GEPT®
LTTC 全民英檢

Language unfolds worlds.
Testing sets standards.

LTTC-GEPT Research Reports RG-04

A Register Analysis of Advanced GEPT Examinees' Written Production

David D. Qian

LTTC® 財團法人 語言訓練測驗中心
THE LANGUAGE TRAINING & TESTING CENTER

A Register Analysis of the GEPT Advanced Level Examinees' Written Production

LTTC-GEPT Research Reports
RG-04

David D. Qian

This study was funded and supported by the Language Training and Testing Center (LTTC), under the LTTC-GEPT Research Grants Program 2012-2013.

LTTC-GEPT Research Reports RG-04
A Register Analysis of the GEPT Advanced Level Examinees' Written Production

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

First published June 2014

Foreword

We have great pleasure in publishing this report: *LTTC-GEPT Research Reports RG-04*. The study described in this report was funded by the 2012-2013 LTTC-GEPT Research Grants. Conducted by Prof. David D. Qian of the Hong Kong Polytechnic University, the study examined the register features of the test-takers' written production across the two GEPT Advanced level writing tasks. It not only identified register features of written output from the tasks, but also justified the employment of two tasks in the test, providing further validity evidence for the GEPT Advanced Level Writing Test.

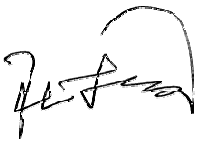
The GEPT, developed more than a decade ago by the LTTC to serve as a fair and reliable testing system for EFL learners, has gained wide recognition in Taiwan and abroad. It has generated positive washback effects on English education in Taiwan. As the GEPT has successfully reached out to the international 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 LTTC has set up the LTTC-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_LTTC/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
LTTC

Author Biodata

Prof. David D. Qian is Professor of Applied Linguistics in the Department of English and Director of the PolyU-Tsinghua U Centre for Language Sciences at The Hong Kong Polytechnic University. Professor Qian teaches master and doctoral level courses in language testing, classroom-based assessment, psycholinguistics, second language acquisition, and research methodology for applied language sciences. His research covers the areas of standardized English language testing, teacher-based assessment, corpus linguistics, discourse and communication in professional and academic contexts, and ESL/EFL vocabulary learning and measurement. As Principal Investigator, he has directed over 20 research projects funded by the Educational Testing Service, Research Grants Council of Hong Kong, Language Training and Testing Center, and The Hong Kong Polytechnic University. His research articles appear in a large number of international peer-reviewed journals, including *Language Assessment Quarterly: An International Journal; Language Learning: A Journal of Research in Language Studies; Language Testing; Assessment in Education: Principles, Policy & Practice; The Canadian Modern Language Review; TESL Canada Journal; and RELC Journal.* He consults in language testing, curriculum development and program evaluation, and serves on the Editorial Advisory Board of *Language Assessment Quarterly: An International Journal*, *Language Testing* and *Reading in a Foreign Language.*

摘要

◆ **研究團隊與研究目的**

本研究由香港理工大學 David Qian 教授主持，針對考生在全民英檢高級寫作測驗兩個部分的作文進行語體分析（register analysis），以了解考生兩部分作文所使用語體（register）的異同，並檢視考生作文中學術寫作論述（academic written discourse）的特徵，為高級寫作測驗提供更多效度證據。

◆ **研究問題**

1) GEPT 高級寫作測驗考生在其兩部分的作文所展現的寫作語體（written registers）相對關係為何？
2) GEPT 高級寫作測驗考生的兩部分作文存在那些語體特徵（register features）？GEPT 高級寫作測驗兩部分作文能否引導考生使用不同語體（registers）？
3) GEPT 高級寫作測驗考生的兩部分作文與學術英語的相似度多高？

◆ **研究方法**

研究者首先以 2010 至 2011 年高級寫作測驗 500 位考生的兩部分作文分別建置語料庫：語料庫一為 500 份考生第一部分的作文；資料庫二為 500 份考生第二部分的作文。接下來，研究者使用根據 Biber (1989)、Biber et al. (2007)，以及 Xiao (2009)的 MD-MF 架構（multi-dimension multi-feature framework）所修訂的 MD-MF 架構分析前述語料。分析方法包含因素分析(factor analysis)、相關性分析、變異數分析(ANOVA)、事後檢定測驗(*post hoc* tests)，以及質化分析。

◆ **研究結果摘要**

- ✓ 與全民英檢高級寫作第二部分相比，考生在第一部分使用的語體較為正式（more formal register）。
- ✓ 考生在全民英檢高級寫作兩部分作文使用的語體無論在字詞文法或語意方面都有明顯差異，顯示本測驗兩部分的題目均能發揮其特定的功能，引導考生寫出不同的語體。
- ✓ 考生在高級寫作兩部分的作文皆使用一定程度的學術字彙（academic words），顯示高級寫作兩部分測驗皆能引導考生使用學術字彙寫作。然而，在 academic phrases 方面，考生在兩部分作文所使用的頻率則不及學術字彙，可能係考生的程度不足，或考生對學術寫作的特徵與重要性了解不足所致。

Abstract

Language testers have come to understand the usefulness of employing multiple writing tasks in language proficiency tests in order to evaluate the multi-facets of test-takers' writing proficiency. One such important facet in the writing assessment is the candidate's ability to use appropriate registers in responding to task prompts during the assessment. It is expected that prompts for different writing tasks may elicit different registers. Without a register analysis on test takers' written production, however, it would be difficult to determine whether, and if so how, different writing tasks are capable of eliciting candidates' response data that can feature different registers. The present investigation into the Advanced level writing component of the *General English Proficiency Test* (GEPT) in Taiwan was conducted under an integrated multi-dimensional multi-feature (MD-MF) framework for register analysis, grounded in analytic approaches advanced by Biber (1998), Biber, Conner, and Upton, (2007) and Xiao (2009). Within this integrated framework, analyses were carried out capturing both lexico-grammatical and semantic features. Exploratory factor analyses were performed to extract latent variables for discerning main register features across two different writing tasks of the GEPT Advanced Level. In a fine-grained manner, the analyses aimed to profile the register features and locate the positions of the registers of 500 GEPT examinees' written output across the two writing tasks against an orality-literacy continuum drawn from previous research. Findings from the present investigation suggest that the two writing tasks of the GEPT Advanced Level are generally capable of eliciting two significantly different registers at both lexico-grammatical and semantic levels. At the lexico-grammatical level, both writing tasks are characterized by a number of dimensions that are mutually distinguishable. Although the location of the register elicited by Task 2 prompts, which contain nonverbal input, was found to be near the oral end of the continuum, this positioning can to some extent be justified by the nonverbal nature of the writing prompts for the task which may elicit certain elements featuring oral discourse. The study also profiled the examinees' written production against the Academic Word List (Coxhead, 2000) and Academic Formulaic List (Simpson-Vlach & Ellis, 2010) to determine the *academicality* in the examinees' response data, or the extent to which the writing prompts have elicited academic features as intended by the test design. The results show that, at the individual word level, the sampled examinees' written output generally exhibits a satisfactory degree of *academicality,* as there is a fairly satisfactory coverage of individual academic words, although the coverage of academic formulae seems to be relatively low for both tasks. Findings from the study have provided evidence to support the validity argument for the GEPT Advanced Level Writing Test.

Table of Contents

1. Introduction

In order to evaluate the multi-facets of test-takers' writing proficiency in a comprehensive manner (Grant & Ginther, 2000; Schoonen et al., 2002; Wolf-Quintero, Inagaki & Kim, 1998), developers of a plethora of international language proficiency tests (e.g. TOEFL® and IELTS®) have come to be aware of the importance of using more than one task in the writing assessment. In the meantime, corresponding rating scales, particularly analytic scales (Hamp-Lyons, 1991, 1995), are assigned to different tasks with a view to probing into various aspects of the candidates' written output, so as to obtain a fuller picture of the candidates' writing ability. Such practice is expected to facilitate the creation of positive washback on teaching and learning and better preparation for further academic studies (Cumming, Grant, Mulcahy-Ernt & Powers, 2004; Cumming, Kantor & Powers, 2001, 2002; Cumming, Kantor, Powers, Santos & Taylor, 2000; Hamp-Lyons & Kroll, 1996; Lee, Kantor & Mollaun, 2002; Rosenfeld, Leung & Oltman, 2001). However, whether multiple writing tasks are able to better assess test-takers' writing ability in different domains or registers still remains to be verified, without which the tasks would either fail to measure certain sub-constructs of the writing ability or render redundantly overlapping measurements due to using measures with similar latent traits repeatedly. Thus, there is a need to probe into the register positioning on a continuum (with the two ends being written and spoken, or academic and non-academic) so as to characterize the registers based on test takers' response data elicited from different writing tasks in a high-stakes English proficiency test, such as the General English Proficiency Test (GEPT) Advanced Level.

This study was designed to (1) investigate the variation of register positioning of candidates' written output across the two GEPT Advanced level writing tasks, (2) examine the register features that differentiate the registers elicited by different writing stimuli, and (3) identify salient features of academic written discourse embedded in candidates' output. It is expected that findings from all this, in an integrated manner, will better inform the scoring criteria for the GEPT Advanced level writing assessment (Roever & Pan, 2008; Shih, 2008) and supply test users with useful information for making decisions on recruitments, admissions and placements.

1.1. Targeting the GEPT Advanced Level Candidates

The test under discussion, GEPT, is a criterion-referenced English language proficiency test developed by the Language Training and Testing Center (LTTC) in Taiwan (Wu, 2007, 2010a, 2010b). This set of multi-level tests is gaining increasing recognition from various social dimensions (Kunnan & Wu, 2010; Roever & Pan, 2008; Shih, 2008). The entire GEPT test battery can be divided into five levels: elementary, intermediate, high-intermediate, advanced and superior (LTTC, 2002). The first three levels of the GEPT have attracted a considerable amount of research (e.g. Chen & Chang, 2008; Ma & Li, 2009; Wu & Chao, 2009; Wu & Chin, 2006; Wu & Liao, 2010; Wu & Ma, 2013) mainly due to the comparatively large candidate population[1]; nevertheless, the tests at the two advanced levels are generally under-researched. To fill in the gap, the present study was designed to investigate the written production of candidates from the GEPT Advanced Level so as to attain a better understanding of candidates' writing proficiency at this particular level.

[1] According to the score reports released from LTTC, the numbers of candidates taking elementary, intermediate, high-intermediate and advanced levels of GEPT in 2009 were 118000, 59000, 14000 and 304 respectively. (accessible online from *http://www.lttc.ntu.edu.tw/academics/results.htm*

1.2. Targeting the Written Production

In examining the response data from this group of test takers, the present study focused on the writing section of the second stage of the test (the first stage consisting of reading and listening sections). It is generally accepted that writing skills are important not only in academic studies but also in workplaces. With a growing need for internationalization in Taiwan, an increasing importance is also being attached to English writing proficiency. The writing tasks in the GEPT Advanced Level involve a number of important aspects that may affect the interpretation of the candidate's real performance on the test. In addition, the writing section of the GEPT Advanced Level differs from the corresponding sections of the three lower-level GEPTs in that the latter also include translation as a sub-section in the writing assessment. The GEPT Advanced level writing assessment, with two tasks, requires candidates to synthesize and summarize information from the given passages and to interpret visually presented information (Roever & Pan, 2008). The two different tasks are supposed to be able to elicit sufficient information on the overall writing ability of the test-takers from the GEPT Advanced Level, an assumption worth confirming.

1.3. Targeting Register Features

In order to provide useful information for further refining the scoring criteria of the GEPT Advanced level writing assessment, this project proposed to examine register features of test takers' written output with three specific objectives.

First, because of the different natures of the two writing tasks in the GEPT Advanced Level, it was assumed that two somewhat different sets of register features might be elicited from these tasks. Based on the existing taxonomies for register analysis (e.g. Biber, 1988; Biber, Connor & Upton, 2007; Xiao, 2009), an integrated framework was formulated for discerning and grouping linguistic features of the test takers' output so that a basis would be provided on which register comparisons and differentiations could be made. The integrated framework would therefore help determine whether the two writing tasks indeed elicited different registers since Task 1 is based on verbal input and Task 2, nonverbal input.

Second, as the GEPT Advanced Level aims to build a validity argument that the test assesses the candidates' suitability for pursuing further academic studies or employment where communication in English is essential, whether candidates' written production carries linguistic and discourse elements featuring English for Academic Purposes (EAP) and English for Specific Purposes (ESP) would be a topic worth investigating. Findings relating to this aspect would hopefully contribute to the promotion of positive washback effects on English language learners during their preparation for the GEPT, if it can be confirmed that an enhancement of academic discourse elements in the written production will meaningfully improve the quality of candidates' written output.

2. Research Background

2.1. Validation of the GEPT

Since the GEPT debuted in 2000, the test battery has undergone rounds of validation in various aspects. At the macro level, a number of studies have been conducted with foci on the alignment of the GEPT with major international English language proficiency tests and the

Common European Framework of Reference[2] (e.g. LTTC, 2003; Wu, 2010b), the concurrent validity of the GEPT with other well-established regional tests (e.g. Chin & Wu, 2001), the consequential validity of the GEPT (e.g. Shih, 2008; Wu, 2009; Wu & Chin, 2006) as well as some level-specific validity studies (e.g. Chin & Kuo, 2004; Wu & Lin, 2008). All this has contributed to the enhancement of validity and beneficial washback of the GEPT.

In addition to the above validation-oriented research, there are also studies focusing on specific language skills. Studies belonging to this group concentrate either on candidates' receptive skills (e.g. Chen & Chang, 2008; Ma & Li, 2009; Wu & Liao, 2010) or productive skills with writing ability as a main focus (Kuo, 2005; Wu & Chao, 2009; Wu, 2003). However, most of the above research relies largely on the test score, with insufficient attention accorded to analyzing the content of the test takers' actual performance, especially on the writing tasks. Therefore, it appears that a gap exists in the GEPT validation research in the area of register and discourse analysis for various writing tasks, as this will involve the analysis of the content of test-takers' writing.

2.2. The GEPT Advanced Level Writing Test

As mentioned earlier, there is not much research on the writing section of the GEPT Advanced Level. Therefore, it is necessary to describe this assessment component before the research questions are stated and design described. As illustrated in Table 1, the two tasks in the GEPT Advanced Level Writing Test require a bi-channel (output based on both verbal and non-verbal input) from test-takers. In particular, candidates are expected to express their own opinions, discuss possible causes and make recommendations in addition to making a summary based on the supplied texts. As both reading comprehension and writing ability are required in the test, the assumed construct of this section can be deemed as multi-componential. Table 2, with a detailed descriptor for each scoring level, lists all different domains to be observed for the two writing tasks. Each piece of writing is assessed from four aspects, namely, relevance and adequacy, coherence and organization, lexical use and grammatical use. As McNamara (2002) and Turner (2000) argue, the rating scale (and the way raters interpret the rating scale) represents the *de facto* test construct; therefore, in the present study, it is necessary to characterize the registers of the test-takers' written production, in order to further validate the existing rating scale.

Table 1. Breakdown of the GEPT Advanced Level Writing Section

Task	Input	Output	Length	Time
Task 1	Verbal input: 2 texts	Summary + expressing opinions	250 words	60 min.
Task 2	Non-verbal input: 2 charts, graphs, tables, or pictures	Summary + discussing possible causes + making recommendations	250 words	45 min.

[2] The detailed description of alignment is available at *http://www.lttc.ntu.edu.tw/E_LTTC/E_GEPT/alignment.htm*

Table 2. The Rating Scale for the GEPT Advanced Level Writing Tasks

	1	2	3 (Pass)	4	5
Relevance and adequacy	• The text lacks relevance and most parts of the task are not addressed. · Nearly all main ideas from the input are missing. · Personal opinions are irrelevant or do not address the task where required.		• The text is generally relevant and most parts of the task are addressed. · One or two main ideas from the input may be missing. · Personal opinions are generally relevant and adequately address the task where required.		• The text is entirely relevant and all parts of the task are fully addressed. · All main ideas from the input are covered. · Personal opinions are entirely relevant and comprehensively address the task where required.
Coherence -coherence -cohesion	• The text shows inadequate coherence and cohesion. · The organizational structure at the text level is not clear. Paragraphs are not separate, logical units. Ideas lack logical sequencing within and between paragraphs. Failures in continuity among ideas are noticeable. · Ideas are poorly connected due to limited/inappropriate linguistic devices.		• The text shows adequate coherence and cohesion. · The organizational structure at the text level is clear. Paragraphs are separate, logical units. In general, ideas are logically sequenced within and between paragraphs. There may be some redundancy, repetition, or lapses in continuity among ideas. · Ideas are adequately connected through the use of appropriate linguistic devices.		• The text shows excellent coherence and cohesion. · The organizational structure at the text level is exceptionally clear. Ideas are logically sequenced within and between paragraphs. There is strong continuity from one clearly stated idea to the next, and there is no repetition or redundancy. · Ideas are well connected through the use of a range of appropriate linguistic devices.
Lexical Use -range -appropriateness	• The range of vocabulary is inadequate to complete the task. • Lexical items are frequently used inappropriately. • The register is inappropriate or mixed, showing that the examinee is unable to distinguish between registers. • Overt plagiarism* and/or overuse of quotation is found in the text.		• An adequate range of vocabulary is used to complete the task. • Lexical items are used appropriately most of the time. • The register is appropriate with only occasional slips. • No plagiarism is found in the text, and quotation is used appropriately.		• A wide range of vocabulary is used to complete the task effectively. • Lexical items are used appropriately. Errors are rare. • The register is appropriate and consistent throughout. • No plagiarism is found in the text, and quotation is used appropriately.
Grammatical Use -range -accuracy	• The range of structures is too limited to complete the task. • Structures are frequently used inaccurately and/or inappropriately.		• An adequate range of structures is used to complete the task. • There may be some inaccurate structures.		• A wide range of structures is used to complete the task effectively. • Structures are used accurately and appropriately. Errors are rare.

* plagiarism: more than three consecutive words are copied from the input without the appropriate use of quotation marks.

2.3. Research Rationale

As described above, the present research project first focused on the register variation across different writing tasks. The prevailing practices concerning register analysis are usually two-fold. One stream focuses on a particular linguistic feature across various registers to describe the changes along the register continuum. For example, Aijmer (2002) renders a detailed description of how discourse particles vary in a variety of spoken registers. In a similar vein, such an approach can also be applied to particular words (e.g. Fortanet, 2004; Lindemann & Mauranen, 2001) or syntactic structures (Hyland, 2002a, 2002b; Marley, 2002). However, one of the limitations of this approach is self-evident in the sense that no panoramic picture of all the possible linguistic features can be captured for characterizing a particular register given that the feature is pre-determined, and the method itself intrinsically belongs to the top-down category. In that context, the other stream, represented by the multi-dimension multi-feature (MD-MF) approach, emerges as a more comprehensive, and therefore more promising, framework, which, instead of focusing on one particular linguistic feature, incorporates all potential linguistic features and groups together features that contribute to one particular factor elucidating one aspect of register characteristics. For instance, Biber (1988) proposed a framework that derives from the results of an exploratory factor analysis. He found that a number of linguistic features were heavily loaded on the factor measuring the degree of spoken and written languages. In his ensuing research, Biber *et al*. (2007) further extended the boundary of linguistic features to include a number of semantic categories to his previous framework (Biber, 1988), which largely dwells on lexico-grammatical features. Xiao (2009) further refined those semantic categories and introduced a few dimensions somewhat different from those of Biber's (1988, also see Biber *et al*., 2007).

For the data analysis of the proposed project, the existing taxonomies (Biber, 1988, Biber *et al*., 2007; Xiao, 2009) were examined, compared and integrated into one. In addition, as some lexico-grammatical or semantic categories might be in low profile, or barely present, in the GEPT data under investigation, the integrated framework was further modified based on the result of the initial analysis of a subsample (about 30%) of the GEPT data so that a suitable framework tailored for the *de facto* written production from the GEPT Advanced level candidates could be formulated as a result. This revised framework was then applied to analyzing the written output from the two GEPT Advanced level writing tasks so that their respective positions on the register continuum could be identified in the context of a broad scope of various written and spoken registers previously investigated (Biber, 1988; Biber *et al*., 2007).

In terms of academic discourse features in the candidates' response data, it has to be acknowledged that academic register studies may focus on not only written but also spoken registers. On the written side, specific grammatical or lexical features of written academic registers has been explored (Douglas, 1997), particularly in the arenas of science or medicine (e.g. Halliday, 1988; Swales, 1990; Thompson & Ye, 1991; Halliday & Martin, 1993; Hyland, 1994; Chih-Hua, 1999, Marco, 2000). In comparison, on the side of the spoken language, research has mainly been conducted on discourse markers and phraseology in academic settings (Biber & Barbieri, 2007; Cutting, 1999; Flowerdew & Tauroza, 1995; Powell & Simpson, 2001; Simpson-Vlach & Ellis, 2010). As one of the intentions for GEPT writing tasks is to assess the extent to which candidates are ready for further academic studies, it was, therefore, of significance to detect the degree of *academicality*, namely, aligning the academic register with candidates' written production. However, no requirement for such *academicality* is transparent in the band descriptors or range finders in the existing rating scale; nor does any highly-inclusive framework of academic discourse features exist to which this project could

readily refer. Therefore, the third step featured a bottom-up, corpus-driven approach to examining the potential features embedded in the GEPT Advanced level candidates' response data from the two writing tasks, such as stance and modality (e.g. Mauranen, 2003; 2004; Mauranen & Bondi 2003; Swales & Burke 2003) and university registers (Biber *et al*., 2002; Csomay, 2005; Reppen & Vásquez, 2007). It was anticipated that by referring to relevant details of the above studies, the present study would be able to reveal some useful academic written register features and map them against candidates' overall performance on the two writing tasks.

2.4. Theoretical Framework

In this section, the MD-MF framework developed in this study is further expounded and relevant studies revolving around it are also discussed in order to indicate the relevance and suitability of the model.

First of all, the evolution of MD-MF frameworks has gone through multiple phases, with Biber's (1988) framework as the forerunner. Because of its inclusiveness, Biber's 1988 model is still regarded as a powerful and influential framework (McEnery, *et al*., 2006), and all his ensuing frameworks are thought of only as extensions to, and refinement of, his (1988) framework. In fact, the MD-MF approach has been applied to a variety of discourse domains. In addition to the established demarcation between spoken and written registers for general use (Biber, 1988), there is also a number of studies focusing on comparing spoken and written university registers (Biber *et al*., 2002), conversational registers between American and British English (Helt, 2001), types of conversational text (Biber, 2008), academic writings between biology and history students (Conrad, 2001), spoken and written registers in elementary school settings (Reppen, 2001), letters written by different learner groups (Connor & Upton, 2003) and grant proposals (Connor & Upton, 2004). However, this framework seems to have been applied more to the analysis of written registers than spoken registers, and few studies have ever used data produced in a testing environment. Therefore, it seems that the adoption of a revised version of Biber's framework may widen the scope of usefulness for the framework.

In Biber's (1988) framework, an almost exhaustive list of linguistic features is covered, with 16 categories: (1) tense and aspect markers; (2) place and time adverbial; (3) pronouns and pro-verbs; (4) questions; (5) nominal forms; (6) passives; (7) stative forms; (8) subordination features; (9) adjectives and adverbs, prepositional phrases; (10) lexical specificity; (11) specialized lexical classes; (12) modals; (13) specialized verb classes; (14) reduced or dispreferred forms; (15) coordination; and (16) negation. In addition, these categories can be further expanded into 67 sub-categories (for a detailed inventory and explanations, see Biber, 1988, pp. 221-245). For example, under the category of lexical specificity, two linguistic features are included: type/token ratio and word length. After the standardized frequencies of all these linguistic features were computed across different registers, factor analyses were conducted to explore the latent factors, which can serve as indicators to predict the characteristics of a range of registers, such as interview, telephone conversation, prose and editorial. In Biber's (1988) work, seven dimensions, or latent factors, were extracted, abstracted and named in accordance with the particular linguistic features loaded. For instance, among the seven dimensions, Dimension 1, *informational versus involved production*, can best testify the positioning of a certain register regarding the degree of oral or written production (Biber, 1988).

Table 3. Biber's (1988) Multi-dimensional Multi-feature Framework

No.	Factors/Dimensions
1	informational *versus* involved production
2	narrative *versus* non-narrative concerns
3	explicit *versus* situation-dependent reference
4	overt expression of persuasion
5	abstract *versus* non-abstract information
6	online informational elaboration
7	academic hedging

In the present study, which adopted a similar MD-MF approach to evaluate all the main linguistic features that can be discerned, an exploratory factor analysis was conducted based on the standardized frequencies of these linguistics features retrieved from the candidates' written production in both writing tasks, from which a profile of register features reflected from linguistic features was captured, and further comparisons of various linguistic features across the writing tasks were made in response to the research questions below.

3. The Present Study

3.1. Research Questions

Based on the above considerations on the positioning of register variations using data elicited by the two writing tasks, the analysis of register features in the written output across a range of proficiency levels, and the determination of the degree of *academicality*, the following research questions were posed.

RQ1: Compared with previous register studies, what are the relative positions of written registers for different types of writing task as evidenced by the output of the GEPT Advanced level candidates?

RQ2: What register features may exist in candidates' written production across the two writing tasks? To what extent can the registers elicited by different tasks be distinguished from each other at lexico-grammatical and semantic levels?

RQ3: To what extent is the GEPT Advanced level examinees' written production from different tasks deemed academic?

3.2. Design and Methods

The flow chart below, indicating the three main research stages corresponding to the three research questions respectively, depicts the general procedures of how this project was conducted. The first stage (shaded in green) was the preliminary formulation of a revised framework tailored for English learner written register based on multidimensional models developed by Biber (1988), Biber *et al*. (2007) and Xiao (2009). The fundamental consideration in this stage was to integrate the linguistic features, especially encompassing the lexico-grammatical and semantic features into the framework. There was a possibility that some linguistic features would be removed from the framework if the results of the initial screening of the GEPT data indicated that these features did not figure prominently in the examinees' data.

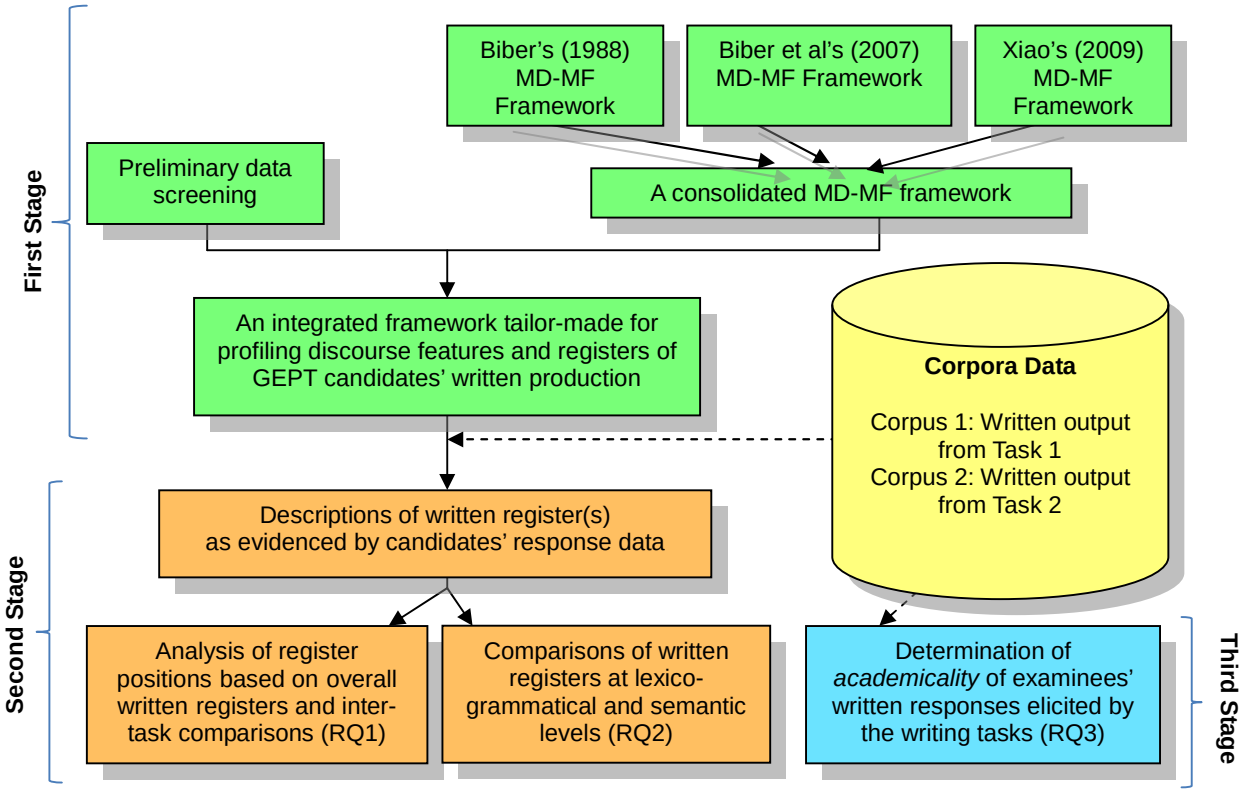


Figure 1. Research design

The second stage (shaded in orange) describes the register features of the GEPT Advanced level examinees' written output. During this stage, work were carried out to apply the revised framework to comparing candidates' response data, in the form of the two corpora constructed from the data elicited by the two GEPT writing tasks, so that relevant linguistic features could be discerned to position the candidates' written register(s) based on different task types. The analysis was focused on register comparisons between the two tasks so that RQ2 could be addressed.

The analysis for the third stage (shaded in blue), instead of relying on the above integrated analytic framework, consulted findings from previous relevant studies on academic register, as reviewed earlier (e.g. Biber *et al*., 2002; Csomay, 2005; Mauranen, 2003; 2004; Mauranen & Bondi 2003; Swales & Burke 2003; Reppen & Vásquez, 2007) and, based on the two corpora which were later constructed from GEPT examinees' written production, adopted a data-driven approach to profiling linguistic features which denote the intensity of *academicality*.

Regarding the research methods for the present study, a mixed-method approach integrating both quantitative and qualitative methods were adopted. The research questions were mainly addressed through statistical analyses, including factor analysis, correlations, non-parametric ANOVA and *post hoc* tests. The statistical processing was conducted by means of *SPSS* (21.0). In profiling the data conforming to the proposed framework, the quantitative retrieval methods of individual and multi-word extraction (*WordSmith Tools 6* as the concordancer), skip-grams and concgrams (*Concgram 1.0* as the automatic phraseology extractor) and semantic and part-of-speech tagging (*Wmatrix* as the automatic tagger) were employed. However, during the third stage, qualitative analysis was also necessary to manually examine a large amount of concordance lines which contain high-profile words and reflect important linguistic and discourse features, to pinpoint useful linguistic and discourse elements denoting

features of academic writing.

3.3. Research Data

For the purpose of generalizability and the thresholds of exploratory factor analysis, the sample frame of the present study was the GEPT Advanced level candidates with some demographic diversity in terms of age, gender and so forth. The test response data on the writing section of the GEPT Advanced Level were collected from 500 GEPT examinees. The dataset includes both physical and digital written scripts, sub-scores on the writing performances and, where possible, the other sub-scores of the selected individual candidates. The non-digital written scripts were first digitized by the project staff; then all the digitized data were compiled separately into Corpus 1, representing output from Task 1, and Corpus 2, Task 2, as foreshadowed in Figure 1. The sample size of 500 candidates (1,000 scripts) across two test administrations (2010 and 2011) ensured a fair representation of the response data on writing prompts (topics).

3.4. Research Findings and Discussion

3.4.1. Register analysis: Orality versus literacy

To begin with, Biber's (1988) first dimension of MD-MF framework, namely *informational vs. involved production*, was used to profile the written output across the two GEPT Advanced level writing tasks in. As the first dimension is a fundamental parameter to mark the relative orality or literacy of a register (Biber, 1988), findings in this aspect mainly respond to RQ1, where test-takers' output elicited by different writing stimuli can be positioned. There are two assumptions in relation to register analysis when written output is profiled for orality or literacy. First, it is expected that the two tasks are capable of eliciting different registers so that the practice of using multiple writing assessments in the GEPT Advanced Level can be justified and validated. Second, regarding the dimension scores deriving from both tasks, it would be ideally assumed that the observed registers fall into certain appropriate range, where similar registers are situated. It should be noted at this point that, in the ensuing analyses, all the frequencies of lexico-grammatical features, as well as those of semantic categories to be expounded below, have been standardized to *per* 1000.

Table 4 and Table 5 list the descriptive features of the 28 lexico-grammatical features relating to the first dimension, which were extracted from the writing scripts of Task 1 and Task 2 respectively. In each table, the first 23 features represent positive loadings on this dimension whilst the last 5 features (shaded) are negatively loaded.

Table 4. Descriptive statistics of the features in Dimension 1 (Task 1)

	Range	Minimum	Maximum	Mean	Std. Deviation
private verb	38.99	.00	38.99	10.5289	6.36684
THAT deletion	7.94	.00	7.94	.7736	1.41928
Contraction	40.27	.00	40.27	3.3069	4.68147
present tense verb	121.12	.00	121.12	70.2962	15.62213
2nd personal pronouns	35.60	.00	35.60	1.9679	5.15090
analytic negation	24.39	.00	24.39	6.6800	4.41497
demonstrative pronouns	29.17	.00	29.17	5.5353	4.92840
DO as pro-verb	12.90	.00	12.90	1.1341	1.83629
Emphatics	26.38	.00	26.38	6.2172	4.65058
1st personal pronouns	60.00	.00	60.00	13.3223	10.19073
it as pronoun	31.31	.00	31.31	8.8882	5.90330
BE as main verb	37.19	.00	37.19	12.6626	6.06421
causative subordination	14.05	.00	14.05	.9690	1.81291
discourse markers	6.38	.00	6.38	.1726	.67708
indefinite pronouns	14.60	.00	14.60	2.1501	2.68794
Hedges	10.83	.00	10.83	.5337	1.35559
Amplifiers	10.49	.00	10.49	1.5534	1.97361
sentence relatives	12.23	.00	12.23	1.3120	2.07183
WH questions	4.30	.00	4.30	.0557	.36750
possibility modals	34.60	.00	34.60	10.0763	6.65579
non-phrasal coordination	9.71	.00	9.71	.4592	1.24491
WH clauses	8.42	.00	8.42	.3760	1.17146
final prepositions	11.49	.00	11.49	1.1119	1.81613
other nouns	328.41	.00	328.41	2.3930E2	26.52489
word length	1.65	4.29	5.93	5.0796	.23706
prepositions	156.05	.00	156.05	1.0301E2	17.65661
type/token ratio	35.48	33.10	68.58	51.4491	5.21278
attributive adjectives	142.52	.00	142.52	71.6881	23.85452

Table 5. Descriptive statistics of the features in Dimension 1 (Task 2)

	Range	Minimum	Maximum	Mean	Std. Deviation
private verb	40.58	.00	40.58	14.7801	6.98235
THAT deletion	16.02	.00	16.02	1.4784	2.36420
contraction	30.16	.00	30.16	4.5308	5.04699
present tense verb	119.41	15.15	134.56	70.9413	16.92921
2nd personal pronouns	57.85	.00	57.85	6.6386	9.40994
analytic negation	28.80	.00	28.80	8.7446	5.34883
demonstrative pronouns	29.06	.00	29.06	6.8877	4.89076
DO as pro-verb	14.93	.00	14.93	2.7036	3.05132
emphatics	26.76	.00	26.76	6.8180	4.69210
1st personal pronouns	61.07	.00	61.07	16.6397	11.20274
it as pronoun	30.25	.00	30.25	8.7364	5.85198
BE as main verb	40.00	.00	40.00	16.1896	6.27055
causative subordination	15.87	.00	15.87	2.0218	2.68913
discourse markers	4.63	.00	4.63	.0885	.49742
indefinite pronouns	21.51	.00	21.51	3.5482	3.69270
hedges	10.44	.00	10.44	1.0819	1.83890
amplifiers	10.44	.00	10.44	1.3293	2.01748
sentence relatives	10.99	.00	10.99	1.3148	1.94620
WH questions	4.65	.00	4.65	.0525	.37432
possibility modals	38.46	.00	38.46	9.0940	5.79770
non-phrasal coordination	11.43	.00	11.43	.8022	1.65546
WH clauses	5.80	.00	5.80	.2331	.77145
final prepositions	10.67	.00	10.67	.6670	1.37602
other nouns	212.14	143.85	355.99	2.2669E2	26.57236
word length	1.43	4.05	5.49	4.6610	.26354
prepositions	113.57	48.93	162.50	1.0751E2	20.00664
type/token ratio	31.08	35.39	66.47	51.2284	5.32940
attributive adjectives	71.94	6.37	78.31	31.4455	11.72939

As suggested by Biber (1988) and McEnery, Xiao and Tono (2006), when dimension scores are computed, researchers may generally follow the formula below, where *N* stands for the number of texts being observed, *SD* for standard deviation and ∑ for summing. Therefore, it can be generally understood that, in the formula, a dimension score of the observed register can be obtained by adding together the mean factor scores of all features with positive weights on a factor and then subtracting the mean factor scores of all features with negative weights on the same factor. In the case of Dimension 1, the first 23 features carry positive weights whereas the last 5 features are negatively loaded on the dimension. It should be noted that the positive or negative sign preceding a value should be retained so that a negative factor score for a feature with negative weight, such as -(-1), would become positive when the dimension score is computed.

$$\textbf{\textit{Dimension score}} = \sum \{\sum [(\text{frequency} - \text{frequency mean})/SD]/\text{N}\}$$

With the statistics in Table 5 and Table 6, the scores for Dimension 1 in accordance with Biber's (1988) framework are thus computed. It is found that the dimension score for Task 1 equals -0.04 whereas that for Task 2 equals 7.20. With reference to the two assumptions outlined above, if the dimension scores representing the degree of orality/literacy diverge, it may be interpreted that multiple writing tasks are able to elicit different registers as anticipated. However, it should be borne in mind that although the dimension scores are different, the elicited register should also approximate what is expected. For example, if a writing assessment expects to elicit two registers, which are news report and seminar notes respectively, their corresponding dimension scores should be positioned somewhere close to the registers profiled in the previous studies.

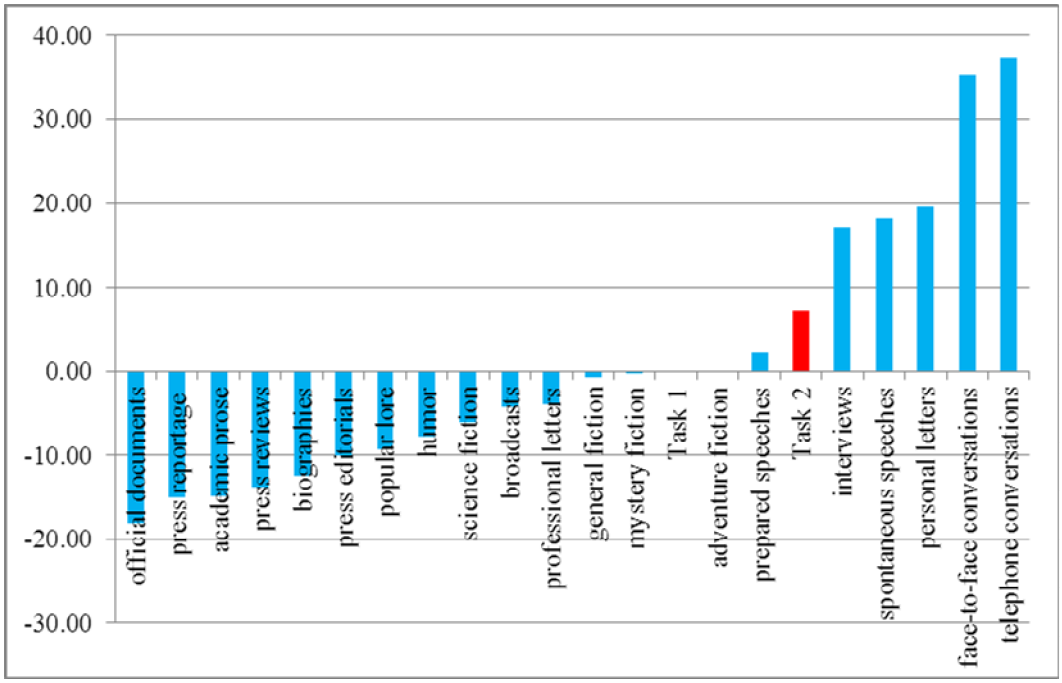


Figure 2. Comparison of orality/literacy of registers

As such, the above computed dimension scores are intended to be compared with other registers (Biber, 1988) along a continuum of orality/literacy. As illustrated in Figure 2, the higher a dimension score for a particular register, the higher the degree of orality it tends to present. From the Figure, it is generally believed that a dimension score of zero can serve as a rough demarcation that delineates written production from spoken discourse (Biber 1988). Task 1 seems to lie amongst a cluster of fictions of various subgenres, such as adventure fiction and mystery fiction. In comparison, Task 2, positioned between prepared speeches and interviews, tends to be closer to the end of orality. Although there is no explicit statement of what register a particular writing task intends to elicit, a vague picture might be conjured up via the instructions preceding the writing stimuli. In both test administrations, Task 1 is generally described as "an essay for a national essay contest" (see Test Papers 2010 & 2011) and Task 2 as "a letter to the Opinion Section of a local English newspaper" (see Test Papers 2010 & 2011). Therefore, the register positioning of Task 1 can be fairly acceptable in that it at least approximates the registers of written production, particularly amidst fictions. However, when it comes to Task 2, the register position seems far detached from professional letters, the register of which should be deemed similar to what is expected from the writing stimulus. It is even far away from press editorials and press reportage that are commonly seen in the newspapers. In other words, while the output elicited by the Task-2 prompts is in written

format, its register position actually anchors between written and spoken registers. This issue deserves attention because there might be two possibilities that have caused this: the test designer purposefully designed this task for an informal written register, as letters appearing in an opinion section of a newspaper often bear some features of oral English; however, it could also be due to a weakness in the design of Task 2 that has led to a written register approximating orality without the test developer's awareness. Whichever was the reason that has caused this orality in the register, there is a need to conduct some further investigation, In order to depict a fine-grained picture of what makes the discrepancies, particularly what leads the Task 2 register to approximate orality, between the two observed registers, this study further compares the means of the individual lexico-grammatical features.

Table 6. Statistically different features between two registers

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
THAT	62.478	.000	-5.716	998	.000	-0.705	0.123	-0.947	-0.463
deletion			-5.716	817.323	.000	-0.705	0.123	-0.947	-0.463
contraction	10.124	.002	-3.976	998	.000	-1.224	0.308	-1.828	-0.620
			-3.976	992.412	.000	-1.224	0.308	-1.828	-0.620
2nd personal	116.953	.000	-9.736	998	.000	-4.671	0.480	-5.612	-3.729
pronoun			-9.736	773.400	.000	-4.671	0.480	-5.613	-3.729
analytic	19.051	.000	-6.656	998	.000	-2.065	0.310	-2.673	-1.456
negation			-6.656	963.384	.000	-2.065	0.310	-2.673	-1.456
1st personal	7.884	.005	-4.898	998	.000	-3.317	0.677	-4.646	-1.988
pronoun			-4.898	989.185	.000	-3.317	0.677	-4.646	-1.988
causative	56.885	.000	-7.259	998	.000	-1.053	0.145	-1.337	-0.768
subord.			-7.259	874.931	.000	-1.053	0.145	-1.337	-0.768
indefinite	49.602	.000	-6.844	998	.000	-1.398	0.204	-1.799	-0.997
pronouns			-6.844	911.878	.000	-1.398	0.204	-1.799	-0.997
hedges	70.368	.000	-5.366	998	.000	-0.548	0.102	-0.749	-0.348
			-5.366	917.693	.000	-0.548	0.102	-0.749	-0.348
Non-phrasal	41.875	.000	-3.703	998	.000	-0.343	0.093	-0.525	-0.161
coord.			-3.703	926.623	.000	-0.343	0.093	-0.525	-0.161
word length	9.396	.002	26.407	998	.000	0.419	0.016	0.388	0.450
			26.407	987.015	.000	0.419	0.016	0.388	0.450
attributive	218.104	.000	33.852	998	.000	40.243	1.189	37.910	42.575
adjectives			33.852	726.965	.000	40.243	1.189	37.909	42.576

Table 6 lists the lexico-grammatical features that are found significantly different (p <.01) across the registers elicited by the two writing tasks. Out of 28 lexico-grammatical features that belong to Dimension 1 in Biber's (1988) framework, 12 features (42.9%) present statistical difference across the two observed registers. Certain features, though statistically different, might not sufficiently contribute to the explanation of why the register compiled from Task 2 written production is close to the spoken genres. As recorded by Biber (1988), *indefinite pronouns*, as markers of generalized pronominal reference, and *non-phrasal*

coordination (also labeled as independent clause coordination) have not been used frequently for register comparison. Despite this, a good number of the above features can account for why the observed register presents a high yet unexpected degree of orality.

First, a few lexico-grammatical features that have affinities with reduction, and that are usually salient in spoken discourse might significantly discern the two observed registers as profiled above. These features may include contractions, THAT-deletion and analytic negation, as found in Table 6. Contractions can be the most frequently cited example of reduced surface form. Biber (1988), along with a number of similar studies, believe that they are quite dispreferred in formal and edited writing, and that contractions are substantially used in informal registers. Unlike contractions, which are merely a form of phonological or orthographic reduction, THAT-deletion is a form of syntactic reduction. As earlier claimed by Frawley (1982) and Weber (1985), the concern for elaborated and explicit expression in written production serves as the driving force preventing this reduction. In other words, when a text is characterized by a high level of formality, the connector *that* is almost never omitted. The third feature under discussion is analytic negation. As the negation of this kind is usually realized in contracted form, such as *isn't* and *won't*, its contribution to approximating spoken registers is not surprising.

Second, first personal pronouns and second personal pronouns also distinguish the two elicited registers noticeably. First personal pronouns, generally treated as markers of ego-involvement in a text, indicate an interpersonal focus and a generally involved style (Chafe, 1982, 1985). In comparison, second personal pronouns require a specific addressee and indicate a high degree of involvement with that addressee (Chafe, 1985). As such, both parameters clearly present an extent of subjectivity and engagement. When comparisons of spoken and written registers are made (e.g. Biber, 1986; Chafe & Danielewicz, 1986; Hu, 1984), the overuse of first and second personal pronouns, also known as writer/reader visibility (McCrostie, 2006; Petch-Tyson, 1988), might trigger the register to move towards the end of orality. However, in terms of letter writing, which is expected to be associated with writing stimuli, this aspect of overuse may be somehow justified in the sense that test-takers need to employ an abundance of self-mentioning and/or address the imaginary newspaper editor(s).

Third, although the de facto use of causative subordinators needs yet to be more closely examined, results from the current analysis suggest that the number of causative subodinators produced by the test takers in their writing for Task 2 is larger than expected. In Biber's (1988) framework, causative subodinator refers to *because* only, which is regarded as the sole subodinator to function unambiguously as a causative adverbial. It might sound confusing why an overuse of *because* would indicate orality. As investigated and recorded previously, words such as *as*, *for*, and *since*, have a range of functions to serve, including being causative subodinators. However, Tottie (1986) and Alterberg (1984), who conducted detailed analyses of these subordination constructions, found that *because* had a more salient presence in spoken discourse while *as* was used more frequently as a causative subodinator in written discourse. Therefore, the register elicited from Task 2 written output serves to exhibit an overuse of causative subodinator *because* for a written register. On one hand, this phenomenon could be the result of the wording of a requirement of the task which says: "the possible reasons for these findings (from graphs and charts)" should be discussed. This wording would naturally elicit a good variety of causative subodinators. On the other hand, it is understood that the use of *because* is solely dependent on test-takers' discretion. Without strong register awareness, they are completely left free to choose any of the available

causative subodinators. However, as what was being investigated was the advanced level of the GEPT, the test candidates should be expected to be able to discern accordingly the shades of differences as outlined above.

Fourth, there also seems to be a significant difference in the use of hedges across the observed registers. As documented by Biber (1986), the use of hedges in conversational discourse indicates an awareness of the limited word choice that is possible under the production restrictions of speech, such as *something like*. Therefore, when hedges are used in great numbers, their contributions for a register to approximate orality also aggregate. A closer look at the concordance lines of hedges in Task 2 written scripts reveals that three words dominate this lexico-grammatical feature: *about (raw frequency = 98)*, *almost (raw frequency = 75)* and *maybe (raw frequency = 36)*. The first two words usually precede the numerals, indicative of aboutness inherent in the charts and graphs of Task 2, whilst *maybe* is more concerned with test-takers' explanations for the possible reasons elicited by the writing prompts. As the task *per se* mainly anticipates a written register, an overuse of these hedges could be somewhat problematic.

Fifth, after a presentation and discussion on the findings regarding the features with positive loadings, the remaining features that cause significant differences within the observed dimension are negatively loaded: *word length* and *attributive adjectives*. It is speculated that the possible reason why the mean word length of Task 2 written production is significantly shorter than that of Task 1 is that a great number of numerals are involved in the writing scripts of Task 2, thus considerably reducing the word length. As a result of this speculation, all the numerals, including percentages and fractions, were removed for another round of independent t-test, which actually elevated the mean word length of Task 2 productions to 4.71, as compared to 4.66 in Table 5. However, a significant mean difference still existed. When it comes to *attributive adjectives*, the top 10 clusters that conform to the constructions of *attributive adjectives*, as listed in Table 7, were extracted so as to probe into the possible reasons. It might dawn upon this study that in Task 1, the top 10 2-word attributive adjectives are dominantly related to the topics, namely tourism and culture, except for *FOREIGN WAY*, which is ranked the 9th. As most of these *attributive adjectives* are present in the writing stimuli, it is very likely that test-takers were enticed to refer to them verbatim repeatedly, due to the factor of keyness of such *attributive adjectives*. As such, it can be felt that the significant difference in attributive adjective frequencies is mainly caused by the writing stimuli of the tasks.

Table 7. Comparisons of attributive adjectives

Rank	**Task 1**		**Task 2**	
	L1	***Center***	***L1***	***Center***
1	CULTURAL_JJ	TOURISM_NN1	OTHER_JJ	PEOPLE_NN
2	LOCAL_JJ	PEOPLE_NN	MAIN_JJ	REASONS_NN2
3	TRADITIONAL_JJ	COMMUNITIES_NN2	POSSIBLE_JJ	PROBLEM_NN1
4	NATIVE_JJ	CULTURES_NN2	CRIMINAL_JJ	REASON_NN1
5	OTHER_JJ	CULTURE_NN1	SOCIAL_JJ	STUDENTS_NN2
6	NEW_JJ	GROWTH_NN1	SERIOUS_JJ	WAY_NN1
7	ECONOMIC_JJ	ENVIRONMENT_NN1	YOUNG_JJ	INVESTIGATION_NN1
8	DIFFERENT_JJ	CULTURAL_JJ	PERSONAL_JJ	CASES_NN2
9	FOREIGN_JJ	WAY_NN1	NEW_JJ	INFORMATION_NN1
10	BIG_JJ	LEGACY_NN1	IMPORTANT_JJ	ISSUE_NN1

As an interim summary, with reference to the first dimension of Biber's (1988) MD-MF framework, this section presents the research findings with regard to the degree of orality/literacy of the observed registers elicited from two writing tasks respectively. It was found that the dimension scores for both registers diverge in a certain degree, with Task 1 register nearing the expected elicitation and Task 2 register approximating the registers of spoken discourse. Therefore, it is felt that Task 2 might have not fully elicited what is expected from test-takers. In order to further explore the potential reasons for this phenomenon, this study further compared the individual lexico-grammatical features within the dimension, and analyzed those features that cause significant differences across the registers.

Notwithstanding the discrepancies found in the dimension scores of the observed registers, we further probed why the register of Task 2 output approximates the end of orality along the continuum. With the above fine-grained analyses on individual lexico-grammatical features, the reasons for the register differences may be broadly categorized into two folds. First, even though the writing stimuli aim to elicit two written registers with assumed divergence, the outcome may present a nuanced picture because of test-takers' *de facto* written output. In case they as a whole do not have strong register awareness that would enable them to produce a written register, the written output could bear many salient features of spoken discourse, such as an overuse of reduced forms and general hedges as well as heavy reliance upon the causative subodinator *because*. Considering the Task 2 register is elicited from advanced-level candidates' written production, a number of features in spoken output are not supposed to be of high profile, given that the writing scripts were selected from a pool of test-takers with representative score ranges. It might be suggested that if the genre of letter writing is still kept intact in the GEPT, perhaps more instructions concerning a professional letter to the newspaper editor(s) could be entertained.

Second, apart from test-takers' proficiency that somehow impeded the successful elicitation of the expected register in Task 2, the writing stimuli in both tasks also need to be slightly reexamined and reconsidered. Although the intention of separating the writing tasks into verbal and nonverbal input can be somehow justified, would it be possible that test-takers are provided with far too many ready-made materials in Task 1 whereas a fairly scant verbal stimulus is present in Task 2? Test-takers, consciously or unconsciously, would digest certain phraseologies in the readings, and further use them verbatim in synthesizing both texts. As found above, the abundance of attributive adjectives in Task 1 register that is caused by topic factor can be a good case in point. In comparison, as no verbal materials are available in Task 2, the register compiled from test-takers' written production seems to present a lesser degree of sophistication, as partially evidenced by a shorter word length. In addition, although test-takers might refrain from plagiarism as the practice of excessive copying is deemed inappropriate, according to the task instructions, yet it is not clear how much copying was actually materialized and how rigorously this was monitored in the marking since the data provided for the project have no indication on the details of marking.

With a view to pulling the Task 2 register somewhat towards the end of literacy, it is suggested that whilst the two writing tasks can still have their different types of verbal and nonverbal inputs, a fine-tuning could be attempted, perhaps by adding more yet controlled details to the writing prompts of Task 2, so that the register of the elicited output from Task 2 prompts will not be too oral or informal.

3.4.2. Register analysis: Latent factors

The research findings in this section are unfolded in three aspects. As the combination of lexico-grammatical and semantic features for running an exploratory factor analysis may cause difficulty of interpretation, this study looked into the latent factors that may characterize the registers from the above two spectrums separately. It should be noted that in this study register features were extracted via principal component analysis, and that the factors were further rotated in a varimax manner so as to maximize the sum of the variances of the squared loadings.

3.4.2.1. Exploring lexico-grammatical factors

This study first explored the latent lexico-grammatical factors across the observed registers. The assumption was that, if the two writing tasks elicited two different registers, their corresponding register features in the domain of lexico-grammar should also be discernible from each other.

Prior to an exploratory factor analysis, whether the datasets fitted such statistical operation needed to be first checked. Following the research of a similar line, Kaiser-Meyer-Olkin (KMO) Test and Bartlett's Test of Sphericity are usually conducted. Kaiser (1974), and Hutcheson and Sofroniou (1999) believe that a KMO value between 0.70 and 0.80 can be regarded as good and above 0.80 ideal. As for Barlett's Test of Sphericity, as long as the test can prove statistical significance, EFA would be appropriate. Both indices should be considered in a triangulated manner, and will also be referred to in the follow-up analyses, where the latent semantic factors are considered.

Table 8. KMO and Bartlett's Test for lexico-grammatical features (Task 1)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.704
Bartlett's Test of Sphericity	Approx. Chi-Square	1.666E3
	df	1378
	Sig.	.000

Table 9. KMO and Bartlett's Test for lexico-grammatical features (Task 2)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.664
Bartlett's Test of Sphericity	Approx. Chi-Square	2.073E3
	df	1378
	Sig.	.000

Table 8 and Table 9 list the output of KMO and Bartlett's Test for Task 1 and Task 2 registers respectively. With reference to the threshold values specified above, both datasets present goodness-of-fit indices for running EFA. Although the KMO value for Task 2 register equals 0.664, which is slightly lower than 0.70, its significant Bartlett's test result (p=0.0001) still qualifies it for EFA. Therefore, we proceed to observe the scree plots generated by the two datasets.

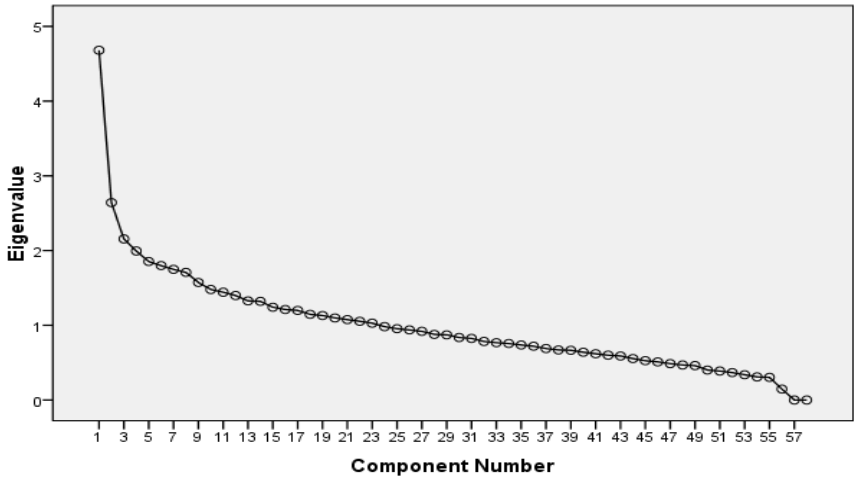


Figure 3. Scree plot of lexico-grammatical features (Task 1)

As can be seen in Figure 3, there are natural break points in the curve of eigenvalues. The number of data points above the break is supposed to be the number of factors to retain. A number of natural breaks can be noted in Figure 3, the most obvious of which are between the first 10 data points, because the 10th point marks an abrupt drop and a subsequent ease-up tendency on the curve. While over-extraction has some drawbacks, under-extraction is also undesirable. On one hand, there is a loss of information in under-extraction in that too many linguistic features will be excluded from final analysis. On the other hand, under-extraction can produce a "confused picture" (Biber, 1988, p.88) of linguistic features when factors collapse, thus complicating the interpretation. As a result, this study followed the recommendations of Costello and Osborne (2005) in running multiple factor analyses by manually setting the number of retained factors as 6 to 10 as indicated in the scree plot. After a comparison of the factorial structures based on 6 to 10 factors in terms of the number of significant loadings (above 0.30) on each factor, cross loadings as well as the ease of interpretation of the extracted factors, this study established an eight-factor factorial structure on the basis of 500 writing scripts for Task 1 register. Table 10 lists the seven extracted factors along with their respective squared loadings. As can be seen, when the number of factors is specified as 7, the cumulative percentage of variance explained reaches 29.092%.

Table 10. Factor extraction of lexico-grammatical features (Task 1)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.681	8.070	8.070	4.681	8.070	8.070
2	2.642	4.556	12.626	2.642	4.556	12.626
3	2.156	3.717	16.343	2.156	3.717	16.343
4	1.994	3.438	19.781	1.994	3.438	19.781
5	1.853	3.195	22.976	1.853	3.195	22.976
6	1.799	3.101	26.077	1.799	3.101	26.077
7	1.748	3.015	29.092	1.748	3.015	29.092

Table 11. Factor extraction of lexico-grammatical features (Task 2)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.457	7.684	7.684	4.457	7.684	7.684
2	3.142	5.418	13.102	3.142	5.418	13.102
3	2.401	4.140	17.241	2.401	4.140	17.241
4	2.138	3.686	20.927	2.138	3.686	20.927
5	2.012	3.468	24.395	2.012	3.468	24.395
6	1.944	3.352	27.747	1.944	3.352	27.747

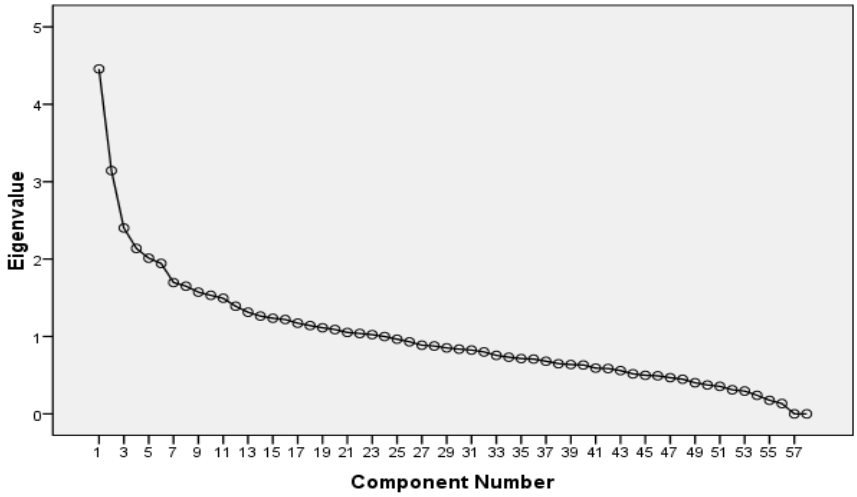


Figure 4. Scree plot of lexico-grammatical features (Task 2)

As reflected in Figure 4, when a similar approach was applied to the Task 2 written output, only six factors were extracted, taking into consideration the significant loadings (above 0.30) on each factor, the effect of cross loadings and the ease of interpretation. The scree plot also indicates a rather sudden drop on the curve at the point of the sixth factor. Accordingly, Table 11 lists the extracted factors of Task 2 register and their squared loadings, with the total percentage of variance explained amounting to 27.747%.

Having extracted the latent factors of lexico-grammatical features across the two observed registers, this study continued to examine how each feature contributes to the corresponding latent variables, and then compared how the extracted factors between the registers may be differentiated. Table 12 reports on the loadings of each feature on the seven extracted factors (henceforth LG1-T1, LG2-T1, LG3-T1, LG4-T1, LG5-T1, LG6-T1 and LG7-T1 respectively) after the features with low loadings (below 0.30) were excluded in the Task 1 register. It can be seen that certain factors are loaded both positively and negatively by the observed features, the phenomenon of which conforms to the results by the previous studies (Biber, 1988; Biber, et al,. 2007; Xiao, 2009). In order to further describe what each factor mainly represents, this study attempted to abstract the lexico-grammatical features within a latent factor.

Table 12. Factor loadings of lexico-grammatical features (Task 1)

	Component						
	LG1-T1	LG2-T1	LG3-T1	LG4-T1	LG5-T1	LG6-T1	LG7-T1
private verb	.608						
emphatics	.408						
1st personal pronouns	.454						
BE as main verb	.335						
indefinite pronouns	.358						
WH questions	.373						
non-phrasal coordination	.332						
WH clauses	.357						
contraction	.435						
present tense verb	.394						
2nd personal pronouns	.442						
other nouns	-.546						
word length	-.593						
prepositions	-.362						
nominalization	-.480						
attributive adjectives	-.682						
demonstrative pronouns		.681					
WH relative clause		.547					
pied piping		.547					
past participial WHIZ		.360					
deletion							
demonstrative		.681					
3rd personal pronouns		-.302					
it as pronoun		-.361					
amplifiers			.384				
hedges			-.365				
public verb				.556			
suasive verb				.487			
necessity modal				.357			
agentless passive				.389			
that clause as verb				.565			
complements							
analytic negation					.433		
DO as pro-verb					.478		
perfect aspect					-.325		
that relative clause					-.383		
past tense verb						-.300	
time adverbial						-.321	
possibility modals							.448
all other adverbs							.410
split auxiliary							.376

As revealed in Table 12, LG1-T1, with heavy loadings (absolute values of loadings above 0.30) on a number of the features that are also included in the first dimension of Biber's (1988) framework, namely, *informational versus involved production*, the dimension can reflect the degree of orality/literacy to a great extent. For example, such features as *private verbs*, *WH questions* and *contraction* are clear indicators of informational production. Similar to Biber's (1988) framework, there are also a few features with negative loadings on this dimension, which can in fact contribute to a higher degree of literacy of the register under investigation. Given the above similarities, LG1-T1 is thus named *interactive versus elaborate discourse*. By "interactive discourse", it mainly refers to a setting, where two-way or multi-way communication is achieved, as opposed to "elaborate discourse", which in most cases refers to only one-way information, such as edited writing of clear and substantial information for readers without much effort in meaning negotiation. However, not all the features previously loaded on the first dimension in Biber's (1988) model are also loaded on LG1-T1. Table 12 shows that *amplifiers* and *hedges* are the only two features loaded on LG3-T1. Although these two features are semantically opposite in the sense that *amplifiers* serve a boosting function whilst *hedges* specify a lesser degree of certainty, their respective positive and negative loadings rightly become a pair in the observed register. Hence, LG3-T1 is named as certainty *versus* uncertainty. Back to LG2-T1 in Table 12, this study found that LG2-T1 is positively loaded with five lexico-grammatical features, yet negative with two features. As the negative features are *pronoun it*, which is the most generalized referent ranging from animate beings to abstract concepts (Biber, 1986; Chafe & Danielewicz, 1986) and *3rd personal pronoun*, which mark relatively inexact reference to persons "outside of the immediate interaction" (Biber, 1988, p. 225), this study somehow associates the name of this factor with the degree of reference. This is because the positively loaded features can, in an opposite direction, also be epitomized into exophoric referents (*demonstrative pronouns* and *demonstratives*), planned referents (*WH relative clause*, *pie-piping construction* and *past participial WHIZ deletion*). As such, LG2-T1 is named as *explicit versus implicit referents*.

LG4-T1, which includes *public verbs*, *suasive verbs*, *necessity modals*, *agentless passives* and *that clause as verb complements*, has all the positive loadings. Considering its main lexico-grammatical function of voicing out persuasive/obligatory discourse, this latent factor is named as such, namely, persuasive/obligatory discourse. Comparatively, LG5-T1 might be a factor with subtle difficulty in interpretability. On one hand, *analytic negation* and *DO as pro-verb*, two features relating to orality, are positively loaded on this factor. On the other hand, it has *perfect aspect* and *that relative clause* as its negatively loaded features. This study proposes to combine the main functions of these features and name the factor *informational reduction versus specification*.

LG6-T1 is also an interesting factor in that it has only two negatively loaded features: *past tense verbs* and *time adverbials*; hence, the factor is named *non-past temporal independence*. LG7-T1 has three negative features, although the absolute abstraction of which might be hard, it can be generalized as *possibilities of various degrees*. This is because *all other adverbs* can modify the conveyance of possibility to different extents in the given register.

Table 13. Lexico-grammatical factor correlation (Task 1)

Component	LG1-T1	LG2-T1	LG3-T1	LG4-T1	LG5-T1	LG6-T1	LG7-T1
LG1-T1	.995	.223	.246	.191	.154	.197	.257
LG2-T1		.958	.123	.049	.141	-.006	-.145
LG3-T1			.912	.192	-.179	-.217	-.055
LG4-T1				.894	.043	.142	-.147
LG5-T1					.941	.226	-.146
LG6-T1						.889	.197
LG7-T1							.841

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 13 lists the correlation results of the extracted lexico-grammatical factors. As can be seen, although the factors were rotated in a varimax manner, they are not significantly correlated. This indicates that the extracted factors are able to not only explain what is latently underlying in terms of lexico-grammatical features but also maintain their independence of each other.

Table 14. Factor loadings of lexico-grammatical features (Task 2)

	Component					
	LG1-T2	LG2-T2	LG3-T2	LG4-T2	LG5-T2	LG6-T2
private verbs	.466					
THAT deletion	.356					
contraction	.472					
present tense verb	.541					
analytic negation	.570					
DO as pro-verb	.468					
it as pronoun	.373					
causative subordination	.371					
indefinite pronouns	.394					
3rd personal pronouns	.441					
conditional subordination	.397					
other nouns	-.666					
prepositions	-.666					
attributive adjectives	-.543					
nominalization	-.490					
past participial WHIZ deletion	-.346					
phrasal coordination		.539				
by-passive		.545				
necessity modal		.456				
seem/appear		-317				
past tense verb		-.372				
demonstrative pronouns			.674			
WH relative clause			.616			
pied piping			.561			
demonstrative			.644			
1st personal pronouns				.438		
agentless passive				-.387		
split auxiliary					.372	
conjunct					.358	
that clause as adjective complements					.352	
predication modal						.357
public verb						.307
infinitive						.358
suasive verb						.304

Table 14 lists the loadings of each feature on the six extracted factors (henceforth LG1-T2, LG2-T2, LG3-T2, LG4-T2, LG5-T2 and LG6-T2 respectively) after the features with low loadings (below 0.30) were excluded in the Task 2 register. As can be seen, LG1-T2 has the largest number of lexico-grammatical features amongst all the factors. In addition, similar to LG1-T1, the features loaded on LG1-T2 are also related to orality/literacy. More specifically, the positively loaded features, such as *private verbs*, *THAT-deletion* and *contraction*, contribute to the degree of orality whilst those with negative loadings, like *attributive adjectives* and *nominalization*, mainly enhance literacy. Nevertheless, certain features project a combination of narrative and descriptive senses, such as present tense verbs and 3rd personal pronouns. Against this, LG1-T2 might adopt a slightly different name from that of LG1-T1: *interactive narration versus elaborative discourse*. LG2-T2 has three positive features and two negative features. It is a bit tricky to interpret this factor as the two polarities do not seem to be absolutely opposing to each other. Considering the main functions of these features, LG2-T2 is named *agent-explicit necessity versus past-specific academic hedging*.

The features heavily loaded (all above 0.60) on LG3-T2 pertain to exophoric referents (demonstratives and demonstrative pronouns) and planned referents (WH relative clause and pied piping construction). Therefore, this factor is named *exophoric and planned referents*. Comparatively, LG4-T2 is loaded with only two features, one positive (1st personal pronoun) and one negative (agentless passive). As these two features indicate whether or not the agent of a sentence is specified, LG4-T2 is named *specified versus unspecified agent*. LG5-T2 is loaded with three positive features, which mainly serve as a complement to the given information of a text. For example, conjunct connects more information with the already available information, and that clause as adjective complements naturally adds more information to a text. On account of these main functions, LG5-T2 would be better treated as *informational additive*. The last factor in Task 2 register is loaded with all positive features. When they are abstracted, an integration of their intended conveyance could be futurity (*predication modals* and *infinitive*) and overt persuasion (*public verbs* and *suasive verbs*). Thus, its name is specified as *futurity-projected overt persuasion*.

Table 15. Lexico-grammatical factor correlation (Task 2)

Component	LG1-T2	LG2-T2	LG3-T2	LG4-T2	LG5-T2	LG6-T2
LG1-T2	.914	.050	.169	.061	.284	.222
LG2-T2		.959	.059	-.140	-.202	.130
LG3-T2			.824	.138	.013	.145
LG4-T2				.894	.047	.189
LG5-T2					.818	-.156
LG6-T2						.983

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 15 lists the factor correlation results for Task 2 register. It can be noticed that similar to what has been found in Task 1 register, the lexico-grammatical registerial factors in Task 2 are also independent of each other even though they are rotated to a certain extent.

Table 16. Lexico-grammatical factor comparisons across registers

	Factors	Names
Task 1	LG1-T1	interactive *versus* elaborate discourse
	LG2-T1	explicit *versus* implicit referents
	LG3-T1	certainty *versus* uncertainty
	LG4-T1	persuasive/obligatory discourse
	LG5-T1	informational reduction *versus* specification
	LG6-T1	non-past temporal independence
	LG7-T1	possibilities of various extents
Task 2	LG1-T2	interactive narration *versus* elaborative discourse
	LG2-T2	agent-explicit necessity *versus* past-specific academic hedging
	LG3-T2	exophoric and planned referents
	LG4-T2	specified *versus* unspecified agent
	LG5-T2	informational additive
	LG6-T2	futurity-projected overt persuasion

The study confirms that, except for one factor in both registers, namely interactive narration *versus* elaborative discourse which has a substantial overlapping coverage, all the other extracted factors would carry discernible and distinct lexico-grammatical foci. This can demonstrate that the writing stimuli in the two tasks are able to elicit different registers. This could also build a validity argument that in terms of lexico-grammatical features, the employment of two writing tasks can be justified and validated.

The above analyses report the findings of the extracted factors across the two observed registers. In particular, the factor names are abstracted after a discrete consideration of the features loading on the factors. In order to probe into the potential dissimilarities in elicitation by the two registers, Table 16 compares the factor names across the registers. It is felt that except the first factors of both registers, all the other extracted factors would carry discernible and distinct lexico-grammatical foci, which can demonstrate that the writing stimuli in the two tasks are able to elicit different registers. This could also build a validity argument that in terms of lexico-grammatical features, the employment of two writing tasks can be justified and validated. The concern over the only overlapping factors between the registers might be addressed by the fact that previous studies under a similar framework also found the first factor dealt with orality/literacy (e.g. Biber, 1988; Xiao, 2009). Therefore, even though the first factors of each register are seemingly the same, it might be acceptable considering similar results from other studies when registers were profiled against the MD-MF framework, where the degree of orality/literacy is usually first demarcated in the first factor.

3.4.2.2. Exploring semantic factors

Having explored the lexico-grammatical features of the observed registers, this section turns to the analysis of semantic factors. It should be noted that as *WMatrix* includes different layers of semantic categorizations, this study only looked at the second layer, where 21 broad categorizations are broken down within themselves. As for more fine-grained sub-categorizations, such as A.5.3 (*accuracy*) under A.5 (*evaluation*), they are beyond the scope of the analyses in this section. This is because an even larger number of semantic categories might cause fragmentary picture of factor analysis, which might bring constraints to the interpretation of the extracted latent factors.

Similarly, prior to running the EFAs, this study checked KMO and Bartlett's Test of Sphericity, the outcomes of which are revealed in Table 17 and Table 18. Given that the indices well fit the required thresholds of EFA, with KMO between 0.7 and 0.8 and Bartlett's Test of Sphericity values being significant, EFAs were then performed on the datasets representative of two observed registers.

Table 17. KMO and Bartlett's Test for semantic categories (Task 1)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.773
Bartlett's Test of Sphericity	Approx. Chi-Square	1.490E4
	df	6105
	Sig.	.000

Table 18. KMO and Bartlett's Test for semantic categories (Task 2)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.736
Bartlett's Test of Sphericity	Approx. Chi-Square	1.462E4
	df	5995
	Sig.	.000

In order to determine the number of factors to be extracted, we first examined the natural breaks on the curve of eigenvalues. As reflected in Figure 5, the natural breaks occur when the factor numbers range from 4 to 8. Following a similar approach as the present study did for extracting lexico-grammatical latent factors earlier, we conducted multiple factor analyses by manually setting the number of retained factors as 4 to 8 accordingly. After comparisons of the factorial structures based on 4 to 8 factors in terms of the number of significant loadings (above 0.30) on each extracted factor, cross loadings as well as the ease of interpretation, this study established a five-factor semantic factorial structure on the basis of 500 writing scripts for Task 1 register. Table 19 lists the cumulative eigenvalues and extraction sums of squared loadings of the extracted factors. After five factors were extracted, the eigenvalues of the extracted factors remained above 2.0, explaining a cumulative variance of 19.744%. Although this cumulative value is lower than what has been reported above on lexico-grammatical latent factors, the results can be deemed as acceptable given the fact that there is a larger number of semantic (sub)categories under observation. The results can also be justified considering the fact that since the semantic categorizations specified by *WMatrix* are already independent of each other, it might be somehow hard to cluster different sub-categories into just a few latent factors.

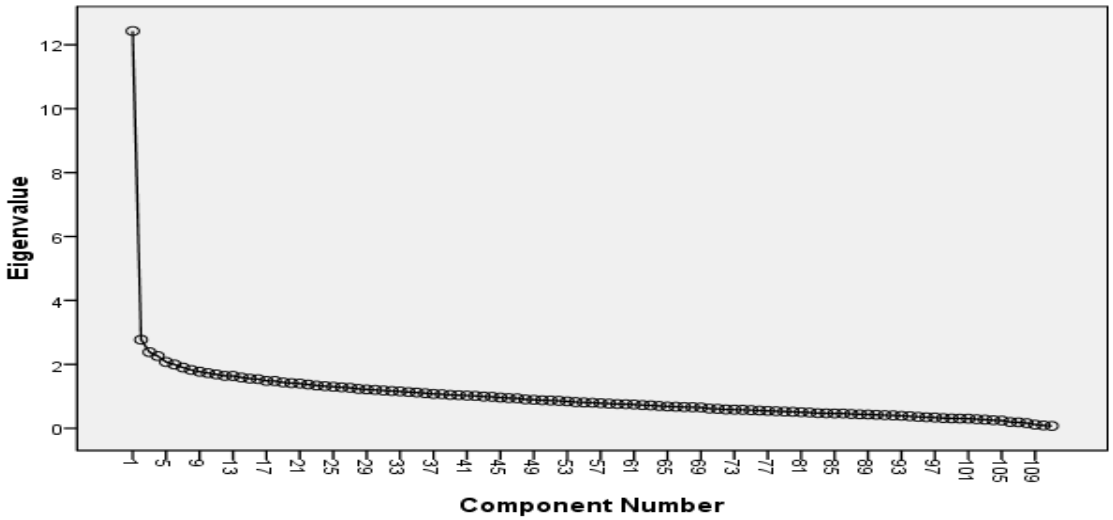


Figure 5. Scree plot of semantic categories (Task 1)

Table 19. Factor extraction of semantic categories (Task 1)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	12.428	11.196	11.196	12.428	11.196	11.196
2	2.772	2.498	13.694	2.772	2.498	13.694
3	2.383	2.146	15.840	2.383	2.146	15.840
4	2.260	2.036	17.876	2.260	2.036	17.876
5	2.073	1.867	19.744	2.073	1.867	19.744

In comparison, after observing the natural breaks (factor number ranging from 4 to 8) of the curve of eigenvalues for Task 2 register as illustrated in Figure 6, we again compared the factorial structures based on 4 to 8 factors in terms of the number of significant loadings (above 0.30) on each extracted factor, cross loadings as well as the ease of interpretation. In the end, this study also established a five-factor semantic factorial structure on the basis of 500 writing scripts for Task 2 register. Table 20 lists all the extracted factors which have eigenvalues above 2.0, as well as the cumulative percentage of the variance explained (19.072%).

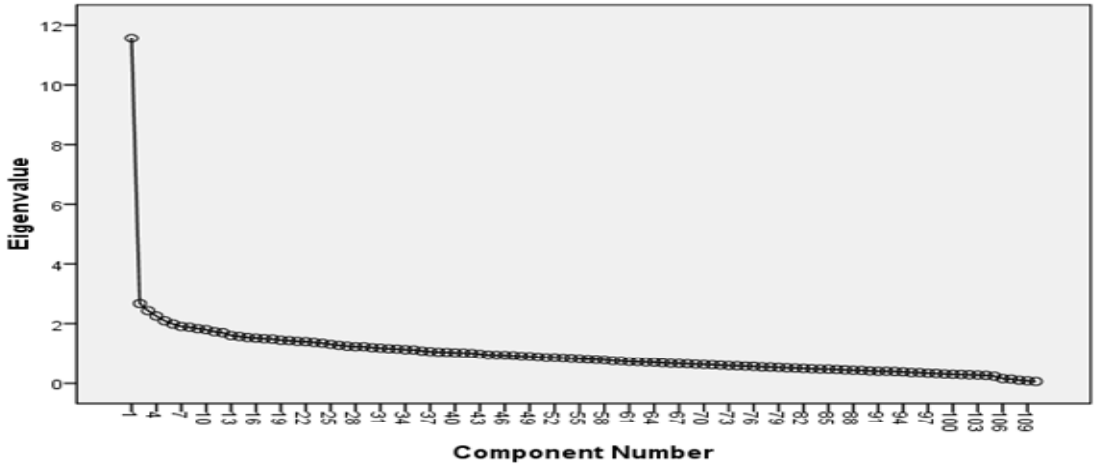


Figure 6. Scree plot of semantic categories (Task 2)

Table 20. Factor extraction of semantic categories (Task 2)

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	11.556	10.505	10.505	11.556	10.505	10.505
2	2.662	2.420	12.926	2.662	2.420	12.926
3	2.424	2.204	15.130	2.424	2.204	15.130
4	2.247	2.043	17.173	2.247	2.043	17.173
5	2.090	1.900	19.072	2.090	1.900	19.072

Based on the above, it can be tentatively concluded that both registers extract the same number of factors, yet whether or not the factors represent the similar interpretations remains to be further explored. As such, in order to expound the extracted factors and further compare the potential overlapping or discrepancies between the observed registers in relation to semantic categories, this study further examined the factor loadings in each register and attempted to name the factors accordingly. Table 21 indicates the distribution of factor loadings and how each semantic category contributes to the latent factors significantly. The extracted factors are labeled as S1-T1, S2-T1, S3-T1, S4-T1 and S5-T1 respectively. For the sake of easier reference, except for the semantic categories contributing to S1-T1, the variables are also attached with what they represent semantically. As S1-T1 explained a large portion of the variance, this factor constitutes most complicated interpretation among all the extracted factors. A glance at the variables contributing to S1-T1 would lead to a confusing picture as a good number of different semantic categories are mixed up. Therefore, this factor may not be easily abstracted simply based on the semantic labels. However, since this factor accounts for a large proportion of the semantic categories of Task 1 written output, it would be highly possible that S1-T1 accordingly should cover the semantic categories of the keywords. Driven by this possibility, this study turned to the keywords as compared with the British National Corpus (BNC). Table 22 lists the top 30 keywords generated by the keyword function of *WordSmith Tools*. After a comparison between the variables positively loaded on S1-T1 and the semantic categories of the top 30 keywords, it is found that an overwhelming majority of the extracted keywords can be accorded with the variables positively contributing to S1-T1, except for *internet* (rank 8th), *hotel* (rank 11th), *their* (rank 18th) and *can* (rank 25th), as marked in the shaded areas, and that those negatively loaded semantic categories (shaded areas in Table 21) cannot contribute to the specified semantic categorizations of the keywords. Therefore, it can be generally understood that although the variables contributing to S1-T1 diversify semantically, they as a whole can represent a semantic entity that is holistically elicited by the writing prompt. Based on the above observation, S1-T1 is named *topic-specific semantic aboutness*.

Table 21. Factor loadings of semantic categories (Task 1)

	Component				
	S1-T1	S2-T1	S3-T1	S4-T1	S5-T1
A.2	.569				
A.5	.758				
A.11	.362				
A.15	.468				
C.1	.852				
F.1	.471				
G.1	.604				
H.3	.388				
I.2	.724				
K.4	.331				
L.1	.326				
M.1	.585				
M.7	.796				
N.3	.419				
O.1	.426				
O.4	.325				
Q.2	.454				
Q.4	.455				
S.1	.713				
S.2	.362				
S.5	.638				
S.8	.370				
T.3	.557				
W.1	.390				
W.3	.442				
X.2	.686				
Y.2	.327				
A.3	-.414				
A.7	-.306				
A.8	-.406				
A.12	-.353				
F.4	-.507				
Q.1	-.616				
Q.3	-.335				
X.6	-.304				
X.7	-.387				
Y.2	-.885				
K.2 music and related activities		.425			
K.6 children's games and toys		.476			
M.6 location and direction		.311			
T.1 time		.352			
H.1 architecture and kinds of houses & buildings			.701		
H.4 residence			.689		
H.2 parts of buildings			.432		
B.2 health and disease				.432	

B.4 cleaning and personal care	.327	
B.5 clothes and personal belongings	.325	
N.1 numbers		.339
N.4 linear order		.388

Table 22. Top 30 key words in Task 1 written production

Rank	Key Word	Keyness	Semantic Category
1	REVIEWS	29948.60547	X.2
2	TOURISM	29702.50391	M.1
3	ONLINE	22511.95898	Y.2
4	CULTURAL	21081.44336	C.1
5	TRAVEL	12741.80371	M.1
6	TRAVELERS	11336.73633	M.1
7	WEBSITES	5357.841797	Y.2
8	INTERNET	4686.21582	Z.1
9	TRAVELING	4190.831543	M.1
10	PEOPLE	3862.563721	S.2
11	HOTELS	3472.573242	H.4
12	TOURISTS	3419.784668	M.1
13	INFORMATION	3409.771729	X.2
14	BUSINESSES	3271.211914	I.2
15	COMMUNITIES	3243.070313	S.5
16	CULTURE	2654.42749	C.1
17	GUIDEBOOKS	2350.073486	S.8
18	THEIR	2282.72876	Z.8
19	LOCAL	2166.955078	M.7
20	NATIVE	2144.293213	M.7
21	CULTURES	2122.098877	C.1
22	RESTAURANTS	2026.178101	F.1
23	OPINIONS	2001.577515	X.2
24	CUSTOMERS	1967.551025	I.2
25	CAN	1909.111938	A.7
26	TRADITIONAL	1885.506104	S.1
27	LOCALS	1700.009766	M.7
28	ARTICLE	1575.886353	Q.4
29	DISADVANTAGES	1440.995972	A.5
30	COMMENTS	1367.164673	Q.2

S2-T1 has four positively loaded variables, namely K.2 (*music and related activities*), K.6 (*children's games and toys*), M.6 (*location and direction*), and T.1 (*time*). As K.2 and K.6 are concerned with entertainment whereas M.6 and T.1 contribute to place and time respectively, this factor is named time- and location-specific entertainment. In a way, it can be imagined that, in producing content related to this semantic factor, test-takers might be mentally involved with scenarios related to traveling and entertainment. This is because in the observed written output of Task 1, test-takers were expected to compare and contrast the opposing views on "cultural tourism" (2010) and "online travel reviews" (2011). Both topics tend to elicit written output concerning time- and location-specific entertainment activities. S3-T1 seems more straightforward in interpretation as the three positively loaded semantic

categories are about architecture and housing. Hence, this factor is also named as such. Likewise, S4-T1 comprises three similar semantic categories, namely, B.2 (*health and disease*), B.4 (*cleaning and personal care*) and B.5 (*clothes and personal belongings*). Therefore, based on the abstracted meanings above, S4-T1 is named hygiene and individual. There are only two semantic categories loaded on the last factor: N.1 (*numbers*) and N.4 (*linear order*). Given the simplicity of this factor, S5-T1 is named numbers and sequences. Table 23 lists the correlation results of the extracted semantic factors for Task 1. Generally these five factors are not highly correlated with each other, which can lend support to the independence of latent factor after rotation and also justify the different naming for the factors.

Table 23. Semantic factor correlation (Task 1)

Component	S1-T1	S2-T1	S3-T1	S4-T1	S5-T1
S1-T1	.985	-.083	-.131	-.005	.077
S2-T1		.846	.011	.161	-.039
S3-T1			.968	.015	-.199
S4-T1				.764	-.114
S5-T1					.759

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 24. Factor loadings of semantic categories (Task 2)

	Component				
	S1-T2	S2-T2	S3-T2	S4-T2	S5-T2
A.1	.304				
A.2	.301				
A.4	.538				
A.5	.423				
A.6	.529				
A.12	.368				
B.1	.635				
B.2	.333				
E.2	.633				
E.3	.885				
E.4	.316				
E.6	.386				
G.2	.865				
I.1	.870				
I.3	.776				
M.7	.566				
N.5	.302				
O.4	.479				
P.1	.900				
S.1	.629				
S.6	.308				
S.7	.376				
T.1	.443				
T.2	.457				
T.3	.549				
A.15	-.368				

I.2	-.464			
M.1	-.438			
M.6	-.358			
N.1	-.698			
O.2	-.353			
Y.2	-.855			
A.3 being/existing		.315		
A.7 definite (+modals)		.341		
A.13 degree		.300		
A.14 exclusivizers/particularizers		.309		
B.5 clothes and personal belongings			.397	
G.1 government, politics and elections			.300	
G.3 warfare, defense and the army			.344	
Q.4 the media			.374	
W.1 the universe				.364
W.3 geographic terms				.328
X.2 mental actions and processes				.394
X.8 trying				.307
X.9 ability				.320
Q.1 communication				-.409
Q.2 speech acts				-.337
S.2 people				.411
S.5 groups and affiliation				.302

Table 24 lists the factor loadings of the semantic categories extracted from Task 2 written production. The factors are respectively labeled as S1-T2, S2-T2, S3-T2, S4-T2 and S5-T2. When the first factor in Table 24 was interpreted, similar difficulty was encountered. This was because S1-T2 seemed to be highly inclusive of various specified semantic categories. Therefore, a similar approach used in Task-1 register above was adopted so as to see if this factor could represent the semantic keyness of the elicited register. Table 25 outlines the top 30 keywords with their semantic categories attached. It can be seen that once again, except for *people* (rank 9th), *should* (ranked 15th), *are* (ranked 18th), *parents* (ranked 24th) and *their* (ranked 26th), the semantic categories of almost all the keywords can be matched with the variables positively contributing to the first factor in Table 24. The reason for this matching might be the same as what has been previously discussed. The combined effect of S1-T2 can account for a semantic entity of the observed register. Therefore, S1-T2 is named *topic-dependent semantic aboutness*.

Table 25. Top 30 key words in Task 2 written production

Rank	Key word	Keyness	Semantic category
1	BULLYING	24124.28516	E.3
2	ATM	23295.00586	I.1
3	FRAUD	14858.82129	G.2
4	TEACHERS	9299.566406	P.1
5	STUDENTS	8634.834961	P.1
6	BULLIED	8395.496094	E.3
7	VICTIMS	5816.87793	A.1
8	CASES	5252.830566	A.4
9	PEOPLE	3707.945068	S.2
10	FIGURE	2874.195557	N.5
11	MONEY	2636.187744	I.1
12	PEERS	2608.404541	A.6
13	FRAUDS	2547.60498	G.2
14	CAMPUSES	2484.849854	M.7
15	SHOULD	2453.363525	S.6
16	DISLIKE	2432.279297	E.2
17	BULLY	2323.390381	E.3
18	ARE	2160.359863	A.3
19	BULLIES	2131.982422	E.3
20	PERSONALITY	2032.258545	S.1
21	PERCENT	1868.73584	N.5
22	CAMPUS	1831.122437	M.7
23	REASONS	1764.30127	A.2
24	PARENTS	1740.398682	S.4
25	AGE	1519.075073	T.3
26	THEIR	1512.750366	Z.8
27	CRIMINALS	1345.303955	G.2
28	NUMBER	1267.777222	N.5
29	REASON	1202.341797	A.2
30	TEMPORARILY	1202.322754	T.1

S2-T2 includes four positively loaded semantic categories, which are A.3 (*being/existing*), A.7 (*definite+modals*), A.13 (*degree*) and A.14 (*exclusivizers/particularizers*). It can be generally felt that except for A.3, the remaining semantic categories mainly serve as determiners. Therefore, when the naming of this factor is considered with the inclusion of A.3, *conditional existence and presentation* might be an appropriate name. This is because the category of *being/existing* also includes a large number of verbs that presents the figures or tendencies as reflected by the given charts of the tasks.

S3-T2 includes four semantic categories that are completely relevant to each other, thus seemingly complicating the naming of this factor. However, with reference to the writing prompts of Task 2 in both test administrations, it can be found that test-takers were required to produce readers' letters to the press (Q.4 *the media*), and that the prompts, which are *ATM fraud* and *bullying on campus*, may be regarded more or less pertaining to crime or offense in society (B.5 *clothes and personal belongings*; G.1 *government, politics and elections*; G.3 *warfare, defense and the army*). Thus, this factor is named *crime-related media facts*. Notwithstanding that this factor seems to somehow overlap with the first broader factor, its

skewed presence shows a more focused content domain of what is elicited. This can be evidenced by the fact that G.2 (*crime, law and order*) is actually loaded positively and heavily on S1-T2 (factor loading equal to 0.865).

The factor S4-T2 is loaded with not only positive but also negative semantic categories. In particular, the two polarities can be detected when a dividing line is drawn in terms of psychological (X.2 *mental actions and processes*; X.8 *trying*; X.9 *ability*) and linguistic (Q.1 *communication*; Q.2 *speech acts*) actions/processes. Apart from these, this factor also seems to be relevant to the environment (W.1 *the universe;* W.3 *geographic terms*). The factor is thus named *environment-specific psychological versus linguistic actions/processes*. When it comes to S5-T2, there are only two positively loaded semantic categories: S.2 (*people*) and S.5 (*groups and affiliation*). In general, this factor is a reference to the criminal/victim because both ATM fraud and bullying on campus might involve two parties. In terms of criminals, the reference could be either an individual or a gang. Likewise, victims can be referred to either specifically or implicitly. Therefore, S5-T2 is named criminal- and victim-related reference.

Table 26. Semantic factor correlation (Task 2)

Component	S1-T2	S2-T2	S3-T2	S4-T2	S5-T2
S1-T2	.991	.058	.029	-.099	.058
S2-T2		.754	.151	-.050	-.025
S3-T2			754	.191	.205
S4-T2				.746	.267
S5-T2					.705

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 26 lists the correlation results of the extracted semantic factors after varimax rotation. It was found that all the factors were generally independent, with all the absolute values of the correlation coefficients below 0.30. This indicates that even though S3-T2 and S5-T2 somewhat overlap in their factor names, they are not related (correlation coefficient equal to 0.205).

Table 27. Semantic category factor comparisons across registers

	Factors	Names
Task 1	S1-T1	topic-specific semantic aboutness
	S2-T1	time- and location-specific entertainment
	S3-T1	architecture and housing
	S4-T1	hygiene and individual
	S5-T1	numbers and sequences
Task 2	S1-T2	topic-specific semantic aboutness
	S2-T2	conditional existence and presentation
	S3-T2	crime-related media facts
	S4-T2	environment-specific psychological *versus* linguistic actions/processes
	S5-T2	criminal- and victim-related reference

Although EFA in relation to semantic categories cannot depict an entire picture of what can all be abstracted semantically, the method can at least portray an epitome of the observed registers. Akin to the synthesized table in terms of lexico-grammatical features, Table 27 lists the semantically latent factors extracted. It should be noted that the first extracted factors

across the registers are similar. However, it should also be noted that all the remaining factors are different from each other. Tentatively, the above factorial structure of both registers can be understood as the first factor (*topic-specific semantic aboutness*) being a trunk with all the others factors serving as branches, in the case of which one tree (Task 1 register) can be distinguished from the other (Task 2 register). Based on the above analyses, another validity argument can thus be articulated that what were elicited in the two observed registers are semantically distinct from each other and that the two tasks of the GEPT Advanced level writing assessment are indeed assessing different facets of candidates writing proficiency.

3.4.3. Register analysis: Academicality

As the last dimension of register analysis in this study centers upon the degree of being academic, this section is primarily concerned with profiling the observed registers at the word and phraseology levels in comparison with AWL (Coxhead, 2000) and the Academic Formulaic List (AFL) (Simpson-Vlach & Ellis, 2010).

Table 28. Academic word coverage

	TOKENS/percentage	TYPES/percentage	FAMILIES
Task 1	17007/ 8.16	1187/12.80	472 (82.81%)
Task 2	9275/ 4.69	1098/12.72	460 (80.70%)

Table 28 shows the results when the observed written output of the GEPT Advanced level candidates were profiled against AWL via a computer software program Range (Heatly, Nation & Coxhead, 2002). It can be seen that, in terms of word families, more than 80% of the academic words in AWL, which contains 570 words in total, also appear in the two mini-corpora of the present study, which suggests that academic words are extensively used by the candidates in their writing. However, families alone cannot provide a full picture of the degree to which the observed texts are academic because even though the academic words are present in the candidates' written output, their frequencies could still be low. Because of this, it is also necessary to refer to relevant information on tokens/percentage (Table 28). Coxhead (2010), when documenting the studies that applied AWL within a decade since the word list was published, pointed out that on average the AWL covered approximately 10% of the vocabulary in the written academic corpora of various disciplines. For example, Coxhead and Hirsh (2007) found that the coverage of academic words in their observed science corpus reached 8.96%. Therefore, the written output from Task 1, which has 8.16% coverage of academic words, basically falls into an acceptable range of academicality. Nonetheless, the Task 2 written output presents a comparatively low percentage (4.69%) of academic words. This may be due to the possibility that even though the candidates were able to use some academic words in Task 2, the presence of such words was sparse in the output. This indicates that if Task 2, with mostly nonverbal input in the writing prompts, is intended for assessing test-takers' written proficiency from the angle of academic engagement, the relevant data provided for this present study have not produced such supportive evidence at the word level. However, this could potentially be attributable to the artifact idiosyncratic to this particular dataset, collected from only two years' administrations; a wider range of writing prompts might result in a different interpretation.

Well above the individual word level, meaningful combinations of multiple words, or phraseologies, was also investigated in this study. Simpson-Vlach and Ellis (2010) generated an academic formulaic list based on the mutual information values and the frequencies of the phraseologies. In this study, the written AFL, which includes top 200 3- to 5-grams, was

referred to. Unlike individual words, which can be filtered by AWL for determining the percentages of coverage of individual academic words, phraseologies needed to be extracted first for further processing. In addition, topic factor may play an even more determining role for the uncontrolled generation of 3- to 5-grams list because, within one writing task for each test administration, a number of phraseologies could be especially topic relevant and might thus be extracted. Based on the above consideration, a frequency threshold of three occurrences was set up, which automatically functioned to extract the 3- to 5-gram phraseologies across the two registers. This threshold was chosen after many times of trial extractions. When the threshold value was set below 3, the chances would be that quite a large number of meaningless phraseologies or those with contractions would be incidentally extracted. However, if the threshold was set above 3, there would be a risk of under-extraction, thus resulting in a lower coverage of formulae than it ought to. When the threshold was set at 3, the total numbers of 3- to 5-gram phraseologies were 10,609 and 11,039 for Task 1 and Task 2 outputs respectively. Since the purpose of this profiling was to compare the observed registers in terms of *academicality* from the perspective of academic formulae, the topic impact was tentatively negligible. This was because even though the top ranking phraseologies in one register might be highly related to the writing stimuli, after they were profiled against AFL and further compared across the registers, those topic-related phraseologies would already be excluded. Therefore, human judgment and manual intervention are necessary in the process of determining what phraseologies may be topic-related and whether they should remain or be removed from the formula lists.

Table 29. Academic formula coverage

	No. (coverage) of academic formulae	Frequency of all academic formulae	Standardized frequency (per 250 words)
Task 1	55/200 (27.5%)	974	1.17
Task 2	63/200 (31.5%)	817	1.01

Table 29 lists the academic formulae coverage for the observed written productions. It can be found that a limited number of academic formulae (55 formulae for Task 1 and 63 formulae for Task 2) out of the 200 formulae in the written AFL could actually be found. When all the academic formulae found in the candidates' writing were standardized in relation to the length of their written production (250 words as required in the instructions), the frequency of occurrences of academic formulae was rather low. For each task, candidates on average used only about one formula with academic characteristics (Table 29).

Even though no existing studies would propose any reference or threshold percentage for a satisfactory coverage by AFL, it might be perceived that the figures above reflect a low degree of academicality. The finding therefore suggests that, even when the observed registers exhibit a satisfactory coverage of individual academic words, the same cannot be said about academic phraseologies. This might be attributable directly to the candidates' written English proficiency. Without an awareness of academic writing, candidates would likely produce English writing just for general communication purposes. This calls for attention of the GEPT test developers. It is suggested that, if the GEPT test developers intend to make Task 2 more academic, its design could be further fine-tuned. For example, instead of instructing candidates to produce letters to the press on certain social issues, Task 2 might contextualize candidates into an academic activity, where describing, summarizing and discussing the nonverbal input can also be materialized.

Table 30 ranks all the individual academic formulae found in the observed written output, along with their respective frequencies. For Task 1, the first three top-ranked academic formulae are basically the three variations of *(on) the other (hand)*, whose cumulative frequencies account for a large proportion (57.29%) of the total academic formulae frequency. This indicates that although Task 1 written production captures 55 academic formulae according to the written AFL, there may still be a huge imbalance in the *de facto* use of these formulae (frequency range equals to 189). Comparatively speaking, Task 2 register does not presents a similar tendency; there seems to be a steady decrease in frequency from the top ranked academic formulae to the bottom ones (frequency range equals to 57).

Table 30. Academic formula frequencies

Rank	Task 1	frequency	Rank	Task 2	frequency
1	on the other	192	1	on the other	60
2	the other hand	185	2	the other hand	53
3	on the other hand	181	3	on the other hand	52
4	are able to	43	4	the most important	34
5	the most important	17	5	his or her	30
6	be used to	15	6	should also be	29
6	important role in	15	7	they do not	27
8	should also be	14	8	to the fact that	26
9	it is important	13	9	it is important	23
10	can be found	12	10	less likely to	22
10	to ensure that	12	10	most likely to	22
12	as a whole	11	12	due to the fact	20
12	can also be	11	12	due to the fact that	20
12	needs to be	11	12	in some cases	20
12	wide range of	11	12	it is clear	20
16	can be used to	10	15	should not be	19
16	his or her	10	16	it is clear that	16
18	to do so	9	17	are able to	15
18	to the fact that	9	17	to do so	15
20	a large number	8	19	can be seen	14
20	a large number of	8	19	if they are	14
20	a wide range	8	19	needs to be	14
20	a wide range of	8	22	it is possible	13
20	they do not	8	22	there are several	13
20	which can be	8	22	there has been	13
26	depend on the	7	25	important role in	12
26	it is important to	7	25	it is possible that	12
26	it is necessary	7	25	shown in figure	12
29	are likely to	6	28	this means that	9
29	can be seen	6	29	are likely to	8
29	due to the fact	6	29	it is important to	8

29	due to the fact that	6	29	it is necessary	8
29	it is necessary to	6	32	as can be seen	7
29	it is worth	6	32	can also be	7
29	there has been	6	32	in most cases	7
36	are as follows	5	32	it is impossible	7
36	if they are	5	32	it is obvious that	7
36	in some cases	5	32	to ensure that	7
36	in this article	5	38	are as follows	6
36	it is impossible	5	38	it is worth	6
36	there are several	5	40	as a consequence	5
36	to ensure that the	5	40	as a whole	5
36	we do not	5	40	as shown in	5
44	are based on	4	40	be explained by	5
44	as a consequence	4	40	even though the	5
44	be seen as	4	40	it is difficult	5
44	it has been	4	40	that there is no	5
44	little or no	4	40	we do not	5
44	whether or not the	4	48	give rise to	4
50	depending on the	3	48	is likely to	4
50	in the form of	3	48	it has been	4
50	is based on the	3	48	it is impossible to	4
50	is likely to	3	48	there are no	4
50	it is impossible to	3	53	be carried out	3
50	that it is not	3	53	be used to	3
			53	can be found	3
			53	can be seen in	3
			53	have shown that	3
			53	is more likely	3
			53	it is likely that	3
			53	that it is not	3
			53	the total number	3
			53	total number of	3

In order to further align the academic formulae in the observed registers with those in the written AFL, this study mapped the corresponding ranks of the identified academic formulae in relation to the top 10 AFL formulae. As can be seen in Table 31, the first ranked formula in the written AFL (*on the other hand*) is ranked the third in both registers, whilst the second ranked formula (*due to the fact that*) is only ranked 29th in Task 1 register and 12th in Task 2 register respectively. What seems somehow strange is that for the remaining high-ranking academic formulae in the written AFL, only one (*a wide range of*) can be found in the above extracted formulae list (ranked 20th in Task 1 only), but all the other formulae are absent from the observed GEPT Task 1 and Task 2 outputs. Even though certain variants of the top-ranked AFL formulae have been identified, such as *it is impossible to* in place of *it is not possible to*, it is rather disappointing to find that a good number of frequently used academic formulae by

native speakers in their written production were not found in the observed outputs. This again suggests that there is much room for GEPT candidates to improve the academic register of their written output since the GEPT writing tasks are mainly intended for assessing candidates' suitability for academic studies.

Table 31. Academic phraseologies in comparison with the written AFL (Top 10)

Academic Formula	Rank in AFL	Rank in Task 1	Rank in Task 2
on the other hand	1	3	3
due to the fact that	2	29	12
on the other hand the	3	/	/
it should be noted	4	/	/
it is not possible to	5	50 (*it is impossible to*)	48 (*it is impossible to*)
a wide range of	6	20	/
there are a number of	7	/	/
in such a way that	8	/	/
take into account the	9	/	/
as can be seen	10	/	19 (*can be seen*)

4. Conclusion and Recommendations

4.1. Summary of Main Findings in Addressing the Research Questions

RQ1: Compared with previous register studies, what are the relative positions of written registers for different types of writing task as evidenced by the output of the GEPT Advanced level candidates?

The dimension scores for the Task-1 and Task-2 registers diverge to some extent, with the Task-1 register nearing the expected elicitation of written academic English and the Task-2 register approximating the register of spoken discourse. Although this approximation of orality in the test takers' response data for Task 2 might be attributable to the candidates' *de facto* performance idiosyncratic to this particular sample rather than the effects of the writing prompts, it is still possible that the test instructions in Task 2 have a role to play in causing this orality in the writing. Compared with Task 2, the verbal input in Task 1 seems to encourage test-takers to refer to certain formal expressions, which explains why the main register elicited by Task 1 tends to approximate the literacy end on the continuum of registers. The same, however, cannot be said for Task 2. Nevertheless, considering that the GEPT Advanced Level assesses proficiency levels of both general English and academic English, a balance of different registers between the two tasks in this way may still be desirable.

RQ2: What register features may exist in candidates' written production across the two writing tasks? To what extent can the registers elicited by different tasks be distinguished from each other at lexico-grammatical and semantic levels?

At the level of lexico-grammatical register features, the registers elicited by the two writing tasks present a wide range of differences in terms of latent factors. Factor analyses confirm that, except for one factor in both registers, namely, *interactive narration versus elaborative discourse,* which has a substantial overlapping coverage, all the other extracted factors carry discernible but distinct lexico-grammatical foci. This result confirms that the writing stimuli

in the two tasks were able to elicit different registers. This finding thus contributes positively to the validity argument that, in terms of lexico-grammatical features, the design of the two writing tasks are suitably justified as different writing tasks are capable of eliciting distinct registers.

At the level of semantic features, the first extracted factors (*topic-specific semantic aboutness*), which appear to represent topic specificity across the two registers of Task 1 and Task 2, are extremely similar. This similarity can be understood and expected although the writing prompts for the two tasks are intrinsically intended for different registers semantically. Meanwhile, all the remaining factors were found to differ from each other between the two task registers. Based on these results, the factorial structure of both registers appears to suggest a double-tree structural relationship, wherein the first factor stands as a trunk with all the others factors serving as branches and, in addition, one tree (Task-1 register) can clearly distinguishes itself from the other (Task-2 register). This finding can articulate another validity argument that what were elicited in the observed registers is semantically distinct from each other, which justifies the employment of two writing tasks in the GEPT Advanced Level Writing Test.

RQ3: To what extent is the GEPT Advanced level examinees' written production from different tasks deemed academic?

With regard to the extent to which the elicited registers appear to be academic, it was found that the written scripts basically present a satisfactory coverage (8.16%) of individual academic words for Task 1, but the percentage of coverage (4.63%) for Task 2 was well below the threshold level as set by Coxhead (2010). On the other hand, the coverage of academic formulae was even lower in the elicited output from both tasks, since the standardized frequencies of such use by each candidate were just 1.17 for Task 1 and 1.01 for Task 2 in a 250-word essay. Therefore, even though the observed registers exhibit an expected coverage of individual academic words, the same cannot be said of the frequency of the use of academic phraseologies for both tasks. This could be due to the fact that the list by Simpson-Vlach and Ellis (2010) was created based on various types of academic writing, whereas there were only two types of essay task in the present study. Another possible reason for this sparse use of academic formulae might be candidates' levels of writing proficiency and lack of awareness of importance and features of academic writing. With the latter in particular, GEPT candidates may produce English writing just for general communication purposes. However, this low coverage could also be due to the artifact idiosyncratic to this particular dataset. Further investigation is needed to better understand the issue.

4.2. Recommendations

Recommendation 1: Promoting the Importance of English Academic Writing

In the international community of English language education at the tertiary level, a central agenda in recent years involves tremendous efforts to align teaching, learning and assessments with curricula in a systematic, effective and productive manner (Cumming, 2009). Steering the pedagogical focus from English for general purposes (EGP) to English for academic purposes (EAP) has been an important effort associated with this agenda. In Hong Kong, EAP teaching has been adopted in practice for decades with some level of success; in the tertiary education system in the Chinese mainland, changing the pedagogical orientation to EAP from EGP has been an important goal for elite universities in recent years. It is our belief that there is also an urgent need to promote EAP, of which English academic writing is an important

component, in schools and universities in Taiwan and raise the awareness of the usefulness of academic English for their future studies so that students will be able to see the need for improving their academic writing and consciously learn to understand how they can write in English more formally and academically.

Recommendation 2: Modifying the Orientation for Writing Task 2

Results of the present study suggest that there is still room for enhancing the level of academicality for the GEPT Advanced Level Writing Task 2. If this is indeed the direction in which the GEPT developers would like to see the test move, it is suggested that the writing stimuli be contextualized in more academic settings so as to encourage test-takers to produce their responses to the stimuli in more formal language. For example, instead of being a piece of writing for the Opinion Section of a newspaper, Task 2 could be contextualized more formally so as to encourage the use of academic language. This modification might lead to a written production from the examinees with more academic flavor, thus creating a beneficial washback to encourage even more practice in academic writing. However, to confirm the necessity for this modification, it is advisable that further empirical research with larger samples and a wider range of writing prompts be conducted.

5. References

Aijmer, K. (2002). *English discourse particles*. Amsterdam: John Benjamins.

Biber, D. (1986). Adverbial stance types in English. *Discourse Processes*, *11*, 1-34.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*, 263-286.

Biber, D., Conrad, S., Reppen, R., Byrd, P., & Helt, M. (2002). Speaking and writing in the university: A multi-dimensional comparison. *TESOL Quarterly*, *36*, 9-48.

Biber, D., Connor, U., & Upton, A. T. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins.

Chen, G., & Chang, M. (2008). Assessing listening at elementary level: A GEPT case study. *Paper presented at the 17th International Symposium on English Teaching*, Taipei.

Chih-Hua, K. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, *18*, 121-138.

Chin, J. & Kuo, G. (2004). A validation report on the superior level GEPT. *Paper presented at the 13th International Symposium on English Teaching*, Taipei.

Chin, J. & Wu, J. (2001). STEP and GEPT: A concurrent study of Taiwanese EFL learners' performance on two tests. *Proceedings of the Fourth International Conference on English Language Testing in Asia* (pp. 22-44), Taipei.

Connor, U. & Upton, T. A. (2003). Linguistic dimensions of direct mail letters. In C. Meyer & P. Leistyna (Eds), *Corpus analysis: Language structure and language use* (pp. 71-86). Amsterdam: Rodopi.

Connor, U. & Upton, T. A. (2004). The genre of grant proposals: A corpus linguistic analysis. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 235-256). Amsterdam: John Benjamins.

Conrad, S. (2001). Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In S. Conrad & D. Biber (Eds.), *Variation in English: Multidimensional studies* (pp. 94-107). Harlow: Longman.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Coxhead, A. (2010). The Academic Word List 10 years on: Research and teaching implications. *TESOL Quarterly*, *35*(2), 355-362.

Coxhead, A., & Hirsh, D. (2007). A pilot science word list for EAP. *Revue Française de linguistique appliquée*, *XII* (2), 65-78.

Csomay, E. (2005). Linguistic variation within university classroom talk: A corpus-based perspective. *Linguistics and Education*, *15*, 243-274.

Cumming, A. (2009). Language assessment in education: Tests, curricula, and teaching. In B. Spolsky (Ed.), *Language policy and language assessment. Annual Review of Applied Linguistics*, *29*, 90-100.

Cumming, A., Grant, L., Mulcahy-Ernt, P. & Powers, D. (2004). *A teacher-verification study of prototype reading and speaking tasks for New TOEFL*. TOEFL Report MS-26. Princeton, NJ: Educational Testing Service.

Cumming, A., Kantor, R., & Powers, D. (2001). S*coring TOEFL essays and TOEFL 2000 prototype tasks: An investigation into raters' decision making and development of a preliminary analytic framework*. TOEFL Report MS-22. Princeton, NJ: Educational Testing Service.

Cumming, A., Kantor, R., & Powers, D. (2002). Decision making while scoring ESL/EFL compositions: A descriptive model. *Modern Language Journal, 86*, 67-96.

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper*. TOEFL Report MS-18. Princeton, NJ: Educational Testing Service.

Cutting, J. (1999). The grammar of the in-group code. *Applied Linguistics, 20,* 179-202.

Douglas, D. (1997). *Testing speaking ability in academic contexts: Theoretical considerations*. TOEFL Report MS-08. Princeton, NJ: Educational Testing Service.

Flowerdew, J., & Tauroza, S. (1995). The effect of discourse markers of second language lecture comprehension. *Studies in Second Language Acquisition, 17,* 435-458.

Fortanet, I. (2004). The use of "We" in university lectures: Reference and function. *English for Specific Purposes, 23,* 45-66.

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing, 9,* 123-145.

Halliday, M.A.K. (1988). On the language of physical science. In M. Ghadessy (Ed.) *Registers of written English,* (pp. 162-178). London: Pinter.

Halliday, M.A.K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.

Hamp-Lyons, L. (1991) Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.) *Assessing second language writing in academic contexts* (pp. 241-76). Norwood: Ablex.

Hamp-Lyons, L. (1995) Rating nonnative writing: The trouble with holistic scoring. *TESOL Quarterly 29,* 759–62.

Hamp-Lyons, L., & Kroll, B. (1996). *TOEFL 2000 ─ Writing: Composition, Community, and Assessment.* TOEFL Report MS-05. Princeton, NJ: Educational Testing Service.

Heatley, A., Nation, I.S.P., & Coxhead, A. (2002). Range and Frequency programs. Retrieved from: http://www.vuw.ac.nz/lals/staff/Paul_Nation (accessed on May 8, 2013).

Helt, M. E. (2001). A multi-dimensional comparison of British and American spoken English. In S. Conrad & D. Biber (Eds.), *Variation in English: Multidimensional studies,* (pp. 157-170). Harlow: Longman.

Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes, 13,* 239-256.

Hyland, K. (2002a). Directives: Argument and engagement in academic writing. *Applied Linguistics, 23*(2), 215-239.

Hyland, K. (2002b). "What do they mean?" Questions in academic writing. *Text, 22*(4), 529-557.

Kunnan, A. & Wu, J. (2010) General English Proficiency Test. In L. Chen & A. Curtis, (Eds.), *English language assessment and the Chinese learner* (pp. 158-174). USA: Routledge.

Kuo, G. (2005). A preliminary corpus study on EFL test takers' writing proficiency. *Proceedings of the Eighth International Conference on English Language Testing in Asia* (pp. 27-35). Hong Kong.

Lee, W., Kantor, R., & Mollaun, P. (2002). Score reliability as an essential prerequisite for validating new writing and speaking tasks for TOEFL. *Paper presented at the annual TESOL Convention,* Salt Lake City, UT.

Lindemann, S., & Mauranen, A. (2001). "It's just really messy": The occurrence and function of just in a corpus of academic speech. *English for Specific Purposes, 20,* 459-475.

Language Training and Testing Center (LTTC). (2003). *Concurrent validity studies of the GEPT Intermediate level, GEPT High-Intermediate level, CBT TOEFL, CET-6, and the English test of the R.O.C. College Entrance Examination.* From http://www.lttc.ntu.edu.tw/academics/thesis_gept.htm (accessed on 16 November, 2011)

Ma, M., & Li, S. (2009). Bridging test construct and beneficial washback effects: Revising the GEPT high-intermediate reading test. *Paper presented at the 26th International*

Conference of English Teaching and Learning in the R. O. C., National Tsinghua University, Hsinchu.

Marley, C. (2002). Popping the question: Questions and modality in written dating advertisements. *Discourse Studies*, *4*(1), 75-98.

Mauranen, A. (2003) "A Good Question." Expressing evaluation in academic speech. In G. Cortese & P. Riley (Eds.), *Domain-specific English: Textual practices across communities and classrooms* (pp. 115-140). New York: Peter Lang.

Mauranen, A. (2004) "They're a little bit different": Variation in hedging in academic speech. In K. Aijmer & A. B. Stenström (Eds.), *Discourse patterns in spoken and written corpora* (pp. 173-197). Amsterdam: John Benjamins.

Mauranen, A., & Bondi, M. (2003). Evaluative language use in academic discourse. *Journal of English for Academic Purposes*, *2*, 269-271.

McCrostie J. (2006). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal*, *32*, 97-114.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London & New York: Routledge.

McNamara, T. (2002) Discourse and assessment. *Annual Review of Applied Linguistics, 22,* 221-242.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.) *Learner English on computer* (pp. 107-118). London & New York: Longman.

Powell, C., & Simpson, R. (2001). Collaboration between corpus linguistics and digital libraries for the MICASE web interface. In R. Simpson & J. Swales (Eds.), *Corpus linguistics in North America*, (pp. 32-47). Ann Arbor: University of Michigan Press.

Reppen, R. (2001). Register variation in student and adult speech and writing. In S. Conrad & D. Biber (Eds.), *Variation in English: Multidimensional studies*, (pp. 187-199). Harlow: Longman.

Reppen, R., & Vásquez, C. (2007). Using corpus linguistics to investigate the language of teacher training. In J. Walinski, K. Kredens & S. Gozdz-Roszkowski (Eds.), *Corpora and ICT in language studies* (pp. 13-29). Frankfurt: Peter Lang.

Roever, C., & Pan, Y. (2008). Test review: GEPT: General English Proficiency Test. *Language Testing 25*(3), 403-18.

Rosenfeld, M., Leung, S., & Oltman, P. K. (2001). The reading, writing listening, and speaking tasks important for academic success at the undergraduate and graduate levels. TOEFL Report MS-21. Princeton, NJ: Educational Testing Service.

Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Snellings, P., Simis, A., & Stevenson, M. (2002). Linguistic knowledge, metacognitive knowledge, and retrieval speed in L1, L2 and EFL writing: A structural equation modeling approach. In S. Ransdell & M. L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 101-122). Dordrecht: Kluwer Academic.

Shih, C. (2008) The General English Proficiency Test. *Language Assessment Quarterly 5*(1): 63-76.

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulae List (AFL). *Applied Linguistics, 31,* 487-512.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J. M., & Burke, A. (2003). "It's really fascinating work": Differences in evaluative adjectives across academic registers. In P. Leistyna & C. F. Meyer (Eds), *Corpus analysis: Language structure and language use* (pp. 1-18). New York: Rodopi.

Thompson, G., & Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers.

Applied Linguistics, *12*, 365-382.

Turner, C. E. (2000). Listening to the voices of rating scale developers: Identifying salient features for second language performance assessment. *The Canadian Modern Language Review 56*(4), 555-84.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii Press.

Wu, R. (2003). Assessing Advanced-level English writing: A report on the case of the GEPT. *Proceedings of the Sixth International Conference on English Language Testing in Asia* (pp. 75-101), Seoul.

Wu, J. (2007). English language assessment in Taiwan: Where do we go from here? *Proceedings of 2007 International Conference and Workshop on TEFL & Applied Linguistics* (pp. 574-86), Taipei.

Wu, J. (2009). Differential item functioning in gender and living background groups in the GEPT. *Paper presented at the Proceedings of the 13th International Conference on Language Education*, National Kaohsiung Normal University, Kaohsiung.

Wu, J. (2010a). Validating GEPT scores against teacher and student assessments. *Paper presented at the International Conference on Applied Linguistics & Language Teaching* (ALLT), Taipei.

Wu, J. (2010b). East meets west: The adoption of the CEFR in Taiwan. *Paper presented at the International Forum on English Language Policies and Practices in Asia*, Taipei.

Wu, J., & Chao, I. (2009). Revision of the GEPT-Intermediate writing rating scales. *Proceedings of 2009 International Conference and Workshop on TEFL & Applied Linguistics* (pp. 516-525), Taipei.

Wu, J., & Lin, A. (2008). Assessing English proficiency at advanced level: Testers' feedback to teaching. *Paper presented at the 25th International Conference of English Teaching and Learning in R. O. C.*, National Chung Cheng University, Chai-yi.

Wu, J., & Ma, T. (2013). Investigating rating processes in an EAP writing test: Insights into scoring validity. *Paper presented at the 35th Annual Language Testing Research Colloquium*, Seoul.

Wu, R., & Chin, J. (2006). An impact study of the Intermediate Level GEPT. *Proceedings of the Ninth International Conference on English Language Testing in Asia* (pp. 41-65), Taipei.

Wu, R., & Liao, C. (2010). Establishing a common score scale for the GEPT Elementary, Intermediate, and High-Intermediate Level listening and reading tests. In T. Kao & Y. Lin (Eds.), *A New Look at Language Teaching and Testing: English as Subject and Vehicle – Selected Papers from the 2009 LTTC International Conference on English Language Teaching and Testing*, pp. 309-29. Taipei: Language Training and Testing Center.

Xiao, R. (2009). Multidimensional analysis and the study of world Englishes. *World Englishes*, *28*(4), 421-450.

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