aforementioned factors (i.e., graph features, test takers' graphicacy and writing ability).

- RQ1: To what extent are there differences in the cognitive processes due to different graph prompts?
- RQ2: To what extent are the cognitive processes affected by test taker's graphicacy?
- RQ3: To what extent are the cognitive processes related to test taker's writing ability?
- RQ4: What is the comparability between GEPT-AWT2 and IELTS-AWT1 tasks in the cognitive processes attributable to the use of different graph prompts (RQ1), test takers' graphicacy (RQ2) and writing ability (RQ3)?

## **Participants**

Thirty-two potential GEPT-Advanced test takers participated in this study as volunteers. As a token of our appreciation, they were paid NT\$1000 each. According to the GEPT Advanced report on the 2009 tests, 36% of test takers were undergraduate English majors, 22% undergraduates of other subject areas, 19% postgraduates, in addition, 60% of participants were from the north. In order to reflect the population of GEPT-Advanced, we proposed to recruit our participants consisting of approximately 50% English majors, 25% undergraduates of other subject areas, and 25% of postgraduates, mainly from universities in the northern area. In our sample achieved, 23 are female and 9 male. Sixteen (50%) declared that English was their major, the rest were in a variety of subject areas (including nursing, 7; management, 2; design with English as a minor, history, neuroscience, public health, Russian, and translation, 1 each; and 1 unknown). Nearly a third of our participants (n=11) came from National Taiwan Normal University, and 9 of them were English majors in that institution. Only two of our participants were studying for a postgraduate degree. Broadly speaking, the sample size and the characteristics of our participants were as proposed.

## **Procedures**

We collected both qualitative and quantitative data at three stages with a number of data collection sessions, as summarised below. At each Stage, data were collected within one half-day with a nominal break between sessions. The number of participants involved in any given session varied, depending on the nature of data to be collected as well as participants' availability. At Stage 1, we did not limit the number of participants in any session; at Stage 2, the maximum number of participants in any session was four; and at Stage 3, most interviews were conducted on a one-to-one basis, however, some interviews involved two or three participants. The intervals between Stages 1 and 2 varied from one to two days, depending on the availability of the participants; Stage 3 interviews were conducted immediately after Stage 2.

### ***STAGE 1: Collection of baseline data***

- a) Every participant was briefed the purpose of the research and asked to sign the consent form (Appendix 4) if he or she decided to participate.
- b) Administration of GEPT-AWT2 (Appendix 5) and IELTS-AWT1 and Task 2 (Appendices 6 and 7) to measure the participants' writing abilities under normal examination

condition.

- c) Administration of Graph Familiarity Questionnaire (Appendix 8) to understand the participants' familiarity and comprehension of different types of graphs. The graphicacy questionnaire also collected the participant's personal information such as their name, gender, university and specialism, and GEPT and IELTS test-taking experience, scores and future plan for taking the two tests.
- d) Training in think-aloud verbal report (Green, 1998) to familiarize the participants with the use of this data collection method (Appendix 9). The participants were allowed to think aloud in English and Chinese. The participants were advised to practise thinking aloud further at their own time after the formal training.

### ***STAGE 2: Collection of cognitive processes***

- a) Opportunity for further familiarization with think-aloud was provided to the participants. Only when they were comfortable with thinking aloud would the tests begin.
- b) Administration of GEPT-AWT2 and IELTS-AWT1 tasks. Three sets of GEPT-AWT2 tasks of different graph prompts and three corresponding IELTS-AWT1 tasks using the same graphs<sup>1</sup> were developed by the authors (see Appendices 10-15). Each participant was randomly assigned two of the three sets of tasks. As a result, Tasks A and B were completed by 21 participants, and Task C by 22. In specific, of the 21 participants who did Task A, 10 of them also did Task B and 11 of them also did Task C. Of the 21 participants who did Task B, 10 of them also did Task A and 11 of them also did Task C. Of the 22 participants who did Task C, half of them also did Task A and the other half did Task B. In addition, we also randomized the order that a participant would start with a GEPT or IELTS task of any given set of tasks. If a participant was randomly assigned to start with a GEPT task in his first set of tasks, then he or she would start with an IELTS task in his second set of tasks; and vice versa. The participants were asked to think aloud continuously, which were audio-recorded, in the duration of the time allocated for the tasks. Field notes were taken by the second author when observing the test taking process. Due to the potential effects of think-aloud on writing, extra 5 minutes were added to each task. Between each task, there was a nominal 5-minute break. In total, there were approximately 80 hours of think-aloud recordings (2.5 hours x 32 participants), which were transcribed verbatim for further analysis.

### ***STAGE 3: Interviews after tests***

Immediately after their writing tests, we interviewed the participants. Depending on the number of participants in the preceding Stage, twelve interviews were conducted on a

---

<sup>1</sup> They were designed on purpose to minimize the effects of potential extra factors that might have made the interpretations of the comparability of the two types of tasks even more complex. By using the same graphs within a set of corresponding tasks, we can focus on the comparability issues of the GEPT and IELTS writing tasks, without having to simultaneously disentangle or worry too much about the effects of different graph features on test performance. However, since this research also aimed to understand the impacts of different graph features on test performance, we used three sets of tasks of different kinds of graphs. We believe that such a design affords more opportunities for us to better understand the two issues (i.e., comparability of GEPT and IELTS and the effects of graph on test performance) than if we had implemented only different graphs in our research design.

one-to-one basis, 7 with two participants, and 2 with three participants. In total, there were about 7.5 hours of audio-recorded interviews. The interviews were conducted mostly in Chinese, asking the participants to comment and reflect on i) in what ways their cognitive processes may be affected by the different graph prompts, their graphicacy and writing abilities, and ii) the comparability in their cognitive processes between GEPT-AWT2 and IELTS-AWT1 (see Appendix 16 for the list of questions that we asked at the interviews). Field notes and think-aloud recordings were used to assist the interviews. All the interviews were transcribed verbatim for further analysis.

In summary, this research comprised three distinct stages. The data from each participant include three written scripts (GEPT-AWT2, IELTS-AWT1 and IELTS Task 2) produced under normal examination condition, one graphicacy questionnaire (Stage 1), four think-alouds while completing four writing tasks (2 GEPT-AWT2 and 2 IELTS-AWT1) and four corresponding written scripts (Stage 2), and interviews (Stage 3).

### **Data preparation**

We followed the standard procedure to prepare the data for further analysis. The recorded think-aloud protocols and interviews were transcribed verbatim, coded and categorized in a qualitative data analysis software. The graphicacy questionnaire data were entered into SPSS; in addition, the participants' responses to the final open ended question were analysed separately.

The written scripts were marked against the respective rating scales for GEPT-Advanced and IELTS-Academic writing tasks. As shown in Appendix 3, the total raw score for GEPT-AWT2 is 20, five points maximum for each of the four category (RA: Relevance and Adequacy, CO: Coherence and Organization, LU: Lexical Use, GU: Grammatical Use). GEPT requests test takers to achieve 3 or above in each category to pass the writing test. In the official score report, it uses 0-5, rather than 0-20; however, for the purpose of the present research, we use the total raw score as well as the sub-scores in our analysis. Although IELTS has similar and more detailed band descriptors, the practice is to give one holistic score (0-9); rather than scores for each of the four sub-categories ("task achievement", "coherence and cohesion", "lexical resource", and "grammatical range and accuracy"); therefore we report only a holistic score for IELTS writings.

Because neither of us are certificated GEPT or IELTS raters (although the first author has extensive experience in marking IELTS writings), we studied carefully the sample scripts, which are available on the websites of the test organisations and were marked by certificated raters, to improve our understanding of how to interpret and apply the rating scales. In the case of GEPT, we found only one sample script (Appendix 17) which was awarded 3 in each of the four categories. In the case of IELTS-AWT1, there are a few sample scripts of different grades available online ([http://www.ielts.org/PDF/113313\\_AC\\_sample\\_scripts.pdf](http://www.ielts.org/PDF/113313_AC_sample_scripts.pdf)). All our participants' scripts were anonymized by giving a number before being marked by the first author. The marking involved three iterative rounds. In the first round, he read all the scripts



of each task, without marking them, in order to get a sense of the overall quality of the writings and also to identify a script that can be used as an anchor (in the case of GEPT, it was a script that can be awarded 3 in each of the four categories). In the second round, he marked the scripts against the rating scale and the anchor script(s). Because of the small number of scripts, he was able to make some comparisons and revise the grades accordingly throughout the marking process within each round. In the third round, after a short break, he marked the scripts again which were randomized so that they were not in the same order as in the first or second round. He then looked at the differences between the second and the third round. Any scripts that had difference equal or larger than 2 in GEPT total score (20 maximum), or 1 in IELTS holistic score (9 maximum), were read again and re-marked. Although we did not use certificated raters or two or more independent raters, we found that our method provided an excellent alternative; and perhaps even a much better way of marking than if two independent raters had very different interpretations of the rating scales. We admit that a potential risk of using only one rater could be grade inflation or otherwise, even if the rating scales have been consistently applied; however, we believe that this risk was mitigated by constantly referring to the anchor script(s).

## **Findings and Discussions**

The data were analysed first of all to understand the characteristics of our participants, i.e., their graphicacy and writing performance under normal examination condition, and under think-aloud condition. We then report our findings to each research question.

### **Participants' graphicacy**

Questions 1-29 and Questions 34-35 of the graphicacy questionnaire (Appendix 8) were used to measure the participants' graphicacy level, and Questions 30-33 to measure their understanding about the relationships between graphicacy and performance of graph-based writing tasks. As Questions 7-9 and 29 were negative questions, we recoded the participants' responses so that a bigger number indicates a higher graphicacy level. As in Yu et al. (2011) which used the English version, the graphicacy scale (31 questions/items) achieved high reliability (Cronbach Alpha =0.922). According to the scale, the maximum possible score would be 186 points (31 x 6), and the minimum, 31 points (31 x 1). The data indicated that our participants have a good level of graphicacy, with a mean of 125.5 (mini.=93, max.=167, std. deviation = 20.9, see Figure 2).

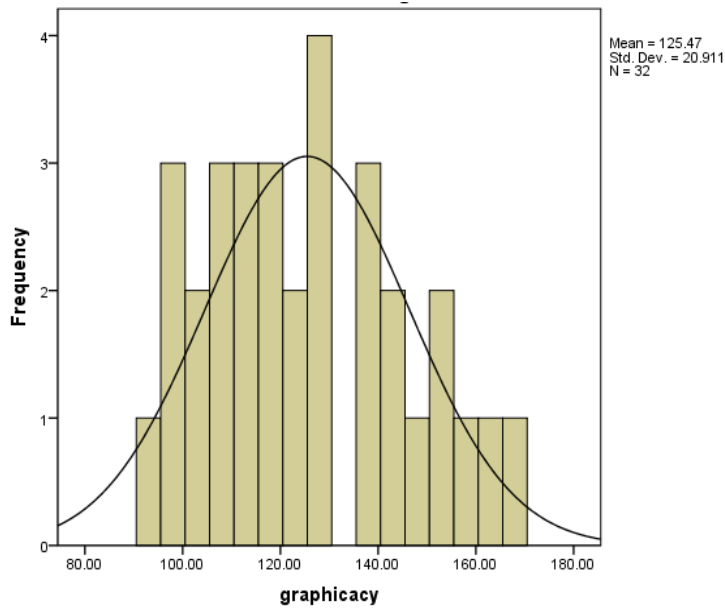


Figure 2: Participants' graphicacy

Questions 10-14 asked for the participants' familiarity and experience with different types of graphs. As Table 1 shows, the participants were more familiar with bar graph and pie chart, which had the same mean and standard deviation, than diagram or statistical table; and they were least familiar with statistical table. The differences between bar/pie graph and statistical table were significant ( $t=2.13$ ,  $df=31$ ,  $p.<.05$ ).

Table 1: Participants' familiarity with different types of graphs

	N	Minimum	Maximum	Mean	Std. Deviation
gf10 (bar graph)	32	1	6	4.47	1.295
gf11 (line graph)	32	2	6	4.19	1.281
gf12 (pie chart)	32	1	6	4.47	1.295
gf13 (diagram)	32	2	6	4.03	1.204
gf14 (statistical table)	32	2	6	4.00	1.459

Questions 30-33 measured the participants' understanding about the relationships between their graphicacy and performance of graph-based writing tasks. In particular, they were asked to what extent the following statements applied to them at a scale of 1 to 6 ("strongly disagree" to "strongly agree")

- Question 30: *I am concerned that I cannot fully demonstrate my writing ability in GEPT-AWT2 because I am not good at describing graphs*
- Question 31: *I may do better in GEPT-AWT2 using familiar graphs than unfamiliar ones.*
- Question 32: *I would prefer one type of graph to be used in GEPT-AWT2*
- Question 33: *Special training on how to describe graphs would be helpful for me to get a higher score in GEPT-AWT2.*

According to the chi-square test statistics, the participants' responses seemed to spread across the six categories/options but the differences among them were not significant. In other words, there did not seem to be any agreement within the group in terms of their views on the relationships between graphicacy and GEPT-AWT1 task performance. This finding is interesting, especially if we compare it with the findings from a similar study on IELTS-AWT1 tasks in Yu, He and Isaacs (in progress). In Yu et al. (in progress), we noticed a very different pattern. With regard to Question 30, the participants in both studies did not think "not good at describing graphs" would prevent them from demonstrating their writing ability. With regard to Question 31, unlike the participants of the present study, the overwhelming majority of the participants in Yu et al. (in progress) believed that that "their familiarity with certain types of graphs would be helpful for them to achieve a better score" in IELTS-AWT1 tasks although they were not so sure if they would prefer one type of graph over another. Similar finding was noted in the present study with regard to Question 32. With regard to Question 33, the participants in Yu et al. (in progress) were almost unanimous to value highly the impacts of special training on how to describe graphs on their performance in IELTS-AWT1; however, this is not the case in the present study on GEPT-AWT2. Such findings might give an impression that the participants in Yu et al. (in progress) might have lower graphicacy level so that they were more concerned about how their graphicacy might affect their test performance. However, this is not the case because the participants in Yu et al. (in progress) had higher graphicacy (mean=138.2, std. deviation=19.3) than the participants of the present study (mean=125.5, std. deviation=20.9). One possible explanation for the difference between these two groups of participants may be that: graph prompts and test takers' graphicacy might be less essential in successful completion of GEPT-AWT2 than IELTS-AWT1 due to the different task requirement (see Appendices 1 and 2, and Introduction section). The participants' views on the complex relationship between graphicacy and test performance is only one side of the coin, though. To what extent their graphicacy affected their actual test performance is the other side of the coin; and similarly, their cognitive processes as demonstrated in the recorded think-aloud protocols would provide further insights to understanding the complex relationship between graphicacy and test performance.

Similarly, the thirteen responses (see Appendix 18) to the final open-ended question are particularly revealing about these participants' familiarity and experience with different types of graphs and their overall graphicacy level, as well as their views on the impacts of graphicacy on performance of graph-based writing tasks.

### **Participants' writing performance under normal examination condition**

Under normal examination condition, the participants completed three writing tasks: GEPT-AWT2, IELTS-AWT1, and IELTS Task Two. Their performance data are presented in Table 2. We understand that IELTS Task Two carries more weight in marking than IELTS-AWT1 (according to the official guidelines of IELTS Partners for test takers, [http://www.ielts.org/PDF/113313\\_AC\\_sample\\_scripts.pdf](http://www.ielts.org/PDF/113313_AC_sample_scripts.pdf)) for the final IELTS writing band; however, we don't know how exactly the weighting is applied for the tasks, therefore we report the two scores separately.

Table 2: Participants' test performance in Stage One writing tasks

	N	Minimum	Maximum	Mean	Std. Deviation
GEPT score (Stage1 task)	32	8	20	13.53	3.243
IELTS score (Stage 1 task 1)	32	4.0	6.5	5.266	.7294
IELTS score (Stage 1 task 2)	32	4.5	7.5	6.000	.7620

The participants' performance in the three tasks are highly correlated. As shown in Table 3, IELTS-AWT1 and IELTS Task 2 had the highest correlation ( $r=0.769$ ), followed by the correlation between IELTS Task 2 and GEPT-AWT2 ( $r=0.751$ ), and between GEPT-AWT2 and IELTS-AWT1 ( $r=0.654$ ). The highest correlation between the two IELTS tasks was as anticipated. The second highest correlation between the GEPT graph-based task (GEPT-AWT2) and the IELTS topic-based argumentative essay task (IELTS Task 2) suggests that these two tasks may share a lot more in their underlying construct of measuring the participants' writing ability than what the GEPT and IELTS graph-based tasks have in common, at least in terms of the test performance data. The two graph-based tasks had the lowest correlation, although the correlation itself is high in statistical sense. Forty-three percent of the performance of IELTS-AWT1 can be explained by the GEPT-AWT2 scores. It is important to note that the two tasks did not use the same graph prompts. It would be interesting to know if the correlations would be even higher if same graph prompts had been used – which was the design of the writing tasks in Stage 2 of this present research. Furthermore, as we argued earlier, the test performance data can provide only a partial picture of the comparability of the two graph-based writing tasks. Test takers' cognitive processes would be essential to gain a better understanding about the comparability of the two types of tasks.

Table 3: Correlations between three Stage One task performances

		GEPT score (Stage1 task)	IELTS score (Stage 1 task 1)	IELTS score (Stage 1 task 2)
IELTS score (Stage 1 task 1)	Pearson Correlation	.654**	1	
	Sig. (2-tailed)	.000		
	N	32	32	32
IELTS score (Stage 1 task 2)	Pearson Correlation	.751**	.769**	1
	Sig. (2-tailed)	.000	.000	
	N	32	32	32

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

As the GEPT scripts were marked analytically in four categories: relevance and adequacy (RA), coherence and organization (CO), lexical use (LU) and grammatical use (GU). We also report the participants' performance in the four categories in Table 4.

Table 4: Participants' performance in Stage One GEPT task

	N	Minimum	Maximum	Mean	Std. Deviation
GEPT1 relevance & adequacy (RA)	32	2	5	3.56	.878
GEPT1 coherence & organization (CO)	32	2	5	3.44	.840
GEPT1 lexical use (LU)	32	2	5	3.22	.906
GEPT1 grammatical use (GU)	32	2	5	3.31	.896

As Table 4 shows, on average, the participants did better in RA than the other sub-categories. They did the worst in LU. The paired samples tests (Table 5) indicated that the differences between RA and LU ( $t=3.232$ ,  $p<.005$ ), between RA and GU ( $t=2.784$ ,  $p<.01$ ), and between CO and LU ( $t=2.239$ ,  $p<.05$ ) are statistically significant.

Table 5: Paired samples tests of Stage One GEPT sub-scores

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error	95% Confidence Interval of the Difference				
	Mean	Std. Deviation	Std. Error	Lower	Upper	t	df	Sig.
Pair 1 GEPT1 RA – GEPT1 CO	.125	.609	.108	-.095	.345	1.161	31	.255
Pair 2 GEPT1 RA – GEPT1 LU	.344	.602	.106	.127	.561	3.232	31	.003
Pair 3 GEPT1 RA – GEPT1 GU	.250	.508	.090	.067	.433	2.784	31	.009
Pair 4 GEPT1 CO – GEPT1 LU	.219	.553	.098	.019	.418	2.239	31	.032
Pair 5 GEPT1 CO – GEPT1 GU	.125	.554	.098	-.075	.325	1.277	31	.211
Pair 6 GEPT1 LU – GEPT1 GU	-.094	.530	.094	-.285	.097	-1.000	31	.325

These significant differences between the sub-scores of GEPT-AWT2 indicated that it would be worth exploring further the correlations between the three tests as Stage 1 tasks (as reported in Table 3), by using the GEPT-AWT2 sub-scores. As shown in Table 6, all the four GEPT sub-scores had higher correlations with IELTS Task 2 than IELTS-AWT1; and with IELTS-AWT1, the highest was “grammatical use” ( $r=0.659$ ), followed by “relevance and adequacy” ( $r=0.641$ ), “coherence and organization” ( $r=0.568$ ) and “lexical use” as the lowest ( $r=0.544$ ).

Table 6: Correlations between GEPT sub-scores and IELTS-AWT1 and IELTS Task 2

		GEPT1 RA	GEPT1 CO	GEPT1 LU	GEPT1 GU	IELTS-A WT1	IELTS Task 2
GEPT1 RA	Pearson Correlation	1	.749**	.773**	.836**	.641**	.724**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
GEPT1 CO	Pearson Correlation		1	.802**	.798**	.568**	.655**
	Sig. (2-tailed)			.000	.000	.001	.000
GEPT1 LU	Pearson Correlation			1	.827**	.544**	.677**
	Sig. (2-tailed)				.000	.001	.000
GEPT1 GU	Pearson Correlation				1	.659**	.709**
	Sig. (2-tailed)					.000	.000
IELTS-AWT1	Pearson Correlation					1	.769**
	Sig. (2-tailed)						.000

\*\* . Correlation is significant at the 0.01 level (2-tailed). N=32

Table 6 also reports the correlations between the GEPT-AWT2 sub-scores. The highest correlation was between “relevance and adequacy” and “grammatical use” ( $r=0.836$ ), followed by “lexical use” and “grammatical use” ( $r=0.827$ ). It was mostly because the “relevance and adequacy”, which essentially measures whether the task has been successfully completed, was largely determined by the participants’ appropriate grammatical and lexical use in order to fulfil the task. The lowest correlation ( $r=0.749$ ) was between “relevance and adequacy” and “coherence and organization”. It was probably because writings which might not be so coherently or logically organized could still convey the key message to readers.

### Participants’ writing performance under think-aloud condition

Under the think-aloud condition, each participant completed two sets of IELTS-AWT1 and GEPT-AWT2 tasks, in random order. Within each set, there was one IELTS-AWT1 and one GEPT-AWT2 task which used the same graphs. As shown in Table 7, Task B writings received the highest mean score, followed by Task C writings and Task A writings. This order is true for both GEPT and IELTS tasks. Task B used a line graph and a horizontal bar graph about credit card debts in Taiwan. Task C used a vertical bar graph and a pie chart about mainland China’s carbon dioxide emission and sources of electricity production. Task A used two tables on some Asian regions’ average test scores of two international English language tests.

Table 7: Participants' performance in Stage Two IELTS-AWT1 and GEPT-AWT2 tasks

	N	Minimum	Maximum	Mean	Std. Deviation
Stage 2 Task A IELTS score	21	4.0	6.5	5.524	.6610
Stage 2 Task A GEPT total score	21	8	18	11.67	2.763
Stage 2 Task B IELTS score	21	4.0	7.5	5.762	.9030
Stage 2 Task B GEPT total score	21	7	18	12.24	3.375
Stage 2 Task C IELTS score	22	4.5	7.5	5.659	.8365
Stage 2 Task C GEPT total score	22	5	19	11.95	3.415

Note: Task A used two tables; Task B, line graph and horizontal bar graph; Task C: vertical bar graph and pie chart

Due to the method of allocating the writing tasks to the participants (see Procedures), around ½ of those who completed a particular writing task, e.g., Task A, also completed Task B and Task C respectively, and vice versa. As shown in Table 8 which reports the correlations in test performance between the six writing tasks, all the correlations within the Tasks (the shaded cells in the table) are statistically significant: Task A IELTS and GEPT had the lowest correlation ( $r=0.634$ ), followed by Task C ( $r=0.736$ ) and Task B which had the highest ( $r=0.864$ ). The across-task correlations (the unshaded cells) were also statistically significant, except for that between Task A IELTS and Task C GEPT ( $r=0.43$ , n.s.), between Task A IELTS and Task B IELTS, and between Task A IELTS and Task B GEPT (in both cases,  $r=0.622$ ,  $p<.06$ , i.e., approaching significance level). The very high correlations between Task A GEPT and Task C GEPT ( $r=0.940$ ,  $p<.001$ ), and between Task B GEPT and Task C GEPT ( $r=0.948$ ,  $p<.001$ ) are interesting (note the small sample size,  $n=11$ ). There seemed to be higher correlations across GEPT tasks than across IELTS tasks.

Table 8: Correlations in test performance between Stage Two tasks

		Task A	Task A	Task B	Task B	Task C	Task C
		IELTS	GEPT Total	IELTS	GEPT total	IELTS	GEPT total
Task A	Pearson Correlation	1	.634**	.622	.622	.823**	.430
IELTS score	Sig. (2-tailed)		.002	.055	.055	.002	.187
	N		21	10	10	11	11
Task A	Pearson Correlation		1	.807**	.809**	.734*	.940**
GEPT total	Sig. (2-tailed)			.005	.005	.010	.000
score	N			10	10	11	11
Task B	Pearson Correlation			1	.864**	.838**	.819**
IELTS score	Sig. (2-tailed)				.000	.001	.002
	N				21	11	11
Task B	Pearson Correlation				1	.763**	.948**
GEPT total	Sig. (2-tailed)					.006	.000
score	N					11	11
Task C	Pearson Correlation					1	.736**
IELTS score	Sig. (2-tailed)						.000
	N						22

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

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

We ran some further analyses on the relationships between the sub-scores (RA, CO, LU and GU) for each GEPT-AWT2 task (Tables 9-11 on the paired samples tests). The correlations between the four sub-scores ranged from 0.642 to 0.844. These were pretty similar to the patterns observed in GEPT-AWT2 under normal examination conditions (see Tables 4-6).

Table 9: Comparisons of GEPT-AWT2 sub-scores (Stage 2, Task A)

		Paired Differences							Sig. (2-tailed )
		Mean	Std. Deviation	Std. Error Mean	95% Confidence		t	df	
					Interval of the				
					Lower	Upper			
Pair 1	RA- CO	.429	.507	.111	.198	.659	3.873	20	.001
Pair 2	RA - LU	.619	.669	.146	.315	.924	4.240	20	.000
Pair 3	RA - GU	.429	.507	.111	.198	.659	3.873	20	.001
Pair 4	CO - LU	.190	.602	.131	-.083	.464	1.451	20	.162
Pair 5	CO - GU	.000	.447	.098	-.204	.204	.000	20	1.000
Pair 6	LU - GU	-.190	.680	.148	-.500	.119	-1.284	20	.214



Table 10: Comparisons of GEPT-AWT2 sub-scores (Stage 2, Task B)

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	RA- CO	.571	.598	.130	.299	.843	4.382	20	.000
Pair 2	RA - LU	.333	.730	.159	.001	.666	2.092	20	.049
Pair 3	RA - GU	.381	.590	.129	.113	.649	2.961	20	.008
Pair 4	CO - LU	-.238	.700	.153	-.557	.081	-1.558	20	.135
Pair 5	CO - GU	-.190	.512	.112	-.423	.042	-1.706	20	.104
Pair 6	LU - GU	.048	.669	.146	-.257	.352	.326	20	.748

Table 11: Comparisons of GEPT-AWT2 sub-scores (Stage 2, Task C)

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	RA- CO	.318	.646	.138	.032	.605	2.309	21	.031
Pair 2	RA - LU	.545	.596	.127	.281	.810	4.294	21	.000
Pair 3	RA - GU	.455	.510	.109	.229	.681	4.183	21	.000
Pair 4	CO - LU	.227	.612	.130	-.044	.499	1.742	21	.096
Pair 5	CO - GU	.136	.640	.136	-.147	.420	1.000	21	.329
Pair 6	LU - GU	-.091	.684	.146	-.394	.212	-.624	21	.540

As the participants were randomly assigned to the Stage 2 tasks, it is important to find out if any group difference in scores was simply due to the differences in the participants' writing ability and graphicacy in the first place. As shown in Table 12, the three Stage 2 task groups (A, B, C) had very similar grades in the Stage 1 tasks (GEPT-AWT1, IELTS-AWT1 and IELTS Task 2). No difference in the average grades of Stage 1 tasks between the three groups (Stage 2) was statistically significant. In other words, the three groups were similar in their English writing ability as measured by IELTS-AWT1 under normal examination condition.

Table 12: Comparison of three Stage Two task groups' performance in Stage One tests

Stage One	Min.			Max.			Mean			Std. deviation		
	A	B	C	A	B	C	A	B	C	A	B	C
GEPT	8	8	8	20	20	20	13.90	13.86	12.86	3.239	3.439	3.013
IELTS1	4.0	4.0	4.0	6.5	6.5	6.5	5.31	5.333	5.159	.6418	.7303	.8075
IELTS2	5.0	4.5	4.5	7.5	7.5	7.0	6.07	6.143	5.795	.6761	.8084	.7662

Similarly, we examined the differences in graphicacy between the three task groups (Table 13). There was no statistically significant difference between the groups either.

Table 13: Comparison of participants' graphicacy between three task groups in Stage Two

	Task A group	Task B group	Task C group
Min.	93.0	93.0	98.0
Max.	167.0	167.0	159.0
Mean	127.9	124.2	124.3
Std. deviation	23.6958	21.1209	17.9613

We are therefore confident to say that any difference in Stage 2 performance between the task groups can be attributed more to the features of the tasks rather than to the initial difference in participants' writing ability or graphicacy which could have been created artificially simply by the research design itself.

In the next section, together with the think-aloud and interview data, we report the findings from the further statistical analysis on the group differences, the effects of graphicacy, graph types and writing abilities on the participants' performance in their Stage 2 tasks.

### **Addressing the research questions**

The findings are reported in the order of the four research questions. In response to each research question, we first report the test performance data, then the think-aloud and interview data.

#### ***Research question one***

*To what extent are there differences in the cognitive processes due to different graph prompts?*

As shown in Appendices 10-15, we used four different types of graph prompts, including statistical tables (Task A: performances of Asian test takers), line graph (Task B: credit card debt in Taiwan), horizontal bar graph (Task B: age distribution of credit card debt) and vertical bar graph (Task C: carbon dioxide emissions in China), and pie chart (Task C: sources of electricity production in China). Table 8 shows that the participants' performance within-tasks (the shaded cells in the table) as well as across-tasks were statistically significantly correlated. Within-tasks, Task A IELTS explained about 40% of Task A GEPT performance; Task C IELTS explained about 54% of Task C GEPT performance; and Task B IELTS explained about 75% of Task B GEPT performance. In our view, these are substantial positive correlations between IELTS-AWT1 and GEPT-AWT2 tasks, and they are generally similar to (in the case of Task A) or higher than the correlation between the IELTS-AWT1 and GEPT-AWT2 tasks in Stage One ( $r=0.654$ , see Table 3). The higher correlations between IELTS and GEPT tasks in Stage 2 than Stage 1 are anticipated because the IELTS and GEPT tasks in Stage 2, unlike the Stage 1 tasks, used the same corresponding graphs. It is also

interesting to note that Task A GEPT and IELTS in Stage 2 which had the lowest correlation in the participants' test scores used statistical tables as prompts – the type of graphs that the participants generally found the most difficult and least familiar with (see *Participants' graphicacy*).

To what extent was the different level of correlations between IELTS and GEPT tasks in Stage 2 attributable to the use of different graph prompts? To what extent was there any significant difference in the average scores between the tasks using different graph prompts? These are two important questions to consider before we analyse the impact of graph prompts on test taking cognitive processes. As GEPT and IELTS used different rating scale, we compared the differences between the average scores of IELTS and GEPT tasks separately. Table 7 reports the average scores of all the six tasks. Due to the specific method of task allocation, it is not appropriate to run paired samples t-tests to examine the differences in means between the task groups, neither was it appropriate to run independent samples t-tests or ANOVA, because the participants for each task were neither independent nor paired equally or the same (see Table 8). We therefore ran one-sample t-tests, using the average score of the other task as a “test value” in SPSS. No statistically significant difference between the three IELTS scores was observed. In other words, the participants had comparable or similar performance across the three tasks using different graph prompts. This holds true for the GEPT total scores as well. As GEPT-AWT2 writings were also analytically marked (see Table 14), we used one-sample t-tests to analyse the differences in the average sub-scores between the three tasks (A, B, C). No statistically significant difference was noted in each GEPT-AWT2 sub-score between the three tasks, either (but see Tables 9-11 on the differences among the sub-scores of GEPT-AWT2, some pairs are statistically significant).

Table 14: Comparison of mean sub-scores of Stage Two GEPT tasks

	Relevance & Adequacy			Coherence & Organization			Lexical Use			Grammatical Use		
	A	B	C	A	B	C	A	B	C	A	B	C
Min.	2	1	2	2	2	1	2	2	1	2	1	1
Max.	5	5	5	4	4	5	5	5	5	4	5	4
Mean	3.29	3.38	3.32	2.86	2.81	3.00	2.67	3.05	2.77	2.86	3.00	2.86
SD.	.845	1.071	.839	.655	.750	.976	.856	.921	1.020	.727	.949	.889

The analysis of the test performance data suggests that there was no statistically significant difference between tasks using different graph prompts. This is true for both IELTS and GEPT graph-based tasks. In this aspect, GEPT-AWT2 seems to be comparable to IELTS-AWT1.

This finding is quite different from Yu et al. (2011) which found that there were significant differences in test takers' performance in eight IELTS-AWT1 tasks of different graph prompts (Note: the present study did not use Yu et al's tasks). In Yu et al. (2011), the highest performance was observed in a line graph task, and lowest in a task using statistical tables, which corresponded well to the test takers' highest familiarity with line graphs and the lowest

familiarity with statistical tables. The fact that there was no significant difference in test performance between graph prompts in the present study could well be attributable to the specific research design. In the present study, the participants had much longer time to process the information when they did the second of the set of two tasks which were randomly assigned to them. For example, when a participant was assigned to do Task A GEPT first, s/he would do Task A IELTS next. When s/he started to do Task A IELTS, she would have already had nearly 50 minutes in processing and re-presenting the information contained in the graphs. Therefore the comprehension process would probably be less essential for successful completion of the second task. Although we randomized the order of tasks (whether to start with IELTS or GEPT) to reduce such effects, it was inevitable in a sense that the participants would probably be much better conversed when they started to do the second task, whichever the task was.

Although the statistical analysis above suggests that the impacts of the use of different graph prompts on test scores were negligible, it does not necessarily mean that the impacts are negligible on the test taking processes. The think-aloud protocols are the window to the participants' cognitive processes of completing the tasks and can provide further essential information to understand the impacts of the use of different prompts on test taking process.

Yu et al. (2011) identified the three key stages of cognitive processes of completing IELTS-AWT1 tasks. This model was used as a guideline to analyse the cognitive processes of completing GEPT-AWT2 and IELTS-AWT1 tasks in this present research. The three key stages are: comprehending non-graphic task directions, comprehending graphic information and re-producing graphic and non-graphic information in continuous discourse. In the case of GEPT-AWT2 tasks, there is the additional stage of personal interpretations, comments and suggestions in relation to the phenomenon observed/demonstrated in the data which is presented through graphs. The recorded think-aloud protocols and interviews are the main sources of qualitative data to investigate the extent to which the differences in the cognitive processes were due to the use of different graph prompts.

The first part of the recorded think-aloud protocols suggests that there was not much difference in how the participants processed the non-graphic (i.e., the verbal) task directions across the tasks. All the participants (except Participant YY. Note: we use the initials of pseudonyms for confidentiality purpose.) started with the non-graphic task directions. Since the participants were allowed to think aloud in both Chinese and English, half of them started reading the instructions in English and then translated parts of what they might have considered important into Chinese – the first language of the participants. The other half translated the task directions directly into Chinese, perhaps simultaneously with their silent reading of the directions. Only Participant YY started with analysing the graphs straightaway for all her writing tasks.

Similar to the findings of Yu et al.'s study (2011), our participants in this study also repeatedly read or referred to (sometimes in Chinese translation) the summary-like sentences about the content of the graphs, as well as the other sentences which specify what the

participants should do for the tasks, but only in terms of the content of their writing, not in terms of the general technical requirement of the tasks (such as time allowed, and number of words required). The summary-like sentences, which provide the participants with the essential information on the topic and theme of the graphs were designed to be the same in the GEPT and IELTS tasks (e.g., “Table 1 shows the IELTS (International English Language Testing System) test taker performance by geographic region in Asia in 2011”). However, the directions of the two types of tasks differ in other aspects. The IELTS-AWT1 tasks have exactly the same sentence: “Summarise the information by selecting and reporting the main features, and make comparisons where relevant”. In GEPT-AWT2 tasks, however, the sentences specifying the similar task requirement were written in such a way that they were closely related to the specific context as well as the topic of the graphs. Take Task A GEPT as an example, the task directions read as follows:

After reading the data, you feel you have some ideas on how Taiwan test takers can do to improve their performance in both tests and decide to write a letter to an Opinion Section of a national English learning magazine in Taiwan.

In your letter:

- First, **briefly explain** why you are writing to the Opinion Section.
- Second, **summarize** what you think are the **main findings** from Tables 1 and 2, and **discuss some possible reasons for these findings**.
- Then, based on these findings, **make suggestions** about what Taiwanese test takers can do to improve their performance in both English tests.

Immediately following this requirement specific to each GEPT-AWT2 task was the more general task directions on the number of words and the expected genre of the writings (which was a letter to a special column of a newspaper), which were the same for all GEPT-AWT2 tasks.

Similar to Yu et al. (2011), the majority of our participants skipped the sentences that stated time allowance and expected length of their writings. This holds true for both GEPT and IELTS tasks. It was very likely because the participants decided that these sentences were not as useful or important as the summary-like sentences that they repeatedly read and referred to.

The difference between the participants in the stage when they processed the non-graphic task instruction is relatively minimal, compared to the next stages when they comprehended graphic information and re-produced graphic and non-graphic information in written discourse. It was at these two stages that the participants demonstrated a number of differences in their cognitive processes. (Note: the differences in the participants’ cognitive processes between GEPT and IELTS tasks are reported in *Research question four*).

The participants' think-aloud protocols and interviews (as well as their responses to the open-ended question in the graphicacy questionnaire, see Appendix 18) indicated that statistical tables were considered the most cognitively demanding among all the graph prompts used in the present study. The majority of the participants pointed out the difficulty of the statistical tables in Task A as compared to line graphs, bar charts and pie charts used in Tasks B and C. They felt that the statistical tables made the task overall more demanding because they had to compare by themselves the performance of the Asian regions to work out certain patterns or trends of the data presented or hidden in the statistical tables. Several participants commented on the extra challenges of statistical tables, for example:

Extract 1 (Participant JY)

I was baffled when I saw the numbers in the tables. I needed to calculate because I wasn't able to analyse them. I was dazzled by all the numbers. I had no idea how to make the comparison so what I did was look at the numbers one by one and said this was higher than that. Overall, I still was not sure how to read (the tables). They were a bit difficult and harder to comprehend.

Extract 2 (Participant SR)

There are several sub-units in the tables. It was difficult for me to look for all the information provided. There are so many sub-units and also so many countries. I did not know where to start from when I started reading and analyzing. It was a bit difficult.

According to the think-aloud protocols, Participant JY along with some other participants who were assigned Task A, spent considerably longer time comprehending the statistical tables than other types of graph prompts. In Task A, they read out all the mean test scores of each Asian region for each language skill and then ranked the regions from the highest to the lowest scores. The participants considered statistical tables more demanding, not necessarily because they were less familiar with tables or numerical data than other graph prompts, but rather we think it was mainly because the extra time required to process the numerical data in the tables. To complete the tasks, the participants had to process the numerical data to identify the overall patterns through a series of calculations and comparisons, much more than merely extracting a specific data point from the statistical tables (Yu et al., 2001). Thus, as some participants commented, the complexity of analysing tables resulted in the increased time they spent on Task A, which echoes Vessey's argument (1991) that it is more time-consuming to process the symbolic information provided by tabular representations.

Extract 3 (Participant YZ)

You have to calculate which is the highest score or which is the total score and it is troublesome. I have to spend more time analysing the task...More time needs to be spent on tables because there are only numbers with no other data.

Extract 4 (Participant WTR)

I think that it would be faster to get a gist of what the graphs are saying, where the graphs are leading to, for pie charts, bar or line graphs. But for tables with numbers, you would have to look at the numbers by yourself and thus, more time-consuming.

Although we did not run word frequency or other discourse analysis on the participants' writings as in Yu et al. (2011), the participants' interviews and think-aloud protocols did demonstrate that the features of graphs or graph types can determine and predict what words the participants would use as well as the content of their writings more generally. For example, in the participants' think-aloud protocols and their writings for Task A, which used two tables describing test taker performances in different regions for IELTS and TOEFL-iBT, the most commonly seen words were 'high', 'low', 'best', 'worst', 'first' or 'last'. The participants' knowledge about the conventions of graphs help them to decide what words should be used, as commented by Participant FYY when asked about whether he would do the same for different graphs in her writings.

Extract 5 (Participant FYY)

I think it's different. It is different to describe trend charts and pie charts. For pie charts, we would probably say 'something makes up for how many percent'. If there are bar graphs, we use 'hierarchy'. We would use different words.

Furthermore, some participants commented that the extent to which graph prompts can determine their lexical choice depends on the tasks. It may be more true or relevant for IELTS-AWT1 than for GEPT-AWT2 tasks.

Extract 6 (Participant YZ)

IELTS requires simple descriptions of the graphs so it's more likely to use specific words, like double, times, or increase. This is what IELTS focuses more on. But for GEPT, you have to be more comprehensive. You need not only words in that field (to describe graphs) like increase, but also some transitions, or other words according to its context.

In summary, the statistical analysis on the participants' test performance seemed to suggest that the impacts of the use of different graph prompts on test scores were negligible (c.f. Yu et al 2011 which found there were statistically significant differences between different graph prompts). The test performance data of this study also suggests that the IELTS and GEPT test scores were highly correlated regardless of what graph prompts were used in the tasks. This seemingly no-difference between different graph prompts in test performance, however, does not have a clear support from the think-aloud and interview data, which indicated that there were substantial differences in the test takers' cognitive processes attributable to the use of different graph prompts, particularly at the stage of comprehending the graphs and re-producing the graph comprehension in continuous discourse in English. At the additional stage of personal interpretations, comments and suggestions in the GEPT-AWT2 tasks, the impacts of the features and types of graph prompts on the cognitive processes seemed to fade away. In other words, we did not notice participants' direct references to their comprehension (or lack of it) or re-production of the graphs *per se* at the personal interpretation stage of the GEPT-AWT2 tasks. We will report other differences between IELTS-AWT1 and GEPT-AWT2 tasks in Research Question Four.

## ***Research question two***

*To what extent are the cognitive processes affected by test taker's graphicacy?*

A number of simple linear regressions were run to understand the relationships between graphicacy and task performance. No significant impact of graphicacy on task performance was noted. In fact, the R squared values were extremely low, which indicates that the participants' graphicacy did not seem to have any direct impact on their performance in IELTS-AWT1 or GEPT-AWT2.

However, the think-aloud and interview data present a very complex picture of the impacts of graphicacy on test taking cognitive processes, which potentially could have affected their test scores. The participants, when asked about the ways their graphicacy affected how they processed and produced their writings about the graphs, revealed that they were generally quite familiar with different types of graphs, but were probably least familiar with statistical tables (see the section on *Participants' graphicacy*), and consequentially their cognitive processes were affected. The intricate relationship between graphicacy and graph prompts makes it almost impossible to separate our discussions on the research findings to question 1 (graph prompts) and 2 (graphicacy), therefore the discussions below should be read jointly with our discussions earlier with regard to research question 1.

### Extract 7 (Participant CHB)

I think that I don't have any troubles reading graphs, very familiar (with them). I am used to talking about pictures (read from pictures).

### Extract 8 (Participant CM)

My familiarity with graphs is higher. I am less familiar with tables, with words and with decimals. Because there are some bar graphs on newspapers or you can see them in the broadcast news but fewer of those that need you to count.

### Extract 9 (Participant YH)

I was stuck because I was not familiar with tables but others were OK.

Similar to Yu et al.'s study (2011), the participants seemed less confident in analysing tables than other graphic representations such as pie charts, bar, or line graphs.

### Extract 10 (Participant YH)

I think tables are the most difficult while the pie charts are the easiest, with the bar graphs in the middle. Because the data seem messy (without patterns) in tables, as if the data are not organized. If you put these (numbers) into bar graphs, you can clearly see the numbers and calculate.

### Extract 11 (Participant JW)

I prefer pie charts because you can see the total amount, which means the percentage (of each part) is apparent, adding up to a hundred percent. This is not the case with bar graphs. Even if the



total amount (in the bar graphs) is hundred, you'd believe more in the pie charts because one can see more clearly the proportion. I prefer this.

#### Extract 12 (Participant KZ)

I personally favor graphs that show trends (referring to bar and line graphs) because I can understand better what the trends are.

The participants' familiarity with and preference towards a particular type of graph is only one of the many manifestations of their graphicacy and how they think their graphicacy may affect their test taking process and product. Another aspect related to participants' graphicacy is how they make sense of the interrelations between two graphs in the tasks. Such skills however probably had more to do with the participants' knowledge about the topic and theme of the graphs, rather than their familiarity with the type of graphs *per se*. For example, three participants had specifically commented on their problems in figuring out the connections between the two graphs in Task C: a bar graph of China's carbon dioxide emission from 1990 to 2008 and a pie chart of sources for electricity production in China in a particular year.

#### Extract 13 (Participant KY)

I got stuck with one of the task. That is, I can't figure out the connection between the 2008 chart and the other. That's why I got stuck. I don't know whether it is because of the graphs or because the link between the two graphs is not established clearly enough.

#### Extract 14 (Participant CHB)

I started with IELTS and then GEPT (for Task C). The first graph is about carbon dioxide emissions and the second is about the ratio of the different ways for electricity production. It says that coal accounts for 80%. And I think. Does the increase of the use of coal result in the increase of carbon dioxide emissions? Are they really interrelated? Is it so? The pie chart displays data for a year, 2008 but the trend chart (bar graphs) shows data from 1990 to 2008. Does that mean the 2008 data can be used to relate to carbon dioxide emissions. The GEPT one did say that they are interrelated.

The issue Participant CHB brought up suggested that the ability to make connections between the two graphs was particularly important for IELTS-AWT1 tasks, as commented by Participant EQ.

#### Extract 15 (Participant EQ)

Regarding 'comparison where relevant' (IELTS-AWT1 task instruction), I will keep thinking and reading the graphs to see if there is any difference (between the two graphs). I will not try to find their link for GEPT. I'd report what I see and the underlying reasons... I think that IELTS requires me to consistently find the connections but not for GEPT.

Such comments suggest that the different task requirements of GEPT-AWT2 and IELTS-AWT1 might have triggered or required different kinds and levels of graphicacy for successful completion of the graph-based writing tasks.

In summary, it was found that the participants' graphicacy did not have statistically significant correlations with test results. This finding holds true for both IELTS-AWT1 and GEPT-AWT2 tasks. However, the dynamic and intricate relationships between graphicacy and graph prompts as two important variables influencing the participants' test taking processes present a much more complex picture of the potential impacts of graphicacy. Furthermore, the think-aloud and interview data suggest that graphicacy may be multi-faceted in the sense that the participants' familiarity with, preference towards and engagement with a certain type of graph(s) in the writing tasks may be content- as well as task-specific. Successful completion of IELTS-AWT1 and GEPT-AWT2 tasks may require or trigger different kinds and levels of graphicacy. We argue that these two types of graph-based writing tasks may differ in their underlying construct with regard to the requirement of graphicacy in task completion, although the test results suggest that graphicacy did not have significant impacts on test performance.

### ***Research question three***

*To what extent are the cognitive processes related to test taker's writing ability?*

In Table 3 we reported some very high correlations in performance between tasks of Stage 1; and in Table 8 we reported similar correlations in performance between tasks of Stage 2. Within this background of high correlations between the tasks within Stage 1 and Stage 2, we ran a number of simple linear regressions to examine the relationships between Stage 1 and Stage 2 tasks, using Stage 1 performance as a proxy of the participants' writing ability under normal examination condition. In specific, we ran regressions on the participants' IELTS-AWT1 performance in Stage 1 and their performance in the IELTS-AWT1 tasks in Stage 2 (Appendices 19-21), between the participants' GEPT-AWT2 performance in Stage 1 and their performance in the GEPT-AWT2 tasks in Stage 2 (Appendices 22-24); and between the participants' IELTS-2 performance in Stage 1 with GEPT-AWT2 performance in Stage 2 (Appendices 25-27). According to the regressions statistics, Stage 1 performance was found to be able to predict a significant part of Stage 2 performance. It holds true for both IELTS and GEPT tasks; and was particularly true for GEPT tasks. The values of R squared were 0.293, 0.360 and 0.568 for IELTS-AWT1 (Stage 1) on IELTS-AWT1 (Stage 2, Task A, B, and C respectively); 0.572, 0.522 and 0.596 for GEPT-AWT2 (Stage 1) on GEPT-AWT2 (Stage 2, Task A, B, and C respectively); and 0.694, 0.664, and 0.603 for IELTS-2 (Stage 1) on GEPT-AWT2 (Stage 2, Task A, B, and C respectively). Details of the regression statistics are presented in Appendices 19-27. The stronger predictive power of IELTS-2 for GEPT-AWT2 is interesting, which is even stronger than the relationships between the Stage 2 IELTS-AWT1 and GEPT-AWT2 tasks, especially Task A and Task C (see Table 8, R squared values are 0.402 between Task A IELTS and GEPT, 0.747 between Task B IELTS and GEPT, and 0.542 between Task C IELTS and GEPT). It is probably because IELTS-2, rather than IELTS-AWT1, may share more similarity in its underlying construct with GEPT-AWT2 (e.g., reasoning skills as required in an argumentative essay writing are essential for both GEPT-AWT2 and IELTS-2), although IELTS-AWT1 and GEPT-AWT2 may share more surface-level task features (e.g., the use of graphs as prompts). Further discussions on the

comparability between GEPT-AWT2 and IELTS-AWT1 are reported in Research Question Four.

Below we report how test takers' English writing abilities, in terms of their knowledge in grammar, vocabulary and structure which might have shaped the way they comprehended the graphs and re-produced their graph comprehension in written discourse. It is important, however, to situate these qualitative analyses within the larger context of the differences and the correlations between the sub-scores within each GEPT-AWT2 task (see Tables 9-11), as well as between the tasks (see Research question one on the potential impacts of different graph prompts on task performance, Table 14). The statistical analysis on students' performance of the GEPT-AWT2 task under normal examination condition (see Tables 4 and 5) indicated that "lexical use" (LU) and "grammatical use" (GU) were probably the two most challenging areas as LU received statistically significantly lower scores than "relevance and adequacy" (RA) and "coherence and organization" (CO); and similarly, "grammatical use" (GU) scores was statistically significantly lower than RA. It should also be noted that these sub-scores were highly correlated (see Table 6). Similar patterns of the differences of the sub-scores within each GEPT-AWT2 task were observed (Tables 9-11). However, as we reported in Table 14, there was no statistically significant difference in each sub-score between the tasks.

From the think aloud protocols, we can see clearly that the participants were constantly monitoring their grammar and word choice or the correct usage of vocabulary. For example, Participant CM constantly monitored her use of grammar while she was writing: "Nevertheless, nevertheless, few days ago, few days ago, I see, I saw, read, I read some, I read some shock data, some shocking data which may overturn our ideas." Some participants (e.g., Participant MC) made revisions on grammatical errors and word choice (see the underlined below) after a longer period of writing as she kept going back and re-read what she had already written.

#### Extract 16 (Participant MC)

In the following letter, first, I'm going to present you two figures and I'll talk about what I found from the figures. Then, I'll discuss some possible reasons for these findings. Then, I'll make some suggestions. Erase 'and then', add 'finally'. And finally, I'll make some suggestions for the Chinese government to reduce the amount of carbon dioxide emissions. First, I'm going to present you, add 'the', two figures.

Participants were also constantly making lexical decisions as which words they should use. For example, Participant XR considered whether 'resources' or 'materials' would be more appropriate when referring to coal for electricity production. In the recorded think-aloud protocols, she said: "79.1% of the electricity production uses coal, uses coal as the resources. Is it material? No, it's not material." As the majority of them did their think-aloud in Chinese, their decisions on word choice also included which word or words in English could best represent what they expressed in Chinese. For example, Participant BY tried to figure out how he should translate "負債利率" (interest rate of debt) into English.

#### Extract 17 (Participant BY)

... cannot higher than the rate of interest, 應該是(should be) debt interest, 應該是(should be) the rate of debt interest, 應該是(should be) the rate of debt interest...

In addition, some participants also considered whether their translations or choice of words or phrases were correct, or academic-like. For example, Participant XR wondered if her translation of ‘社會新鮮人 (fresh graduates)’ to ‘social freshmen’ was an appropriate English phrase.

#### Extract 18 (Participant XR)

因為 20 到 29 歲的人通常, 通常剛出社會不久, 所以對金錢的這些還沒有很夠, 所以這個要怎麼講呢(Because people between 20 to 29 years old have just graduated so they don't have much sense of money. How am I going to say this)? social freshmen, 不知道這樣子講可不可以(I don't know if this is correct.) social freshmen, 不要, 聽起來好 chinglish (no, sounds so Chinglish). 就說他們很 young (I'd just say they are very young).

Participant YM referred to how her teacher had reminded her of an outdated usage:

#### Extract 19 (Participant YM)

然後這邊要寫說 As for me (Then, I should write 'As for me' here) 因為我的老師曾經說過, as far as I am concerned 是非常老派的寫作方法, (because my teacher once said 'as far as I am concerned' is a very old-fashioned usage), 講起話來很像現代人在講古語 (it will sound like modern people speaking old English), 所以我說 as for me (so I say, 'as for me'), 或是 for me to mean (or 'for me to mean'), 我就不用 as far as I am concerned (I would not use 'as far as I am concerned')

The interview data provide further evidences on the participants' views on the relationships between their English writing ability and performance of the GEPT-AWT2 and IELTS-AWT1 tasks. According to the interview data, two significant aspects were mentioned by the participants: lexical knowledge, and skills in organising their writings with clarity and coherence.

#### Extract 20 (Participant ZR)

Huge effect, like the knowledge of word usage, vocabulary and grammar. Sometimes you don't know what to use and can't express yourself. English writing ability doesn't affect me in reading (task instructions and graph descriptions) but does matter in my writing. There will be problems when I try to express myself.

#### Extract 21 (Participant WL)

The amount of vocabulary. I use Chinese to describe what I want to say but when I use English, I'd always get stuck with one or two words. Then you'd need to use easier ways to translate but may not get the effect that I want. That's why I think the amount of vocabulary becomes the obstacle when I write.

Extract 22 (Participant WR)

Grammar and structure in particular. It's because we need to think about the structure in a short time limit. If you have been very familiar with this, there's nothing to worry about. For me, I get stuck when I write. It feels strange. I think I should have more practice before taking such tests.

Participant MC pointed out the importance of structure for both GEPT and IELTS tasks even though she did not think her writing ability affected too much her completion of the tasks and her performances.

Extract 23 (Participant MC)

English writing ability? I don't think it influences a lot because you can just copy from the information given on the graphs. You can do it as long as you have got hold of the structure, and you know what you are doing.

There were other participants who gave similar responses as Participant MC; and we noted that the majority of them were English majors from National Taiwan Normal University, with higher English language proficiency. Among the three variables: types of graph prompts, graphicacy and writing ability, they tended to rank writing ability as the least important for their writing performance. This may suggest that these students' English writing proficiency is sufficient for them to consider graphicacy or graph types more significant than their writing ability in successfully analyzing and completing the GEPT-AWT2 and IELTS-AWT1 tasks. Quite contrarily, participants whose writing performances were comparatively lower tended to attribute their lower writing performances more to the deficiency of their writing ability than to other factors (e.g., graphicacy).

Extract 24 (Participant YJ)

Even if you are not familiar with graphs, you just need to spend time to understand the graphs. I personally prefer graphs because words (written words) are too abstract for me. Also, with the restriction of my English ability, it is difficult to think of words to write. That's why I think writing ability matters more. It turns out that I can't put what I have read into written words or what I write is incomplete. Then it's not good. In fact, the analysis is not difficult. The graphs are not difficult.

In summary, the participants' test performance data and the recorded think-aloud protocols and interviews all point to the importance of English writing ability for successful completion of the IELTS-AWT1 and GEPT-AWT2 tasks, although the level of significance of the writing ability varied among the participants, perhaps depending on their writing ability as well as the interplays between their writing ability, graphicacy and features of graph prompts. The three most frequently observed indicators of writing ability in the qualitative data included lexical knowledge and decision-making, grammar knowledge, and skills in organising writings with clarity and coherence, which correspond to some extent the analytical rating scale for GEPT-Advanced writings. The statistical analyses of the sub-scores of GEPT task performances indicate that "lexical use", which received the lowest sub-scores quite consistently across the three tasks, may be the most demanding in the rating scale or the least

developed area in the participants' writing ability. Furthermore, the strong predictive power of IELTS-2 in Stage 1 on GEPT-AWT2 in Stage 2 may suggest that the writing ability measured in IELTS-2 (topic-based argumentative essay writing) may share more similarities with GEPT-AWT2 in their underlying construct (e.g., the reasoning skills required in argumentative essay writing) than IELTS-AWT1 even though the two graph-based writing tasks share more in surface-level task features (e.g., the use of graphs as prompts).

#### ***Research question four***

*What is the comparability between GEPT-AWT2 and IELTS-AWT1 tasks in the cognitive processes attributable to the use of different graph prompts (RQ1), test takers' graphicacy (RQ2) and writing ability (RQ3)?*

Throughout our discussions in Research Questions 1-3 above, we have presented some indications of the comparability between GEPT-AWT2 and IELTS-AWT1, and GEPT-AWT2 and IELTS-2. The comparability between the tasks was examined with reference to the extent to which the variability in graph features of the tasks, the participants' graphicacy and English writing ability influenced their test-taking cognitive processes. In this section, we explore further in details the comparability between the two tasks in terms of different graph prompts, test takers' graphicacy and writing ability.

The participants' overall comparisons on their experiences in completing GEPT-AWT2 and IELTS-AWT1 tasks are quite illuminating about the comparability of the two tasks. The majority of the participants could be divided into two groups in terms of their test taking experiences. The participants who considered GEPT to be more challenging underlined the specific task feature of GEPT-AWT2: the need to discuss beyond the graphs about possible reasons and make suggestions concerning the problems.

##### **Extract 25 (Participant KZ)**

The challenges in GEPT are that we need to discuss the underlying meanings of the graphs, the reasons behind such trends, and finally, we are tested on what suggestions we can make to improve the problems. So I think GEPT is difficult concerning these aspects, background information. GEPT advanced is more challenging, when it comes to involving background information.

##### **Extract 26 (Participant XR)**

I think GEPT is more challenging because the expected length of the writing is longer and there are more aspects that require me to think, such as possible reasons or to give suggestions. For IELTS, I can go ahead and writing what I see from the graphs.

On the other hand, those who felt themselves more challenged by IELTS-AWT1 pointed out the difficulty of integrating data and finding relations and patterns solely from the graphs.

Extract 27 (Participant ZM)

IELTS is more challenging to me because I don't know to which direction the data is leading me to (referring to the theme or focus of the writing)...You have to select from the data and make interpretations. Other than that, you have to collect and sort out the related data to draw a conclusion as the task requires to make comparison where relevant.

Nevertheless, other participants like Participant WR acknowledged that there were different facets of challenges posed by the two tests.

Extract 28 (Participant WR)

I think the two tests are challenging in different ways. IELTS challenges you to detailed analysis but with less time pressure. For GEPT, you can write a lot more. I think that you can pick the points you want to focus on and make examples. This is what I consider as less challenging. Yet, you don't have enough time to write so many stuffs and you cannot think of ideas in such a short time. Sometimes, when you want to come up with ideas at the last moment, you can't think of any or should I say I don't have enough ideas to meet the expected length. For me, I feel hurried by the time limit.

In RQ1, we discussed the effects of different graphic prompts on the cognitive processes of completing GEPTS-AWT2 and IELTS-AWT1 tasks, highlighting that the complexity in analyzing and integrating data from two statistical tables made them the most challenging graph prompts in this study (see also Yu et al. 2011). However, the participants claimed that there was little difference in the way they analysed the same tables for the GEPT and IELTS tasks even if the tables were indeed more cognitively demanding. Nevertheless, some participants suggested that their graphicacy, the comprehension of graphs in general, was more likely to affect their performance on IELTS-AWT1 than on GEPT-AWT2.

Extract 29 (Participant CM)

Because I have to write completely based on the data for IELTS but for GEPT, I only need to select specific information required by the task. I don't need to care about other information and I can write whatever I can write. For example, we were asked to calculate and compare the English proficiency levels of the six countries. For the GEPT task, I was to give suggestions to Taiwanese people. The only few pieces of information I needed to know were Taiwanese test takers performed neither poorly nor very well, which countries ranked before us and Taiwanese test takers' weakness on writing and listening. Then I could give suggestions. Yet, with the same information, I would not be able to complete the IELTS task. I would need to refer to (the countries with) the highest or the lowest scores to complete (the expected length)

Extract 30 (Participant JW)

I think graphicacy influences more on IELTS because if you don't understand what the graphs are about, then you are doomed. As for GEPT, you can make things up. The length allows you to (make things up) since only some information from the graphs is needed (to complete the task).

As GEPT and IELTS tasks used the same graphs, whether a participant started with a GEPT or an IELTS task (randomized by design) could have some impacts on the participants' cognitive processes of doing the subsequent IELTS or GEPT task. As we argued in *Research question one*, such order effects might be inevitable. However, the participants' views on the order effects are not unanimous as shown Extracts 31-33.

Extract 31 (Participant JW)

I think it was more difficult for me to start with GEPT for my second round was due to the fact that IELTS required us to use 150 words for describing two graphs, without including reasons and suggestions. Thus, we would have to be thorough in comprehending the graphs. We have to find similarities and differences and describe in details and also how we can compare. As for GEPT (if we started with IELTS task first), there were three parts. The only part related to graph description was the second part so it would be fast (meaning speed up the writing process).

Extract 32 (Participant CHA)

I think it would be easier to start with IELTS then GEPT...The analysis of graphs is done in IELTS task so you can concentrate on other parts when you are writing the GEPT task. You won't need to spend time analyzing the graphs...It seems that I can't complete the GEPT task if I start with it first.

From a quite opposite point of view, Participant BY commented:

Extract 33 (Participant BY)

If you finish GEPT then you don't need to think more when you do IELTS. But vice versa, you'd have to think about reasons additionally. For my second round, I started with IELTS. I didn't think about the questions that would be asked afterwards (referring to 'possible reasons' and 'suggestions') and directly went ahead to write. That was why I almost skipped thinking about the graphs when I did GEPT (but spent more time thinking about other aspects.)

In *Research question three* we reported how the participants' English writing abilities might have influenced the way they re-produced their graph comprehension in continuous written discourse in English in IELTS and GEPT graph-based writing tasks. Here, we further explore whether the influences on GEPT-AWT2 and IELTS-AWT1 tasks was similar or different. As some participants commented, their English writing ability seemed to play a more important role on their completion of the GEPT task.

Extract 34 (Participant WF)

GEPT, because GEPT requires you to express more, perhaps on suggestions. Then you'd discover the problems of the two graphs or the difference (between the two graphs). These are all additional thoughts (beyond the graphs) and you have to put these thoughts into words. But you may not be able to write what you think. IELTS only asks you to explain about the graphs. You can find some of the words from them (the graphs, the task instructions).

GEPT-AWT2 asks test takers to extend beyond the information contained in the graphs as well as their graph comprehension, which the participants considered demanding to their



writing ability. One participant, who had high writing proficiency as evidenced in her performances in the Stage 1 and 2 tasks, explained:

Extract 35 (Participant MC)

There are some information that cannot be found from the graphs in GEPT. You'd have to think by your own, to write an outline...content-wise, things are fixed with graphs, which means you can find the information up there...(However), things like reasons, causes or suggestions. You'd have to think yourself and English writing ability can be influential.

The other test feature that is specific to GEPT-AWT2 tasks, the letter format, may also demand for better writing ability, as one participant claimed.

Extract 36 (Participant TR)

GEPT requires you to write a complete article, which means it needs to have an introduction, transition and then ending. Why I am writing this letter and then I analyse, objective analysis then I express my thoughts. IELTS, on the other hand, is like a short paragraph or two, with only analysis...If you have been practicing English writing regularly, you'd be able to do this task well...If you don't practice writing, you may be able to handle IELTS task but GEPT is like a complete (letter).

Her claim that regular practice of English writing (not specified as which types of writing) may help to perform well on GEPT-AWT2 tasks is interesting. The link she made between practicing writing a complete article in English and the GEPT-AWT2 task is another piece of evidence suggesting that the underlying constructs of GEPT-AWT2 and IELTS-2 (argumentative essay writing) may be more similar than what they appear to be.

On the other hand, some participants argued that their English writing ability would have played a more significant and important role in their successful completion of IELTS-AWT1 than GEPT-AWT2, because in their view IELTS-AWT1's focus on graph description is more formal and academic-like and therefore more challenging than the letter format of the GEPT-AWT2 task.

Extract 37 (Participant YJ)

IELTS requires us to analyse the graphs so we need specific lexical knowledge. The part that I have trouble writing is the part to describe the graphs because it requires us to make in-depth descriptions with details. But if you don't know how to describe, you'd have to beat about the bush, which may result in what you have written is not what the graphs are about. As for GEPT, being colloquial is ok.

Extract 38 (Participant AA)

IELTS is apparently like an essay, in which you should write and describe with clarity.

Extract 39 (Participant XR)

Because I think that IELTS is more formal. It's like writing for a journal.

In addition to the expected formal register, the lack of detailed task instructions in IELTS-AWT1 posed another challenge to the participants' writing ability; while the clear instructions on the three steps or components of the GEPT-AWT2 writings might have lent some much-needed scaffolding to the participants for them to have an easier and quicker start in planning and organising their writings.

Extract 40 (Participant ZR)

IELTS, from the tasks I have done today, should be IELTS. It's because it doesn't give me a direction of which I can follow...As for GEPT, I can write according to the procedures it has prescribed, pick out the association of the graphs and then go on and write.

Extract 41 (Participant SR)

(Writing ability) affects more on me writing IELTS because GEPT has set the framework for me. It has stated what to write for the first, second and the third part.

The abovementioned suggests that the difference in non-graphic task instructions between GEPT-AWT2 and IELTS-AWT1 tasks may account for how differently English writing ability affected the participants' cognitive processes when completing the tasks.

In this study, the surface-level comparability of the GEPT-AWT2 and IELTS-AWT1 tasks is based on the fact that both use graph as prompts to measure test takers' writing abilities. However, some participants believed that GEPT-AWT2 tasks were assessing more than their graph comprehension and re-production of graph comprehension in writing, as Participant BY commented.

Extract 42 (Participant BY)

I don't know why I should write to it (opinion section). Why should I write this part? It's weird. I don't understand why...It's not necessary. This part is very strange. Is it because this is GEPT? But that's nothing to do with your ability to understand and interpret graphs, nothing. Then why do I need to briefly explain (why I am writing to the opinion section).

Others argued that background or contextual knowledge is a lot more important for the GEPT-AWT2 than IELTS-AWT1 tasks.

Extract 43 (Participant KZ)

For me, the lack of background knowledge is what hinders my writing. I have certain level of English writing proficiency. But if background knowledge is needed and I don't know what to write about, my writing ability may be affected as well. Background knowledge is really important...When writing GEPT task, I spend time on (recalling) background knowledge then I think about how to describe the reasons and make related suggestions.

Extract 44 (Participant EQ)

I think when you analyse the graphs you would need to know the background. It is without forms (unlike graphs). It's like in history lessons, you need to know the reasons behind the incident and its influence. These are beyond the graphs...GEPT took me a lot of time to complete because I think I don't have enough background knowledge.

The participants' comments on the importance of background knowledge for successful completion of GEPT-AWT2 tasks raise a question on the underlying construct of the GEPT-AWT2 tasks as well as a question on the comparability between GEPT-AWT2 and IELTS-AWT1 or indeed between GEPT-AWT2 and IELTS-2 (or any other argumentative essay writing tasks). In addition to having to rely on background knowledge to figure out the reasons and make reasonable or logical personal interpretations, suggestions and comments, some participants believed that the extra length (in terms of number of words) required in GEPT-AWT2 also makes the test more challenging than IELTS-AWT1.

In summary, there were noticeable differential impacts of features of graph prompts, the participants' graphicacy and English writing ability on their test taking cognitive processes in the two graph-based writing tasks. The participants found the IELTS-AWT1 and GEPT-AWT2 tasks challenging, but in different ways or aspects. Generally speaking, test takers' graphicacy and graph comprehension were considered more essential for their successful completion of IELTS-AWT1 than GEPT-AWT2 tasks. The personal interpretations of the graphic information in relation to test takers' background knowledge made the GEPT-AWT2 more challenging for some participants, even though the GEPT-AWT2 tasks provide clear directions (i.e., the three steps/guidelines) on what test takers should include in their writings. However, the participants who considered IELTS-AWT1 more challenging looked at the issue from a different perspective. In their view, the more formal register (i.e., more academic-like) as implied by IELTS and the restriction to graph analysis and description only (in other words, no personal interpretations or comments) made the IELTS-AWT1 task more challenging than GEPT-AWT2, because the IELTS task was either too open for them to find a focus or too limited to expand on issues beyond what available in the graphs. This challenge is further complicated in IELTS-AWT1 because it lacks the three steps as specified in the GEPT-AWT2 task directions. The comparability between IELTS-AWT1 and GEPT-AWT2 may remain at the surface-level feature, that is, both tasks use graphs as prompts; however, our findings on the greater role of background knowledge and scientific reasoning skills required in GEPT-AWT2 than IELTS-AWT1 suggest that GEPT-AWT2 may be more comparable in terms of its underlying construct with IELTS-2 or other argumentative essay writing tasks.

## Conclusion

This research investigated the comparability between GEPT-AWT2 and IELTS-AWT1 tasks. In specific, it examined the extent to which test takers' performance and cognitive processes were affected by their graphicacy, English writing ability, and the use of different graph prompts (bar, line graph, pie chart, statistical tables, etc.). Thirty-two potential GEPT-Advanced test takers in Taiwan participated in this study as volunteers. Data were collected from three distinct stages. In Stage 1, the participants' graphicacy and their writing ability under normal examination condition were collected as baseline data. In Stage 2, the participants completed four writing tasks (2 IELTS-AWT1 and 2 GEPT-AWT2) randomly assigned to them out of 6 tasks, while thinking-aloud their writing processes. In Stage 3, post-test interviews were conducted with all the participants, either on one-to-one basis, in pairs, or in threes.

Below we recap the main findings in relation to the four research questions.

Firstly, the participants' test performance data indicated that the impacts of the use of different graph prompts on test scores were negligible and the IELTS and GEPT test scores were highly correlated regardless of what graph prompts were used in the tasks. The use of different graph prompts were substantially related to the test takers' cognitive processes, particularly at the stage of comprehending the graphs and re-producing the graph comprehension in writing. At the additional stage of personal interpretations, comments and suggestions in the GEPT-AWT2 tasks, the impacts of the features and types of graph prompts on the cognitive processes seemed to fade away, while their background knowledge stepped in for the successful completion of the GEPT-AWT2 tasks. Secondly, the participants' graphicacy did not have significant correlations with their IELTS-AWT1 or GEPT-AWT2 test results, but the complex and intertwining influences of graphicacy and graph prompts on test taking cognitive processes indicate that successful completion of IELTS-AWT1 and GEPT-AWT2 tasks may require different kinds and levels of graphicacy. Thirdly, all the quantitative and qualitative data demonstrate the importance of English writing ability for successful completion of the IELTS-AWT1 and GEPT-AWT2 tasks, although in different strengths and shapes or forms. Lexical knowledge and decision-making, grammar knowledge, and skills in organising writings with clarity and coherence were considered the most essential for successful completion of the tasks. Fourthly, in terms of the comparability between the two graph-based writing tasks, there were a number of noticeable differential impacts of graph prompts, graphicacy and English writing ability on test taking cognitive processes. Graphicacy and graph comprehension were considered more essential for successful completion of IELTS-AWT1 than GEPT-AWT2 tasks. The comparability between IELTS-AWT1 and GEPT-AWT2 may remain at the surface-level, that is, both tasks use graphs as prompts; however, our data suggest that the writing ability measured in the topic-based argumentative essay writing task – IELTS-2 may share more similarities with GEPT-AWT2 in their underlying construct. Furthermore, it should be pointed out that GEPT-AWT2 tasks may require additional processes (e.g., interpreting and commenting on

the data) but the additional processes do not necessarily mean that GEPT-AWT2 is more difficult or challenging than IELTS-AWT1. Both graph-based tasks are challenging, but for different reasons and in different aspects.

Conceptually speaking, these findings present a complex picture of the impacts of graph prompts, graphicacy and writing ability on test taking process as well as test scores, and equally important, they point to a number of similarities and differences between GEPT-AWT2 and IELTS-AWT1 tasks in both test taking process and product. Methodologically, these findings, in particular, on the subtle but noticeable differences between the test results data and the test taking cognitive processes data highlight the importance and usefulness of examining test takers' cognitive processes, in addition to test scores, when we conduct any test comparability or alignment studies.

Further detailed analysis on the discourse features of the written scripts, as in another LTTC-GEPT funded research (Qian, personal communication) and Yu et al (2011), would provide further evidence on the comparability of the two graph-based writing tasks. In terms of data collection methods, the use of think-aloud could have affected or even altered the way that the participants would normally do in a normal examination condition (even after proper training); therefore, the use of unobtrusive eye-trackers to record and analyse test takers' eye-movements (see Yu, He and Isaacs, in progress) would be essential to better understand how test takers process and use the graphic and non-graphic information in the tasks.

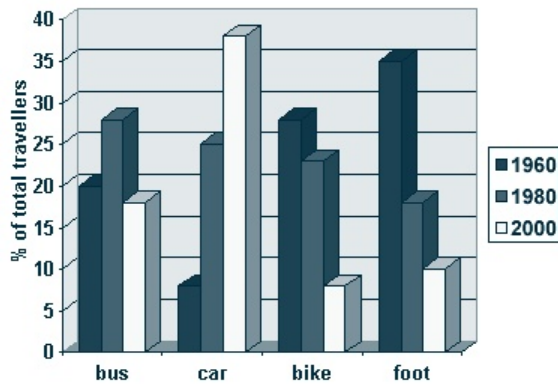
## References

- Ginther, A. (2002). Context and content visuals and performance on listening comprehension stimuli. *Language Testing*, 19(2), 133-167.
- Green, A. (1998). *Verbal Protocol Analysis in Language Testing Research: A Handbook*. Cambridge: Cambridge University Press.
- Katz, I. R., Xi, X., Kim, H.-J., & Cheng, P. C. H. (2004). *Elicited Speech From Graph Items on the Test of Spoken English*. ETS Research Report 74. Princeton, NJ.: Educational Testing Service.
- O'Loughlin, K., & Wigglesworth, G. (2003). Task Design in IELTS Academic Writing Task 1: The Effect of Quantity and Manner of Presentation of Information on Candidate Writing. In R. Tulloh (Ed.), *IELTS Research Reports* (Vol. 4, pp. 89-130). Canberra, Australia: IELTS Australia Pty Limited.
- Vessey, I. (1991). Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22(2), 219-240.
- Wainer, H. (1992). Understanding Graphs and Tables. *Educational Researcher*, 21(1), 14-23.
- 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*. Taipei: LTTC-GEPT.
- Xi, X. (2005). Do visual chunks and planning impact performance on the graph description task in the SPEAK exam? *Language Testing*, 22(4), 463-508.
- Xi, X. (2010). Aspects of performance on line graph description tasks: influenced by graph familiarity and different task features. *Language Testing*, 27(1), 73-100.
- Yang, H.-C. (2012). Modeling the relationships between test-taking strategies and test performance on a graph-writing task: Implications for EAP. *English for Specific Purposes*, 31(3), 174-187.
- Yu, G., Rea-Dickins, P. M., & Kiely, R. (2011). The cognitive processes of taking IELTS academic writing task one *IELTS Research Reports Volume 11* (pp. 373-449). London: IDP: IELTS Australia & British Council.

## Appendix 1: A sample of IELTS Academic Writing Task One

You should spend about 20 minutes on this task.

The graph below shows the different modes of transport used to travel to and from work in one European city in 1960, 1980 and 2000.



Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

Source:

[http://www.ielts.org/test\\_takers\\_information/test\\_sample/academic\\_writing\\_sample.aspx](http://www.ielts.org/test_takers_information/test_sample/academic_writing_sample.aspx),  
Accessed on 1<sup>st</sup> November 2010)

## Appendix 2: A sample of GEPT Advanced Level Writing Task Two

A local English newspaper has just printed some worrying statistics on the traffic accidents that occurred in the downtown area in June. The data are shown in Figures 1 and 2 below. As a citizen, you would like to help improve the situation.

Write to the Opinion section of this local English newspaper:

- Firstly, summarize what you think are the **main findings** from the reported data and discuss the possible causes.
- Secondly, make suggestions about what can be done to reduce the number of accidents in the downtown area.

Your report must be about **250 words**. You have **45 minutes** to complete Task 2.

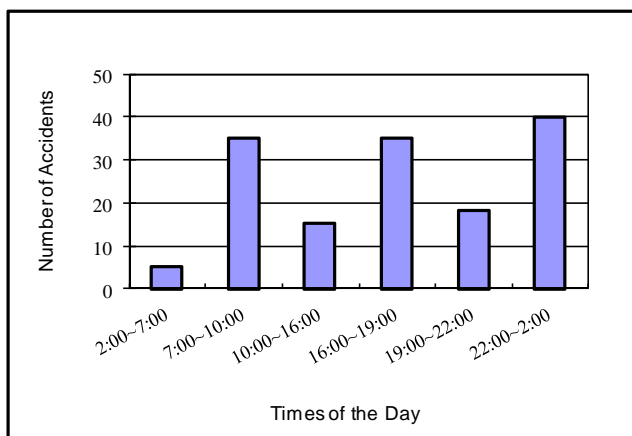


Figure 1. Number of accidents in the downtown area in June

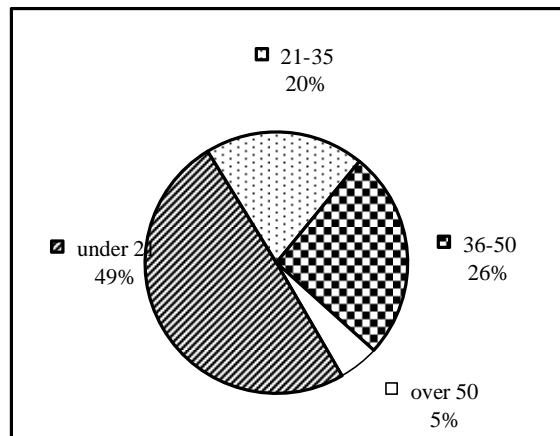


Figure 2. Percentage of accidents involving drivers in different age groups

Source: <http://www.lttc.ntu.edu.tw/geptad97/writing/writing.htm> accessed on 1<sup>st</sup> November 2010



### Appendix 3: Rating scale of GEPT Advanced Writing

	1	2	3 (Pass)	4	5
Relevance and Adequacy (RA)	Parts of the task are not addressed or are unclear. The content is inadequate or lacks relevance.		All parts of the task are clearly addressed. The content is relevant and adequate.		All parts of the task are effectively addressed. The entire content is relevant and more than adequate.
Coherence and Organization (CO)	The text lacks logical organization and coherence which may lead to confusion.		The text is coherent and logically organized in general.		The text is coherent and logically organized throughout.
Lexical Use (LU)	The range of vocabulary is inadequate to complete the task and/or lexical items are often used inappropriately.		The range of vocabulary is adequate to complete the task and lexical items are used appropriately most of the time.		A wide range of vocabulary is used to complete the task effectively.
Grammatical Use (GU)	Structures used are often inaccurate and/or inappropriate.		Structures used are accurate and appropriate for the most part.		A wide range of structures is used to complete the task accurately, appropriately and effectively.

## Appendix 4: Consent form in Chinese

### 學生同意信

親愛的同學們，

謝謝您同意參與財團法人語言中心所資助的全民英檢研究計畫，此次研究計畫是由英國布里斯托大學(University of Bristol)的研究人員進行資料的蒐集與分析，這個計畫的主旨在於了解與試者在進行全民英檢高級寫作第二部分考題與雅思考試學術類寫作第一部分考題時，解答圖表題目的認知和答題的過程，並進行比較。此研究蒐集資料的過程，您將參與的部分如下：

1. 進行測驗：全民英檢高級寫作第二部分圖表題目與雅思考試學術類寫作第一部分圖表題目
2. 填寫問卷：此問卷用來測量您對於圖表的熟悉度與理解度
3. 進行測驗並同時進行有聲思考：描述在面對全民英檢高級寫作第二部分圖表題目與雅思考試同類型題目時的認知與答題過程
4. 進行一對一或小組面談

您在進行有聲思考與面談時，將會有人錄音。為了感謝您的參與，我們將提供您 實驗獎金，金額不多，但參與此研究對於即將參加全民英檢或雅思考試的您，也會有實質的幫助。您有權利在任何您想要退出的時候退出此研究計畫，沒有任何的附加條件，但我們仍希望您能全程參與，直到整個資料蒐集的程序結束。

我們依照國際語言評量學會與英國應用語言學學會所推薦的研究倫理規範，在此正式要求您的同意，這次研究所蒐集到的所有資料（包含您的年齡、性別、圖表熟悉度、考試表現、有聲思考與面談的錄音檔等），將以匿名的方式存檔，我們將此研究的資料用公正和尊重的態度，以研究報告、學術發表等的方式呈現。我們將嚴格遵英國數據保護法(1998)來保護您的資料。

在仔細閱讀完此同意書后請在以下簽名處簽名

姓名：\_\_\_\_\_ 簽名：\_\_\_\_\_ 日期：\_\_\_\_\_

若您有任何關於此研究或此同意書的疑問，請與我聯繫。

敬祝，學業進步。

林淑雯 [dl2628@gmail.com](mailto:dl2628@gmail.com), [andrelsw@ntunhs.edu.tw](mailto:andrelsw@ntunhs.edu.tw)  
財團法人語言訓練測驗中心全民英檢研究計畫

## Appendix 5: Stage One GEPT task

Figure 1 shows the number of vehicle accidents in Taiwan (2000-2011), and Figure 2, the percentage of accidents involving drivers of different age groups in one major city in 2011. After reading the data, you are concerned and would like to advise young drivers on how to avoid vehicle accidents and decide to write a letter to an Opinion Section of an English newspaper in Taiwan.

In your letter:

- First, **briefly explain** why you are writing to the Opinion Section.
- Second, **summarize** what you think are the **main findings** from Figures 1 and 2, and **discuss some possible reasons for these findings**.
- Then, based on these findings, **make suggestions** to young drivers on how to avoid vehicle accidents.

Your letter **should be at least 250 words** in length. You have **50 minutes** to complete this task.

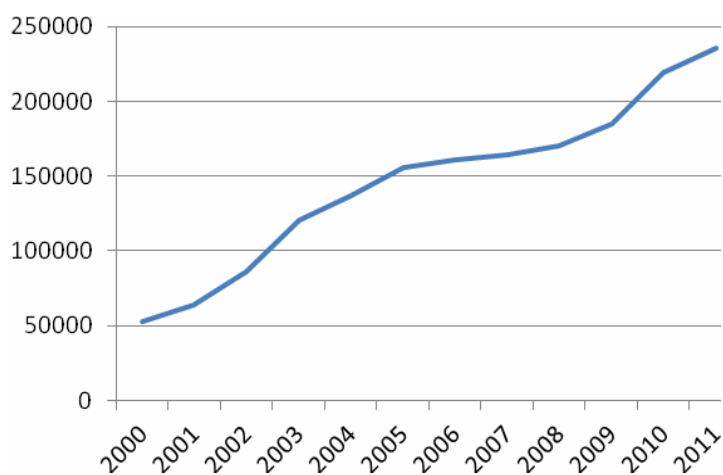


Figure 1: Number of vehicle accidents in Taiwan (2000-2011)

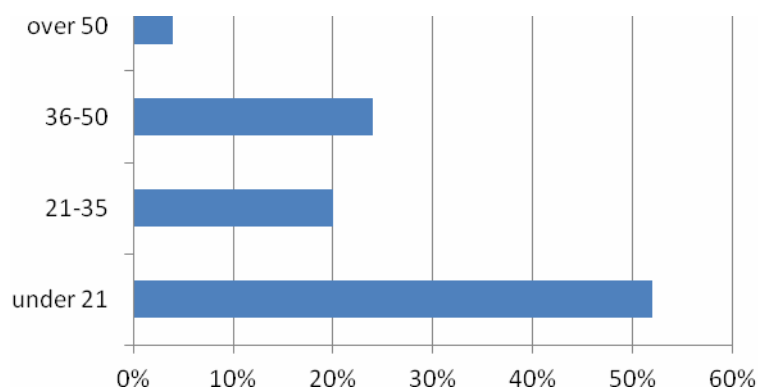


Figure 2: Percentage of vehicle accidents involving drivers of different age groups in one major city (2011)

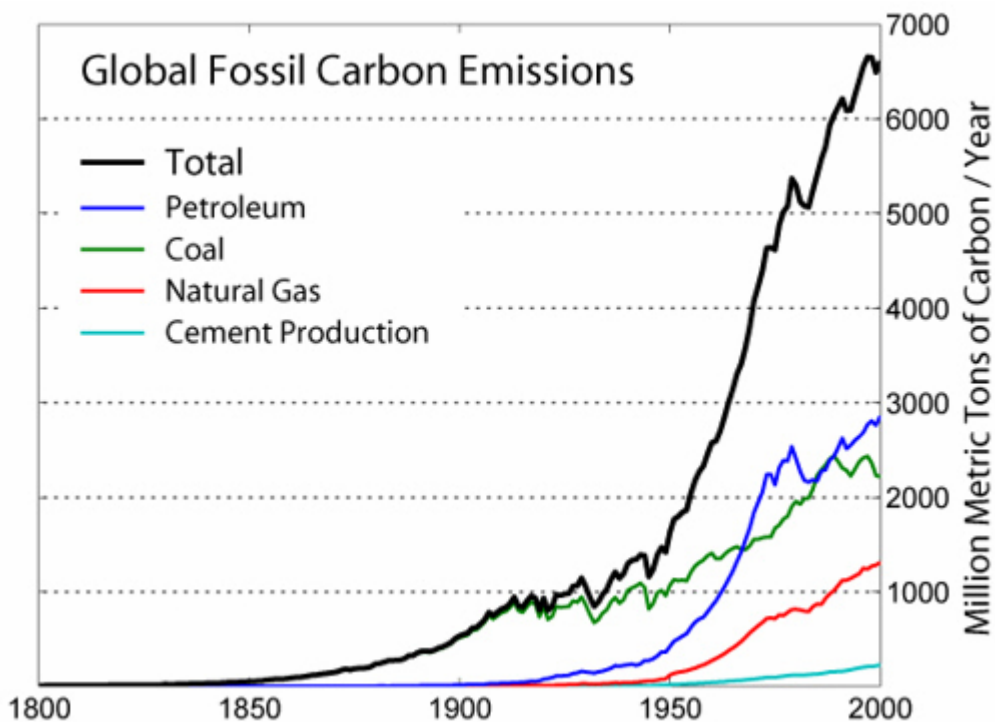
## Appendix 6: Stage One IELTS AWT1 task

You should spend about 20 minutes on this task.

The following graph shows the global fossil carbon emissions from 1880 to 2000.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.



**Appendix 7: Stage One IELTS independent writing task**

**You should spend about 40 minutes on this task.**

Write about the following topic:

*Once children start school, the teachers would have more influence in their intellectual and social development than parents.*

*To what extent do you agree or disagree?*

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

**Write at least 250 words.**

## Appendix 8: Stage One Graphicacy questionnaire in Chinese

### 問卷

此問卷將會收集您的某些個人資料以及您對各類圖表（包括數字統計表格、示意圖、流程圖等）的熟悉程度。請勾選或填空最符合您個人現況的答案，獨自完成問卷，答案不分對錯。

答題範例: 男性 (✓)

若有任何疑問，請詢問在場的研究員。

個人資料

您的手機號碼: \_\_\_\_\_

您的 email 地址: \_\_\_\_\_ (請大寫)

您的姓名: \_\_\_\_\_

您的性別: 男性 ( ) / 女性 ( )

大學/學院/ 系所 / 科系 \_\_\_\_\_

年級: (1) / (2) / (3) / (4) / (5) / (6) / (7)

大學部 ( ) / 研究所 ( ) → 若是研究所，是碩士 ( ) 或博士 ( )

您曾經考過全民英檢嗎? 有 ( ) / 沒有 ( )

若您答有，請問您最近的一次何時考的? ( )，寫作成績是: ( )

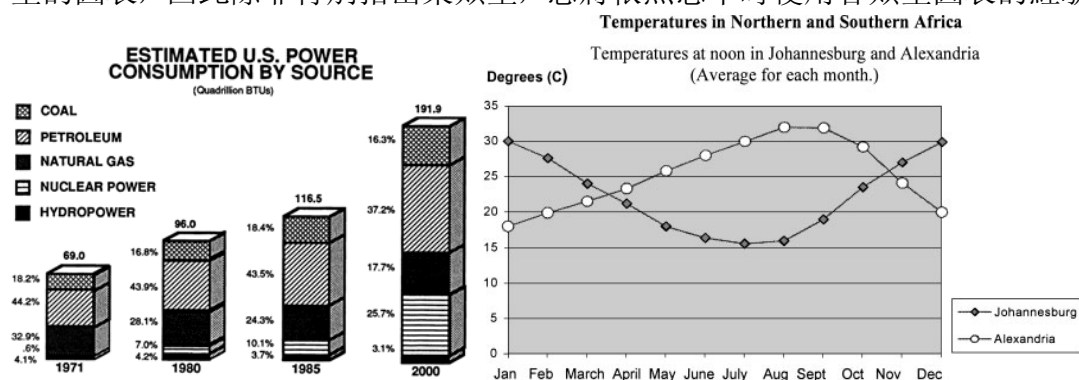
若您答沒有，請問您會在最近半年內考嗎? 是 ( ) / 否 ( )

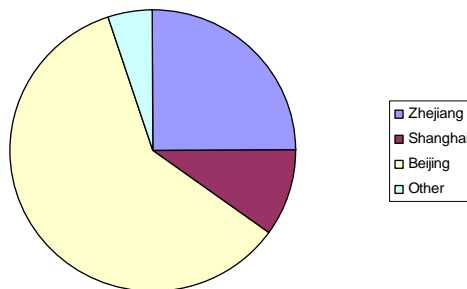
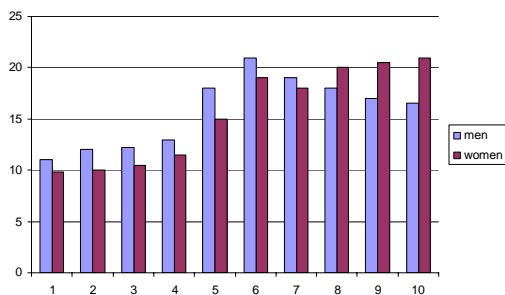
你曾經考過雅思考試嗎? 有 ( ) / 沒有 ( )

若您答有，請問您最近的一次何時考的? ( )，寫作成績是: ( )

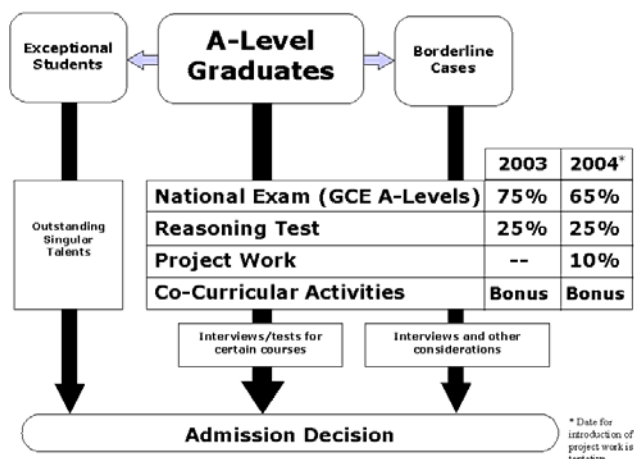
若您答沒有，請問您會在最近半年內考? 是 ( ) / 否 ( )

以下的句子將跟您對各類圖表（包括數字統計表格、示意圖、流程圖等）的熟悉程度。我們於下方提供了六種圖表作為範例，這份問卷中，我們將用『圖表』兩字來代表各類型的圖表，因此除非特別指出某類型，您將依照您平時使用各類型圖表的經驗來回答。






Countries	packaging	
	Tonnes exported in bags	Tonnes exported in containers
China	652	2001
India	4361	5002
New Zealand	82	44032



請勾選一個最適合用來描述您現況的一個號碼，答案不分對錯。

從不 
→
 經常  
 [1] [2] [3] [4] [5] [6]

1. 我使用專門的電腦軟體來繪製圖表。 [1] [2] [3] [4] [5] [6]
2. 做為我學業(或學術研究)的一部分，我需要繪製圖表。 [1] [2] [3] [4] [5] [6]
3. 做為我學業(或學術研究)的一部分，我需要看圖表後做詮釋。 [1] [2] [3] [4] [5] [6]
4. 我在大眾傳播媒體中(例如雜誌、報紙等)閱讀圖表。 [1] [2] [3] [4] [5] [6]
5. 當我閱讀圖表時，我試著找出圖表所呈現的主要趨勢或大概內容。 [1] [2] [3] [4] [5] [6]
6. 當我閱讀圖表時，我試著去思考造成這些主要趨勢或大概內容背後的可能原因。 [1] [2] [3] [4] [5] [6]

7. 當我閱讀圖表時，我不會研究圖表的細節內容。  
[1] [2] [3] [4] [5] [6]
8. 當我在大眾傳播媒體的文章裡看到圖表時，我會忽視或略過圖表。  
[1] [2] [3] [4] [5] [6]
9. 當我在學術文章裡看到圖表時，我會忽視或略過圖表。  
[1] [2] [3] [4] [5] [6]
- 強烈不同意  強烈同意
10. 我熟悉長條圖。  
[1] [2] [3] [4] [5] [6]
11. 我熟悉線圖。  
[1] [2] [3] [4] [5] [6]
12. 我熟悉圓餅圖。  
[1] [2] [3] [4] [5] [6]
13. 我熟悉顯示過程的流程圖。  
[1] [2] [3] [4] [5] [6]
14. 我熟悉呈現數字資料的表格。  
[1] [2] [3] [4] [5] [6]
15. 當我看大眾傳播媒體文章時，我能發現有關圖表的錯誤或錯誤的詮釋。  
[1] [2] [3] [4] [5] [6]
16. 當我看我的領域的學術文章時，我能發現有關圖表的錯誤或錯誤的詮釋。  
[1] [2] [3] [4] [5] [6]
17. 我知道圖表中的不同構成要素（例如 X、Y 軸、說明、顏色）。  
[1] [2] [3] [4] [5] [6]
18. 我知道圖表中不同構成要素，如何相輔相成來呈現資料。  
[1] [2] [3] [4] [5] [6]
19. 我知道圖表中圖表與其代表的數字資料之間的關聯性。  
[1] [2] [3] [4] [5] [6]
20. 我能識別單一圖表中所呈現的關係或類型。  
[1] [2] [3] [4] [5] [6]
21. 我能識別在類似主題下，多個圖表中所呈現的關係或類型。  
[1] [2] [3] [4] [5] [6]
22. 我知道在呈現某些資料時，一種圖表比另一種圖表更好。  
[1] [2] [3] [4] [5] [6]
23. 我能認出建構差的圖表。  
[1] [2] [3] [4] [5] [6]
24. 我能修改並使得建構差的圖表變好。  
[1] [2] [3] [4] [5] [6]
25. 我能以文字來描述圖表所呈現的走勢或大概內容。  
[1] [2] [3] [4] [5] [6]
26. 我能使用圖表來形容或表達數字資料所呈現的走勢或大概內容。  
[1] [2] [3] [4] [5] [6]
27. 我覺得圖表可以很生動地呈現數字資料。  
[1] [2] [3] [4] [5] [6]
28. 我覺得圖表幫助我記憶數字資料中的重要資訊。  
[1] [2] [3] [4] [5] [6]
29. 圖表在文章裡浪費空間。  
[1] [2] [3] [4] [5] [6]
30. 我擔憂因為我對於詮釋圖表並不擅長，因此無法在全民英檢高級複試的寫作考題中，完整地證明我的寫作能力。  
[1] [2] [3] [4] [5] [6]
31. 在全民英檢高級複試的寫作考題中，熟悉的圖表比起不熟悉的圖表，可以讓我表現得更好。  
[1] [2] [3] [4] [5] [6]
32. 我比較希望在全民英檢高級複試的寫作考題中，就只使用一種圖表出題。  
[1] [2] [3] [4] [5] [6]
33. 在詮釋圖表上能受到專門的訓練，將有助於我在全民英檢高級複試的寫作考題中得到更高分。  
[1] [2] [3] [4] [5] [6]





## Appendix 9: Stage One think-aloud training

In this project, you will be asked to think-aloud while doing the GEPT Advanced Writing Task 2 (GEPT-2) and IELTS Academic Writing Task One (IELTS-1). The main purpose of collecting your think-aloud protocols is to understand your test-taking process.

### How to do think-aloud:

- **The most important thing is to keep talking**, i.e., verbalizing what you are doing during the whole process including:
  - ❖ **what you are reading,**
  - ❖ **what you are thinking** and
  - ❖ **what you are writing.**
- **You can use English and/or Chinese.**
- **The researcher will remain silent unless you stop talking for more than 10 seconds.** In that case, the researcher will show you a white paper/card with TALK written on it to remind you to talk.
- **You are required to start recording with:**
  - ❖ My name is ?????.
  - ❖ It is now ??? (time), on ??? (date).
  - ❖ The task is ??? (Read the first line of the page, e.g., GEPT-Task-A, GEPT-Task-B, GEPT-Task-C, or IELTS-Task-A, IELTS-Task-B, IELTS-Task-C)
  - ❖ Then think-aloud when you do the writing task.

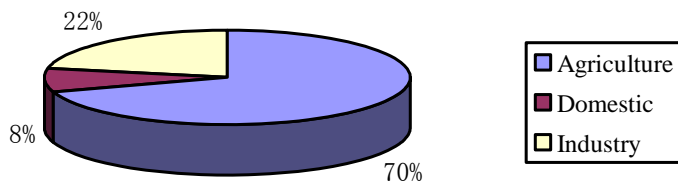
### Practise think-aloud:

#### (1) Some maths tasks for you to practise think-aloud

- $124 + 3546 = ?$
- $124 \times 378 = ?$
- The average mass of 3 parcels is 6 kg. Two of the parcels have a mass of 4.6 kg and 6 kg. Find the mass of the third parcel.

#### (2) The following is an example of a think-aloud protocol when doing the IELTS Academic Writing Task One.

World water use, 2000



Water use by selected countries, 2000

	China	India	New Zealand	Canada
Agriculture	69%	92%	44%	8%
Domestic	9%	5%	46%	12%
Industry	22%	3%	10%	80%

The think-aloud protocol when doing the task above.

... The chart and table below give information about the way in which water was used in different countries in 2002, 2000. So we have a chart and a table. And then I have to summarise the information by selecting and reporting the main features, and make comparisons where relevant. So the tasks are the same. But we have different graphs. One is a pie chart, the other is a table. Now I am looking at the pie chart. The title of the pie chart is World Water Use, 2000. And we have 3 parts, agriculture, domestic and industry. And...ok, and each has different proportions. And then I am looking at the second graph which is a table. It is about water use in selected countries in the year 2000. So we have China, India, New Zealand, and Canada in the agriculture, domestic, industry. Ok. So it seems that there is connection between these two graphs. One is about the world water use in 3 categories. The other one is about eh, water use in selecting countries in these 3 areas. Yes. Now we have got a rough picture of what picture like. Now I am trying to find the main features.

In the first one, obviously, agriculture took the main proportion, 70%, and then it is industry and then domestic use. In the second graph, I found that in different countries, the proportion of the three categories are different. Like in India, agriculture took about 92% while Canada is only 8%. Now I can compare these figures. But there is much more information in the table than in the pie chart. So I will concentrate more on the table. Ok, I think I am going to give, make a very very brief draft. I will firstly give a opening paragraph, and then I am going to talk about, yeah, the first pie chart and then second table. If I have time, I will draw a conclusion. If I don't have time, just forget it. Ok, now start.

Opening paragraph, eh... we have some sentence pattern like report. Ok, it is reported that,

ok, it is reported that in the pie chart that, oh, no no I made a mistake. Because I just directly to the second paragraph. I should have the opening paragraph. So, I should firstly say the chart and table, eh...I see, chart and table, below, yes, ok, below, describe, we use describe instead of giving information about, describe, how water was used in the whole world. I am changing the wording and paraphrasing in the world as well as in five different countries in the year 2000. Because it is in year 2000, actually it is past tense. This is very important...past tense. And then, describing the first graph. As something shows, as the pie chart shows, eh, agriculture, I should have a phrase here, take, account for is better. Agriculture accounted, past tense, accounted for 70% of the world water use in 2000. Now I am comparing so I can use the link word while, while domestic, while industry, industry took 22% and domestic use, here use is a noun, domestic use, 8%. Because I only give the figure, and then I should give some comments, so the amount of the water used by agriculture was, say, was more than twice. Because 70% and together, eh, agriculture took 70% but together domestic and industry only took 30%. So you can say the amount of water used by agriculture was more than twice the amount of industrial and domestic use. Ok, seems enough.

Now, I am moving on to the table. This is more difficult and more complex. Therefore, I should pay more attention to it. Let's say. According to the table, I think I can, Ok, I think I can firstly give a brief account of the main information given in the table like in the five countries. No, four countries, one two three four, make a mistake, four countries, not five countries, just now miscounted. One of the four countries, yes, China and India used more water in agriculture while New Zealand and Canada not. And ok, so in China and India, agriculture accounted for the most water user while in New Zealand and Canada, it is not the case. Ok, ok, we will start with this. According to the table, India and China, agriculture, oh, maybe I should just refine wording in the last sentence of the second paragraph. I should say, the amount of the water used by agriculture was more than twice the amount of industrial and domestic use together. I can add together to make it more accurate. Now I am continuing. According to the table of India and China, agriculture, eh, took the largest proportion of water use in the year 2000. eh, yes, ok, yes, right...ok...took the largest proportion of water use in the year 2000. India used 92% while China, I am comparing, while China 69%. Eh, however, in New Zealand and Canada, because Canada use the smallest amount of water in agriculture, I can put Canada first. In Canada and New Zealand, oh, I should change, it didn't mention, because it is different, in New Zealand, domestic use accounts for biggest proportion in Canada, it's industry. I should change. I should describe it one by one. However, in Canada, industry took the, I should, I should change the wording, paraphrase, industry took the, took the most amount of water use. The most amount, the largest amount. Industry took the largest amount of water use reaching 80%. While in New Zealand, I can use the sentence pattern to emphasize. It was, while in New Zealand, it was domestic use that accounted for the biggest water use, ok, so about largest water user in the four countries.

And now, I should discuss, I think, I can discuss the smallest water user. Yeah, ok, Eh, In India, industry only used 3%, thus, eh, thus becoming the smallest water user. And domestic only, only, a kind of, this is a kind of redundant. Use the same phrase again and again. In

India industry only used 3%, thus becoming the smallest water user and domestic just 5%. China, China dedicated 22% of this water to industry while only 90% in domestic. New Zealand, New Zealand, eh, actually, I think I should talk about, more about the water use in New Zealand. Because the amount of water use in agriculture and domestic were all the same. I think, I can add information, ah, I can add information here to New Zealand. That is water use in domestic is ok. While in New Zealand it was domestic use that accounted for the biggest water use, making 46%...making 46%. Eh, I can say, It is noted that the water use in agriculture was 44% and close to industry in New Zealand. It is good. New Zealand...only...I am continuing writing about the smallest water user. New Zealand only use 10% of water in its industry. Ok, last and least turn to Canada. Canada, Canada gave a very small proportion of its water to agriculture. As low as 8% and only 12% for domestic use. Ok, Ok. So I am almost done. I still have four minutes. I can just give a conclusion. Eh, in conclusion, the proportions of different water uses in the world vary from country to country. It appears that in developing countries, more water was used in agriculture while in developed countries, more in industry in the year 2000. Ok, I think, that is...I think I add one more sentence, as I still have one more minute. In conclusion, the proportions of different water uses in the world vary from country to country. Although, agriculture account, accounted for a significant majority in world water use. it appears that in developing countries, more water was used in agriculture while in developed countries, more in industry in 2000. Ok.

*Source: Li 2006, MSc dissertation supervised by Yu at University of Bristol, UK*

## Appendix 10: Stage Two GEPT Task A

### GEPT, Task A

Table 1 shows the IELTS (International English Language Testing System) test taker performance by geographic region in Asia in 2011; and Table 2, TOEFL-iBT (Test of English as a Foreign Language, Internet-based Test) test taker performance in 2012. After reading the data, you feel you have some ideas on how Taiwan test takers can do to improve their performance in both tests and decide to write a letter to an Opinion Section of a national English learning magazine in Taiwan.

In your letter:

- First, **briefly explain** why you are writing to the Opinion Section.
- Second, **summarize** what you think are the **main findings** from Tables 1 and 2, and **discuss some possible reasons for these findings**.
- Then, based on these findings, **make suggestions** about what Taiwanese test takers can do to improve their performance in both English tests.

Your letter **should be at least 250 words** in length. You have **50 minutes** to complete this task.

Table 1: IELTS test taker performance by geographic region (2011). Note: Maximum score for each skill and overall is 9.

Geographic Region	Listening	Reading	Writing	Speaking	Overall
China, People's Republic of	5.8	5.9	5.2	5.3	5.6
Hong Kong	6.7	6.4	5.9	6.2	6.4
Japan	6	5.6	5.5	5.8	5.8
Korea, South	6.2	6.1	5.4	5.7	5.9
Malaysia	7.2	7	6.2	6.6	6.9
Taiwan	5.9	6	5.5	5.9	5.9

Table 2: TOEFL test taker performance by geographic region (2012). Note: Maximum score for each skill is 30, and Total is 120.

Geographic Region	Listening	Reading	Writing	Speaking	Total
China, People's Republic of	18	20	20	19	77
Hong Kong	20	19	22	21	82
Japan	17	18	18	17	70
Korea, South	21	21	22	20	84
Singapore	25	24	25	24	98
Taiwan	19	20	20	20	79

## Appendix 11: Stage Two IELTS Task A

### IELTS, Task A

**You should spend about 25 minutes on this task.**

*Table 1 shows the IELTS (International English Language Testing System) test taker performance by geographic region in Asia in 2011; and Table 2, TOEFL-iBT (Test of English as a Foreign Language, Internet-based Test) test taker performance in 2012.*

*Summarise the information by selecting and reporting the main features, and make comparisons where relevant.*

**Write at least 150 words.**

Table 1: IELTS test taker performance by geographic region (2011). Note: Maximum score for each skill and overall is 9.

Geographic Region	Listening	Reading	Writing	Speaking	Overall
China, People's Republic of	5.8	5.9	5.2	5.3	5.6
Hong Kong	6.7	6.4	5.9	6.2	6.4
Japan	6	5.6	5.5	5.8	5.8
Korea, South	6.2	6.1	5.4	5.7	5.9
Malaysia	7.2	7	6.2	6.6	6.9
Taiwan	5.9	6	5.5	5.9	5.9

Table 2: TOEFL test taker performance by geographic region (2012). Note: Maximum score for each skill is 30, and Total is 120.

Geographic Region	Listening	Reading	Writing	Speaking	Total
China, People's Republic of	18	20	20	19	77
Hong Kong	20	19	22	21	82
Japan	17	18	18	17	70
Korea, South	21	21	22	20	84
Singapore	25	24	25	24	98
Taiwan	19	20	20	20	79

## Appendix 12: Stage Two GEPT Task B

### GEPT, Task B

A national English-language newspaper in Taiwan has recently published the results of a survey concerning credit card debt. Figure 1 reports the total amount of debt nationwide between 2003 and 2007, while Figure 2 reports the age distribution of people with credit card debt in 2007. After reading the data, you feel concerned about the way people use their credit cards and decide to write a letter to the Opinion Section of this English newspaper.

In your letter:

- First, **briefly explain** why you are writing to the Opinion Section.
- Second, **summarize** what you think are the **main findings** from Figures 1 and 2, and **discuss some possible reasons for these findings**.
- Then, based on these findings, **make suggestions** about what people can do to avoid credit card debt.

Your letter **should be at least 250 words** in length. You have **45 minutes** to complete this task.

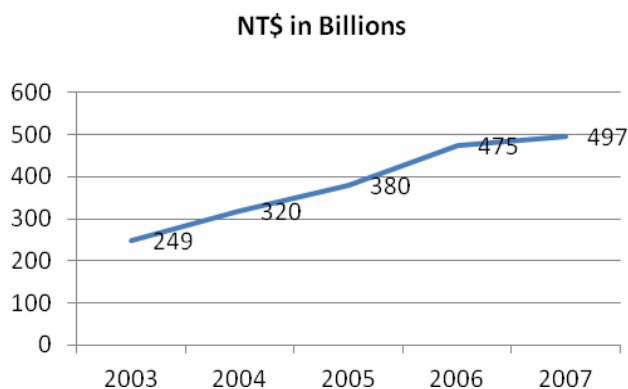


Figure 1: Total amount of credit card debt nationwide between 2003 and 2007

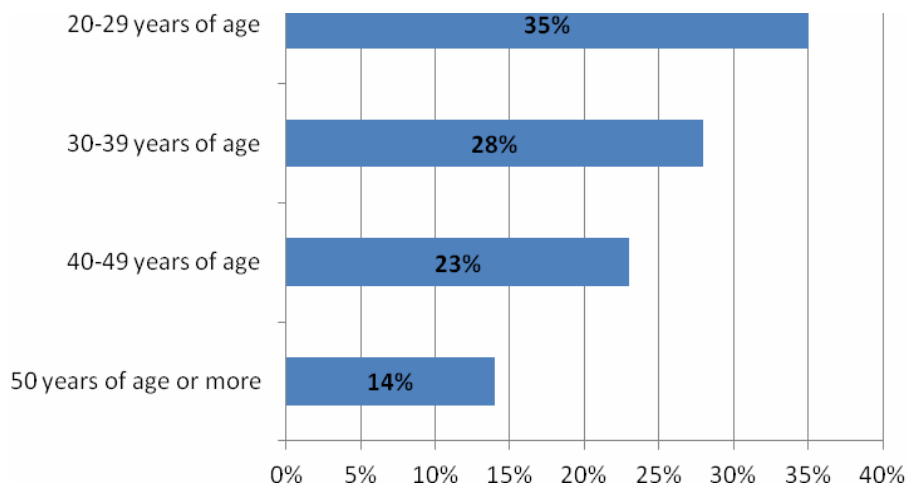


Figure 2: Age distribution of people with credit card debt in 2007



## Appendix 13: Stage Two IELTS Task B

### IELTS, Task B

You should spend about 25 minutes on this task.

Figure 1 reports the total amount of debt nationwide in Taiwan between 2003 and 2007, while Figure 2 reports the age distribution of people with credit card debt in 2007.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

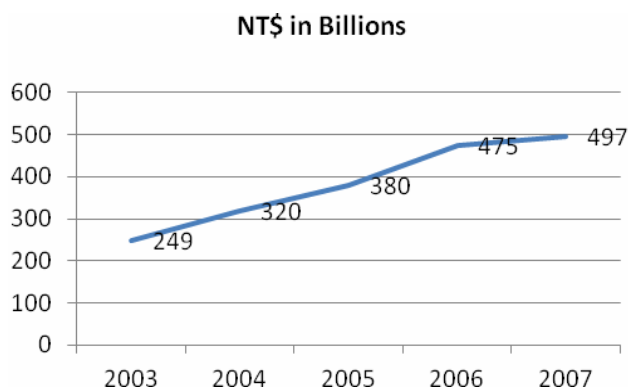


Figure 1: Total amount of credit card debt nationwide between 2003 and 2007

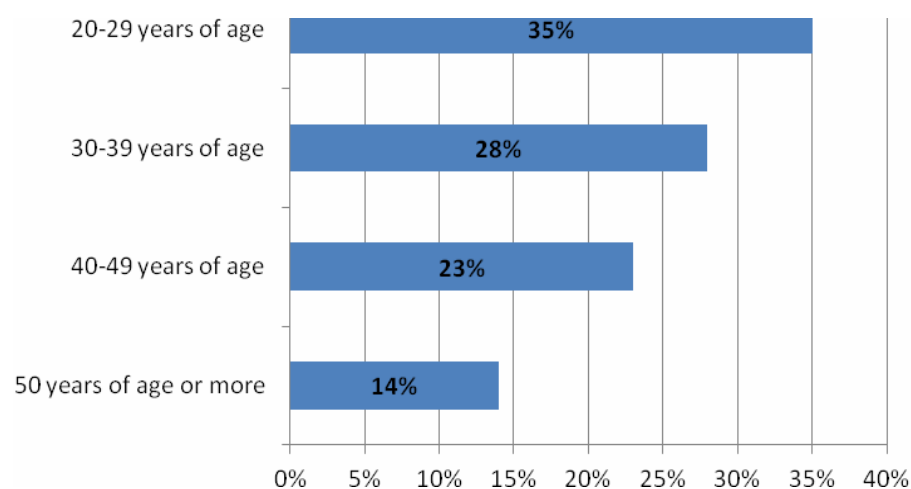


Figure 2: Age distribution of people with credit card debt in 2007

## Appendix 14: Stage Two GEPT Task C

### GEPT, Task C

Figure 1 shows the carbon dioxide (CO<sub>2</sub>) emissions (1990-2008); and Figure 2, the sources for producing electricity (2008) in mainland China. After reading the data, you feel you have some ideas on how the Chinese government can do to avoid its over-reliance on coal for electricity production – a leading contributor to CO<sub>2</sub> emissions, and decide to write a letter to an Opinion Section of a national English newspaper in mainland China.

In your letter:

- First, **briefly explain** why you are writing to the Opinion Section.
- Second, **summarize** what you think are the **main findings** from Figures 1 and 2, and **discuss some possible reasons for these findings**.
- Then, based on these findings, **make suggestions** about how the Chinese government can do to reduce its over-reliance on coal for electricity production.

Your letter **should be at least 250 words** in length. You have **50 minutes** to complete this task.

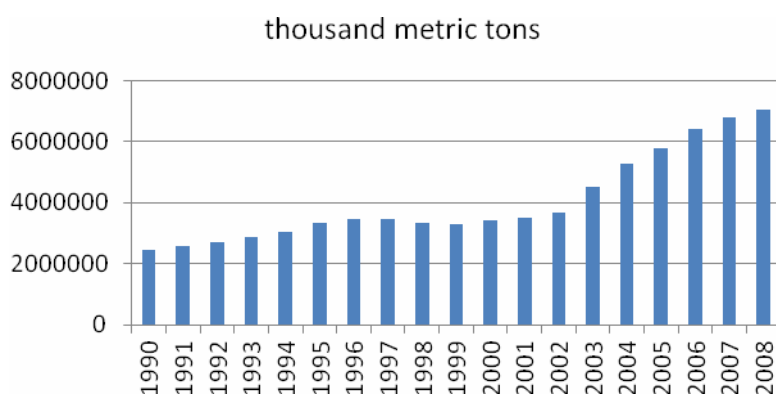


Figure 1: China's carbon dioxide emissions. (Source: United Nations Statistics Division)

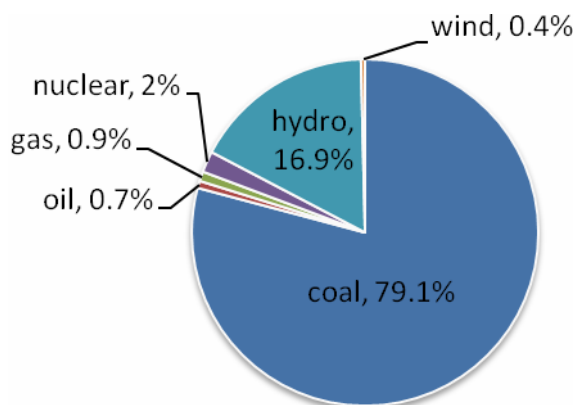


Figure 2: Sources for electricity production in China (2008)

## Appendix 15: Stage Two IELTS Task C

### IELTS, Task C

You should spend about 25 minutes on this task.

Figure 1 shows the carbon dioxide (CO<sub>2</sub>) emissions (1990-2008); and Figure 2, the sources for producing electricity (2008) in mainland China.

Summarise the information by selecting and reporting the main features, and make comparisons where relevant.

Write at least 150 words.

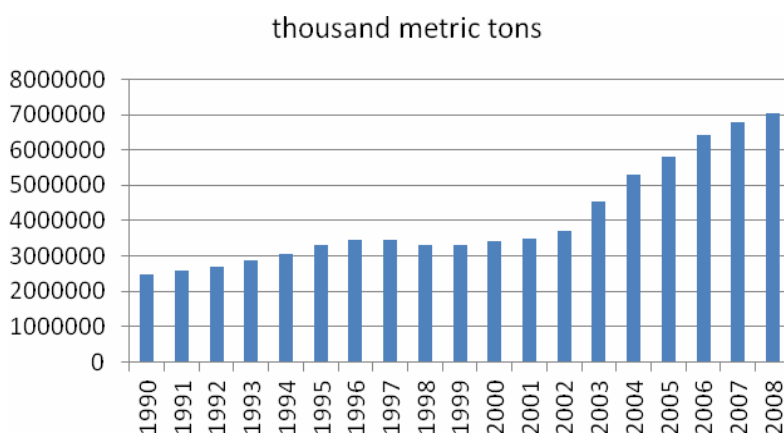


Figure 1: China's carbon dioxide emissions. (Source: United Nations Statistics Division)

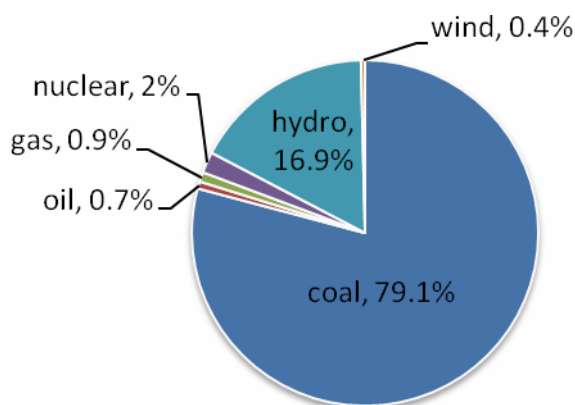


Figure 2: Sources for electricity production in China (2008)

## Appendix 16: Questions and guidelines for the semi-structured group interviews

1. **Briefing** the participants the purpose of the group interview: to better understand their thinking process when doing GEPT Advanced Writing Task Two (GEPT-2)
2. **General impression of the similarity and difference between GEPT-2 and IELTS-1:** Asking the students to talk about their experience of doing the GEPT-2 and IELTS-1 tasks:
  - What is their general impression of the tasks?
  - Which task(s) do they find more challenging and why?
  - Overall, as you understand, what are the major similarity and difference between GEPT-2 and IELTS-1 writing tasks?
  - What are the similarity and difference in your approaches/strategies when doing GEPT-2 and IELTS-1 writing tasks?
3. Questions on the **use of different graphs:**
  - In what ways, do you think your GEPT-2 and IELTS-1 writing processes may be affected by different types of graphs (e.g., line, bar, pie)?
  - Did you work differently for different types of graphs for GEPT-2 and IELTS-1 writing tasks, and how?
4. Questions on **graphicacy:**
  - In what ways, do you think your GEPT-2 and IELTS-1 writing processes may be affected by your familiarity and comprehension/understanding of graphs?
  - To what extent is there any difference or similarity between GEPT-2 and IELTS-1 tasks in how your graphicacy may affect your writing processes?
5. Questions on **English writing abilities:**
  - In what ways, do you think your GEPT-2 and IELTS-1 writing processes may be affected by your writing ability?
  - To what extent is there any difference or similarity between GEPT-2 and IELTS-1 tasks in how your writing processes may be affected by your writing abilities?
6. **Two summary questions:**
  - In summary, in what ways were your cognitive processes affected by the different graph prompts, your graphicacy and English writing abilities?
  - In summary, what are the major similarity and difference in your cognitive processes between GEPT-2 and IELTS-1 writing tasks?
7. Any other comments

### Notes:

- Ideally 3 students maximum in a group interview. However, depending on their availability, it is fine to have 1-1 individual interview or a slightly larger group interview.
- The interviews should be recorded.
- At the beginning of the interview, the researcher should say/record WHEN and WHERE the interview is conducted and WHO are in the interview.
- The interview will last 40-60 minutes.
- The interview can be conducted in Chinese and/or English.
- During the interview, the recorded think-aloud protocols and field notes taken during the test sessions may be used, as prompts or check-points, if necessary.

## Appendix 17: Sample rating of GEPT-AWT2

*We can easily see two apparent findings from figure one and two. First, we find three time periods when most accidents happen. Two of them are the rush hours during the day, when people are going to work or school, and when people are getting home. The other time is pretty late at night. In figure two, we find that nearly half of the drivers are under twenty-one. In the words, half of the traffic accidents are caused by these young drivers.*

*For each findings, we can find some possible causes. First, most accidents during the rush hours may be a result of the heavy traffic. Everyone is busy and in a hurry. And the drivers may become impatient in the traffic. But why do some many accident occur in the late night when the traffic is not as heavy? Perhaps, some drivers may become careless just because of the good traffic. Or, darkness may be another reason. Still, the most likely reason is drunk driving. Secondly, there are some reasons accounting for young drivers' accidents. Maybe, they are usually more impatient. Or, they have less experiences so that they are not used to the traffic so well. Or, some young drivers are too confident in their driving skills which are not as skillful as they think. These are some possible causes for this traffic accidents.*

*Maybe we can learn some lessons from the data and these passible causes. First, we should try to reduce the number of cars in the downtown aren during rush hours. To encourage more people to take mass transportation, the mass transportation system and the service should be improved. Severe punishments should be strictly enforced. Thirdly, the standards of giving drivers licenses should be more strict. We have to make sure every drivers on the roads are well-trained. If we can do all these mentioned, a large number of accidents can be increased.*

### READER'S COMMENT

---

	<u>R&amp;A</u>	<u>O&amp;C</u>	<u>LU</u>	<u>GU</u>
Pass	3	3	3	3

All parts of the task are sufficiently and clearly addressed. In general, the paper is logically organized, with appropriate paragraphing and use of linking devices. The range of vocabulary and structures is adequate, and their usage is mostly correct. However, there are still some errors (e.g. *But why do some many accident occur....*)

---

## Appendix 18: Participants' responses to the open-ended question of the graphicacy questionnaire

閱讀報章雜誌時會留意圖表, 在做報告時約有 40%的機會會製作圖表, 大致而言, 對圖表並沒有非常熟悉
高中的時候因為上地理課的關係比較有機會解讀圖表, 上大學後通常在報告時會有需要繪製圖表的機會, 但我發現在解讀方面, 若圖表形式是我不擅長的或不熟悉的, 在解讀時也會遇到比較大的挑戰, 表現可能會沒那麼好!
平常除了閱讀報章雜誌或是觀看新聞會接觸之外, 或在考試中遇到類似題型, 其他時間較不會主動接觸, 因此也不是很熟悉
會利用圖表呈現報告, 不同的項目會使用不同的圖
看得懂就好, 很少認真詮釋圖表內容, 自己畫出美麗的表格時會很爽!
可能與所讀科系有關, 只有在做教學評量或做比較時才會接觸和使用圖表
我最常使用, 製作的圖表為樹狀圖, 圓餅圖和流程圖. 對於這類的圖表我的判別能力較強, 也經常用這些來整理思緒. 其他圖表相對上使用及閱讀頻率少很多
我每天會看 MLB 的外電, 其中專欄作家每天會使用圖表來分析打者, 投手型態, 所以基本上每天都會閱讀一些圖表. 不過本身並不擅長製作, 只會使用 Excel 做一些簡單表格
用 PPT 報告時常會以數字或趨勢表格呈現我的重點. 我是模擬聯合國的社員, 社課常討論他國的青年失業人數, 由數字可知, 年年上升
IELTS 考試, 系上報告時會用到圖表加以解說, 讓人一目了然. 但對流程圖較不熟悉
較常看見圓餅圖及長條圖, 所以對這兩種圖熟悉度較高. 而因較不常接觸流程圖所以相較於上述兩種圖表會多花上一些些時間, 但也不至於無法理解. 其他如線圖, 表格等, 雖不比圓餅, 長條圖來得常接觸, 但對我而言都還算容易理解
歷史系資料中的圖表多以數字呈現, 因此化成線圖應該能展現趨勢, 所以在試題中, 若遇到線圖較好發揮
做研究時多以流程圖為實驗架構主軸, 長條圖為資料分析的呈現, 線狀圖則為趨勢表現

**Appendix 19: Regressions statistics of writing performance in Stage 1 and Stage 2 (Task A: IELTS)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.542 <sup>a</sup>	.293	.256	.5701

a. Predictors: (Constant), IELTS score (Stage 1 task 1)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.563	1	2.563	7.887	.011 <sup>a</sup>
	Residual	6.175	19	.325		
	Total	8.738	20			

a. Predictors: (Constant), IELTS score (Stage 1 task 1)

b. Dependent Variable: Stage 2 Task A IELTS score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.562	1.062		2.413	.026
	IELTS score (Stage 1 task 1)	.558	.199	.542	2.808	.011

a. Dependent Variable: Stage 2 Task A IELTS score

**Appendix 20: Regressions statistics of writing performance in Stage 1 and Stage 2 (Task B: IELTS)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.600 <sup>a</sup>	.360	.327	.7410

a. Predictors: (Constant), IELTS score (Stage 1 task 1)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5.876	1	5.876	10.700	.004 <sup>a</sup>
	Residual	10.434	19	.549		
	Total	16.310	20			

a. Predictors: (Constant), IELTS score (Stage 1 task 1)

b. Dependent Variable: Stage 2 Task B IELTS score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.804	1.221		1.477	.156
	IELTS score (Stage 1 task 1)	.742	.227	.600	3.271	.004

a. Dependent Variable: Stage 2 Task B IELTS score



**Appendix 21: Regressions statistics of writing performance in Stage 1 and Stage 2 (Task C: IELTS)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.754 <sup>a</sup>	.568	.547	.5631

a. Predictors: (Constant), IELTS score (Stage 1 task 1)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.350	1	8.350	26.331	.000 <sup>a</sup>
	Residual	6.343	20	.317		
	Total	14.693	21			

a. Predictors: (Constant), IELTS score (Stage 1 task 1)

b. Dependent Variable: Stage 2 Task C IELTS score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.630	.794		2.053	.053
	IELTS score (Stage 1 task 1)	.781	.152	.754	5.131	.000

a. Dependent Variable: Stage 2 Task C IELTS score

**Appendix 22: Regressions statistics of writing performance in Stage 1 and Stage 2 (Task A: GEPT)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.756 <sup>a</sup>	.572	.549	1.855

a. Predictors: (Constant), GEPT score (Stage1 task)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	87.294	1	87.294	25.371	.000 <sup>a</sup>
	Residual	65.373	19	3.441		
	Total	152.667	20			

a. Predictors: (Constant), GEPT score (Stage1 task)

b. Dependent Variable: Stage 2 Task A GEPT total score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.698	1.826		1.477	.156
	GEPT score (Stage1 task)	.645	.128	.756	5.037	.000

a. Dependent Variable: Stage 2 Task A GEPT total score

**Appendix 23: Regressions statistics of writing performance in Stage 1 and Stage 2 (Task B: GEPT)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.722 <sup>a</sup>	.522	.497	2.394

a. Predictors: (Constant), GEPT score (Stage1 task)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	118.899	1	118.899	20.743	.000 <sup>a</sup>
	Residual	108.911	19	5.732		
	Total	227.810	20			

a. Predictors: (Constant), GEPT score (Stage1 task)

b. Dependent Variable: Stage 2 Task B GEPT total score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.414	2.219		1.088	.290
	GEPT score (Stage1 task)	.709	.156	.722	4.554	.000

a. Dependent Variable: Stage 2 Task B GEPT total score

**Appendix 24: Regressions statistics of writing performance in Stage 1 and Stage 2 (Task C: GEPT)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.772 <sup>a</sup>	.596	.576	2.223

a. Predictors: (Constant), GEPT score (Stage1 task)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	146.090	1	146.090	29.554	.000 <sup>a</sup>
	Residual	98.864	20	4.943		
	Total	244.955	21			

a. Predictors: (Constant), GEPT score (Stage1 task)

b. Dependent Variable: Stage 2 Task C GEPT total score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized	t	Sig.
		B	Std. Error	Coefficients		
1	(Constant)	.692	2.125		.326	.748
	GEPT score (Stage1 task)	.876	.161	.772	5.436	.000

a. Dependent Variable: Stage 2 Task C GEPT total score

**Appendix 25: Regressions statistics of writing task performance of Stage 1 (IELTS-2) and Stage 2 (GEPT-AWT2, Task A)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.776 <sup>a</sup>	.603	.582	1.787

a. Predictors: (Constant), IELTS score (Stage 1 task 2)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	91.984	1	91.984	28.801	.000 <sup>a</sup>
	Residual	60.682	19	3.194		
	Total	152.667	20			

a. Predictors: (Constant), IELTS score (Stage 1 task 2)

b. Dependent Variable: Stage 2 Task A GEPT total score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-7.591	3.610		-2.103	.049
	IELTS score (Stage 1 task 2)	3.172	.591	.776	5.367	.000

a. Dependent Variable: Stage 2 Task A GEPT total score

**Appendix 26: Regressions statistics of writing task performance of Stage 1 (IELTS-2) and Stage 2 (GEPT-AWT2, Task B)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.802 <sup>a</sup>	.644	.625	2.067

a. Predictors: (Constant), IELTS score (Stage 1 task 2)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	146.670	1	146.670	34.345	.000 <sup>a</sup>
	Residual	81.139	19	4.270		
	Total	227.810	20			

a. Predictors: (Constant), IELTS score (Stage 1 task 2)

b. Dependent Variable: Stage 2 Task B GEPT total score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-8.339	3.540		-2.356	.029
	IELTS score (Stage 1 task 2)	3.350	.572	.802	5.860	.000

a. Dependent Variable: Stage 2 Task B GEPT total score

**Appendix 27: Regressions statistics of writing task performance of Stage 1 (IELTS-2) and Stage 2 (GEPT-AWT2, Task C)**

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.833 <sup>a</sup>	.694	.679	1.935

a. Predictors: (Constant), IELTS score (Stage 1 task 2)

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	170.097	1	170.097	45.446	.000 <sup>a</sup>
	Residual	74.857	20	3.743		
	Total	244.955	21			

a. Predictors: (Constant), IELTS score (Stage 1 task 2)

b. Dependent Variable: Stage 2 Task C GEPT total score

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-9.571	3.220		-2.973	.008
	IELTS score (Stage 1 task 2)	3.714	.551	.833	6.741	.000

a. Dependent Variable: Stage 2 Task C GEPT total score



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](mailto:geptgrants@lttc.ntu.edu.tw)  
Website: [www.lttc.ntu.edu.tw](http://www.lttc.ntu.edu.tw)



©LTTC 2014