IELTS Research Reports Online Series
ISSN 2201-2982
Reference: 2013/2

# An investigation of the relations between test-takers' first language and the discourse of written performance on the IELTS Academic Writing Test, Task 2

***Author:*** *A. Mehdi Riazi and John S. Knox, Macquarie University, Australia*

***Grant awarded:*** *Round 16, 2010*

*Keywords:* *IELTS Academic Writing, test validation, computational text analysis, systemic functional linguistics, genre, appraisal theory*

## Abstract

This project examines the responses of IELTS candidates to Task 2 of the Academic Writing Test, exploring the relations between candidates' first language, their band score, and the language features of their texts. The findings show that candidates' first language is one of several factors related to the band score they achieve.

The scripts came from candidates representing three L1 groups (Arabic L1, Hindi L1, and European-based L1) and three band scores (band 5, 6, and 7). Quantitative analysis was conducted on 254 scripts, measuring text length, readability of the scripts, Word Frequency Level (WFL), lexical diversity, grammatical complexity, incidence of all connectives, and two measures of coreferentiality (argument and stem overlap).

Discourse analysis was conducted on a subset of 54 texts, using genre analysis and Appraisal Theory from Systemic Functional Linguistics.

Descriptive statistics of textual features indicate that, overall, scripts with higher band scores (6 and 7) were found to be more complex (using less frequent words, greater lexical diversity, and more syntactic complexity) than cohesive. Significant differences were also found between the three L1 categories at the same band scores. These included: readability at band 7 between European-based L1 and Hindi L1 scripts; lexical diversity at band scores 5 and 6 between European-based L1 and Hindi L1 scripts; word frequency at band 7 between Hindi L1 and European-based L1 scripts; cohesion at band 6 between Arabic L1 and European-based L1 scripts; and cohesion also at band 7 between Hindi L1 and Arabic L1 scripts.

Some differences were also found in the discourse analysis, with scripts of European-based L1 candidates more likely to use a typical generic structure in higher bands, and the scripts of Hindi L1 candidates showing slightly different discursive patterns in Appraisal from the other two groups.

A range of measures (quantitative and discourse analytic) did not show any difference according to L1. The measures found to be good indicators of band score regardless of candidate L1 were text length, reading ease and word frequency in the quantitative analysis, and genre and use of Attitude in the discourse analysis.

There were also several unexpected findings, and research is recommended in areas including the input of scripts (handwriting versus typed), the relations between task and genre, and the 'management of voices' in candidate responses in relation to academic writing more generally.

## Publishing details

**Published by IDP: IELTS Australia © 2013.**
This online series succeeds *IELTS Research Reports Volumes 1–13*, published 1998–2012 in print and on CD.

**Web: www.ielts.org**

## AUTHOR BIODATA


### A. Mehdi Riazi

Associate Professor Mehdi Riazi is the convenor of the postgraduate units of language assessment and research methods in the Department of Linguistics, Macquarie University. He is currently supervising eight PhD students and one Master's student. One PhD thesis on test validity and five Master theses have been completed under his supervision at Macquarie University.

Before joining Macquarie University, he taught Master and Doctoral courses at Shiraz University, Iran, where he supervised 14 PhD and approximately 40 Master dissertations on issues related to ESL teaching and learning. Four of the PhD dissertations and a relatively large number of the Master theses were related to language testing and assessment (including one on Iranian IELTS candidates' attitudes to the IELTS Test – see Rasti 2009).

Associate Professor Riazi was also team leader of the project which developed the Shiraz University Language Proficiency Test (SULPT). He was the centre administrator for the TOEFL–iBT at Shiraz University for two years (2007–2009). He has published and presented papers in journals and conferences on different issues and topics related to ESL pedagogy and assessment.

### John Knox

Dr John Knox is a Lecturer in the Department of Linguistics, Macquarie University, Australia. He has published in the areas of language assessment, language pedagogy, language teacher education, systemic functional linguistics, and multimodality.

He has been an IELTS Examiner (1997–2006), an IELTS item writer (2001–2006), an UCLES main suite Oral Examiner (1995–1999), and an UCLES Oral Examiner Trainer Coordinator (1999–2000).

Dr Knox has also been a consultant to the Australian Adult Migrant English Program's (AMEP) National Assessment Task Bank project (2003–2006, 2013), and a consultant to the AMEP Citizenship Course Project as an item writer for the Australian Citizenship Test, (December 2005–January 2006).

## INTRODUCTION FROM IELTS

This study by Mehdi Riazi and John Knox from Macquarie University was conducted with support from the IELTS partners (British Council, IDP: IELTS Australia, and Cambridge English Language Assessment) as part of the IELTS joint-funded research program. Research funded by the British Council and IDP: IELTS Australia under this program complement those conducted and commissioned by Cambridge English Language Assessment, and together inform the ongoing validation and improvement of IELTS.

A significant body of research has been produced since the program began in 1995 – over 90 empirical studies have received grant funding. After undergoing a process of peer review and revision, many of the studies have been published in academic journals, in several IELTS-focused volumes in the Studies in Language Testing series (http://research.cambridgeesol.org/research-collaboration/silt), and in *IELTS Research Reports*, of which 13 volumes have been produced to date.

The IELTS partners recognise that there have been changes in the way people access research. Since 2011, IELTS Research Reports have been available to download free of charge from the IELTS website, www.ielts.org. However, collecting a volume's worth of research takes time. Thus, individual reports are now made available on the website as soon as they are ready.

This report looked at IELTS Academic Task 2, using multiple methods to look for similarities and differences in performances across a range of band scores and first language backgrounds. In terms of aims and methods, it is most similar to Mayor, Hewings, North & Swann (2007), but looking at candidates from different L1 backgrounds and who had obtained different band scores. Both reports contribute to research conducted or supported by the IELTS partners on the nature of good writing and the description thereof (e.g. Banerjee, Franceschina & Smith, 2007; Hawkey & Barker, 2004; Kennedy & Thorp, 2007).

Riazi and Knox replicate many of the previous studies' outcomes, finding for example that more highly rated scripts use less common lexis, evidence greater complexity, employ fewer explicit cohesive devices, and show expected genre features, among others. Apart from providing support for the ability of IELTS to discriminate between writing of different quality therefore, this replication across studies across different data samples provides evidence for the consistency with which IELTS has been marked over the years.

It is also interesting to note that, in the literature reviewed in this report, the same features as above are generally the same ones which distinguish texts produced by language learners and English L1 in various testing and non-testing contexts, including writing in the university setting. That is to say, for all the limitations imposed by the testing context on what can or cannot be elicited, IELTS is able to discriminate candidates on many of the same aspects as in the target language use domain.

Methodologically, the quantitative analysis was aided by the use of Coh-Metrix, a relatively new automated tool capable of producing more indices of text quality, which is already being used and will continue to help researchers in the coming years. Nevertheless, as the authors acknowledge, these indices do not capture all the features described in the IELTS Writing band descriptors, and thus only captures in part what trained examiners are able to do in whole.

The limits of automated analysis provide the *raison d'etre* for the qualitative analysis in the research, which will also continue to be important for researchers to do so as to provide a more complete and triangulated picture of what is being investigated. Resource limitations unfortunately prevented greater overlap and comparison between the quantitative and qualitative components of the study, and represent an obvious direction for future studies in this area to take.

Indeed, as new tools produce more indices and new frameworks point out more features, the greater challenge will be to determine what each measure is able to tell us and not tell us, and how these measures combine and interact with one another to reliably identify examples of good writing. This research points us in the right direction.

**Dr Gad S Lim**
**Principal Research and Validation Manager**
**Cambridge English Language Assessment**

## References to the IELTS Introduction

Banerjee, J, Franceschina, F, and Smith, AM, 2007, 'Documenting features of written language production typical at different IELTS band score levels' in *IELTS Research Reports Volume 7*, IELTS Australia, Canberra and British Council, London, pp 241-309.

Hawkey, R, and Barker, F, 2004, 'Developing a common scale for the assessment of writing' in *Assessing Writing*, *9*(3), pp 122-159.

Kennedy, C, and Thorp, D, 2007, 'A corpus-based investigation of linguistic responses to an IELTS Academic Writing task' in L Taylor and P Falvey (Eds), *IELTS Collected Papers: Research in speaking and writing assessment*, Cambridge ESOL/Cambridge University Press, Cambridge, pp 316-377.

Mayor, B, Hewings, A, North, S, and Swann, J, 2007, 'A linguistic analysis of Chinese and Greek L1 scripts for IELTS Academic Writing Task 2' in L Taylor and P Falvey (Eds), *IELTS Collected Papers: Research in speaking and writing assessment*, Cambridge ESOL/Cambridge University Press, Cambridge, pp 250-313.

# TABLE OF CONTENTS

## List of tables

## List of figures

## GLOSSARY

| | |
|---|---|
| **Affect (within Appraisal theory** | Affect deals with the expression of human emotion (Martin and White 2005, pp 61ff) |
| **Appraisal theory** | Appraisal theory deals with "the interpersonal in language, ... the subjective presence of writers/speakers in texts as they adopt stances towards both the material they present and those with whom they communicate" (Martin and White 2005, p 1). It has three basic categories: Attitude, Engagement, and Graduation |
| **Appreciation (within Appraisal theory)** | Appreciation deals with "meanings construing our evaluations of 'things', especially things we make and performances we give, but also including natural phenomena" (Martin and White 2005, p 56) |
| **Attitude (within Appraisal theory)** | Attitude is concerned with "three semantic regions covering what is traditionally referred to as emotion, ethics and aesthetics" (Martin and White 2005, p 42). Emotions are dealt with in the sub-system entitled Affect; ethics in the sub-system entitled Engagement, aesthetics in the sub-system entitled Appreciation |
| **CC** | Coherence and Cohesion |
| **Coh-Metrix** | Software that analyses written texts on multiple measures of language and discourse that range from words to discourse genres |
| **Coreferentiality** | Stem overlap and argument overlap |
| **CTA** | Computational Text Analysis |
| **Engagement (within Appraisal theory)** | Engagement is concerned with "the linguistic resources by which speakers/writers adopt a stance towards the value positions being referenced by the text and with respect to those they address" (Martin and White 2005, p 92). The two primary sub-divisions in Engagement are Monogloss and Heterogloss |
| **FRE** | Flesch Reading Ease |
| **GRA** | Grammatical Range and Accuracy |
| **Heterogloss (within Appraisal theory)** | Any expression which recognises that the position stated is not the only possible one, including devices such as reporting verbs, modality, and negation |
| **Judgement (within Appraisal theory)** | Judgement deals with meanings around the evaluation of human behaviour, and whether it is esteemed or sanctioned behaviour. Broadly, it is about the semantic regions of 'right and wrong' |
| **LR** | Lexical Resource |
| **Monogloss (within Appraisal theory)** | 'Bare assertions' that do not overtly recognise the possibility of alternate positions to the one expressed |
| **SFL** | Systemic Functional Linguistics |
| **TR** | Task Response |
| **TTR** | Type/Token ratio |
| **WF** | Word Frequency |

# 1 INTRODUCTION

## 1.1 Context and rationale

Higher education has become increasingly internationalised over the last two decades. Central to this process has been the global spread of English (Graddol 2006). As students enter English-medium higher education programs, they must participate in the discourse of the disciplinary community within which their program of study is located. Increasingly, such disciplinary discourses are understood as involving distinct discursive practices, yet the fact remains that there are discursive demands in academic English which are shared by the different disciplinary communities as part of the broader discourse community of academia (Hyland 2006).

Tests like the IELTS Academic Writing Test aim to assess the extent to which prospective tertiary students, who come from anywhere in the world and who speak any variety of English, are able to participate in the written activities of the broad discourse community of English-language academia, regardless of individual and social variables. In the case of IELTS, the approach taken to achieve this aim is direct testing of candidates' writing ability by assessing their performance on two writing tasks.

As Taylor (2004) contends, the inclusion of direct tests of writing in high-stakes and large-scale English-language proficiency tests reflects the growing interest in communicative language ability and the importance of performance-based assessment. The strong argument for performance-based assessment (writing and speaking sections) in tests such as IELTS is that, if we want to know how well somebody can write or speak, it seems natural to ask them to do so and to evaluate their performance. The directness of the interpretation makes many competing interpretations (e.g., in terms of method effects) less plausible (Kane, Crooks and Cohen 1999).

Another positive aspect of performance-based testing is the effect this approach has on teaching and learning the language, or the positive washback effect (Bachman 1990; Bachman and Palmer 1996; Hughes 2003). A positive washback effect promotes ESL/EFL curricula (instructional materials, teaching methods, and assessment) that foster oral and written communication abilities in students. Other benefits of using performance-based assessment can be found in Brown (2004, p 109). However, the mere appearance of fidelity or authenticity does not necessarily imply that a proposed interpretation is valid (Messick 1994 cited in Kane et al. 1999). The interpretation of the test scores, especially when it comes to proficiency levels and test-takers' characteristics, needs to be considered more carefully to ensure the validity of test score interpretations.

This report details research into candidate responses to Task 2 of the IELTS Academic Writing Test, in the hope of contributing to a greater understanding of the validity of this test, and its contribution to the overall social aims of the IELTS Test in the context of higher education and internationalisation.

## 1.2 Design

The research reported here is broadly conceptualised within a test validation framework, and intends to contribute to ongoing validation studies of the IELTS Academic Writing Test, with a focus on Task 2 as stated above. Two variables are addressed in the study:

1. three band scores (5, 6, and 7) on the IELTS Academic Writing Test
2. three test-taker first languages (L1s) (Arabic, Hindi, and European-based L1).

The reason for choosing the three language groups is that, based on IELTS Test-taker Performance 2009 (IELTS 2010), Dutch and German L1 candidates obtained the highest mean score on the IELTS Academic Writing Module (6.79 and 6.61 respectively), Arabic L1 candidates the lowest (4.89), and Hindi L1 candidates an intermediate mean score (5.67). In sourcing candidate responses to the IELTS Academic Writing Test, there were not sufficient numbers of German and Dutch scripts, so the 'European-based L1' group was expanded to include scripts from Portuguese L1 (mean score: 6.11) and Romanian L1 (mean: 6.31) candidates. These 'European-based L1' scripts were treated as a single group.

We stress that, as a result of the issues in data collection stated above, the grouping of different languages under the 'European-based L1' label is based on the mean performance of candidates on IELTS Task 2, and is not based on linguistic similarity or language family. In all cases, candidates' L1 is identified by the candidates' self-reporting to IELTS, and IELTS' subsequent reporting to the researchers. Potential issues with the operationalisation of L1 in this study are discussed in Section 4.2, below.

## 1.3 Aims of the study

This research project has three aims. The first aim is to identify probable systematic differences between scripts assessed at different band levels (namely, 5, 6, and 7). What linguistic features do band 5 scripts have in common, band 6 scripts, and band 7 scripts? What systematic differences are there in linguistic features of scripts between the different bands?

The second aim is to investigate the impact of test-takers' L1 on the linguistic features of scripts assessed at the same band level. Do the scripts of candidates with the same band score, but different L1s, display any systematic linguistic variation?

The third aim is to explore the interaction between band score and test-takers' L1, and whether the impact of test-takers' L1 (if any) differs in degree and/or kind at different band scores. Does test-takers' L1 have a different impact at different band scores?

Are scripts at some band levels linguistically more homogenous across L1 groups than scripts at others?

This presents us with a matrix for comparison with nine 'blocks' of scripts as shown in Table 1.1.

As Taylor (2004, p 2) argues, "Analysis of actual samples of writing performance has always been instrumental in helping us to understand more about key features of writing ability across different proficiency levels and within different domains". Accordingly, this project focuses on the linguistic features of the test-takers' scripts, using both computer-based quantitative analyses of the lexico-syntactic features of the scripts as employed in Computational Text Analysis (CTA), and detailed discourse analysis of genre and Appraisal from Systemic Functional Linguistics (SFL).

The impact of Computational Text Analysis (CTA) within applied linguistics research is well known (Cobb 2010). CTA provides a relatively accurate and objective analysis of text features, which can be used to compare texts, and to relate them to other features of interest such as level of proficiency, and test-takers' L1. The textual features included in the analysis, and the computer program used to perform these analyses are explained in Section 2.

Systemic Functional Linguistics (SFL) is a social theory of language which takes the text as its basic unit of study. In SFL, meaning is made at different levels: the whole text, stretches of discourse 'above the clause', clause level grammar and lexis. SFL has made a significant contribution to the theory and practice of language education (e.g. Christie and Derewianka 2008; Christie and Martin 1997; Halliday and Martin 1993; Hood 2010; McCabe et al. 2007; Ravelli and Ellis 2004) and language assessment (e.g. Coffin 2004a; Coffin and Hewings 2005; Huang and Mohan 2009; Leung and Mohan 2004; Mohan and Slater 2004; Perrett 1997).

Two of the most widely recognised contributions of SFL to language education are genre theory (e.g. Martin and Rose 2008) and Appraisal theory (e.g. Martin and White 2005). The current study reports on analysis of these two 'levels' of language, both of which are grounded in a lexicogrammatical analysis of a subset of the total scripts collected, consisting of six texts from each 'block' (see Table 1.1), or 54 texts in total.

As noted, the aim was to collect 270 scripts from the IELTS Academic Writing Test, Task 2 (30 scripts from each of the nine 'blocks' identified in Table 1.1). Ideally, all scripts would have come from a single task, but this was not possible, and the scripts responded to 26 different tasks (see Table 3.2). Thirty scripts were collected for most blocks, but not all. In total, 254 texts were analysed using CTA (see Section 2), and 54 texts were analysed using SFL as planned (see Section 3).

All scripts were transcribed from handwriting into word-processing software. This aspect of the research was surprisingly challenging, and the researchers had to work much more closely with the secretarial assistants than anticipated on this stage of the research process. Decisions constantly had to be made related to:

- punctuation (e.g. was a mark intended as a comma, a full-stop, or had the pencil simply been rested on the page?)
- capitalisation (some candidates wrote scripts completely in capitals; some always capitalised particular letters (e.g. "r") – even in the middle of words; some 'fudged' the capitalisation of proper nouns so it was unclear whether a word was capitalised or not)
- paragraphing (paragraph breaks were not always indicated by line breaks)
- legibility (some candidates had idiosyncratic ways of writing particular letters, some candidates simply had very bad handwriting).

While many of these decisions were relatively minor, others had ramifications for grammatical and discursive understanding of the scripts. Handwriting was not the focus of the research, but it became clear that many candidates used the 'flexibility' of handwriting to their advantage, in a way that would not be acceptable in submitting academic assignments (which are now usually required to be submitted typed in most English-medium universities).

| First Language / Band score | Arabic | Hindi | European-based |
|---|---|---|---|
| 7 | 'Block A'<br>30 scripts (Task 2) | 'Block D'<br>30 scripts (Task 2) | 'Block G'<br>30 scripts (Task 2) |
| 6 | 'Block B'<br>30 scripts (Task 2) | 'Block E'<br>30 scripts (Task 2) | 'Block H'<br>30 scripts (Task 2) |
| 5 | 'Block C'<br>30 scripts (Task 2) | 'Block F'<br>30 scripts (Task 2) | 'Block I'<br>30 scripts (Task 2) |

***Table 1.1: Matrix of comparison: L1 and assessed writing band score***

The issues with handwritten scripts were foregrounded due to the need to transcribe the scripts, and this made visible potential issues in scoring and reliability that may not always be apparent in rating, and even in rater training and moderation (cf. Weigle 2002, pp 104–6). The issue of handwriting versus computer entry is taken up again in Section 4 from a different perspective. Once the scripts were transcribed, they were subjected to Computational Text Analysis and Systemic Functional Linguistic discourse analysis.

## 1.4 Previous research

The impact of a number of variables on candidates' performance on the IELTS Academic Writing Test has been studied, including background discipline (Celestine and Su Ming 1999), task design (O'Loughlin and Wigglesworth 2003), and memorisation (Wray and Pegg 2005).

Other variables, more directly relevant to the current study, have also been researched. Mayor, Hewings, North, Swann and Coffin's (2007) study examined the errors, complexity (t-units with dependent clauses), and discourse (simple and complex themes, interpersonal pronominal reference, argument structures) of Academic Writing Task 2 scripts of candidates with Chinese and Greek as their first language (see also Coffin 2004; Coffin and Hewings 2005).

Mayor et al. analysed 186 Task 2 scripts of high- (n=86) vs. low-scoring (n=100) Chinese (n=90) and Greek (n=96) L1 candidates. Scores at band 7 and 8 were considered high scores, and those at band 5 as low scores. Their analysis of the scripts included both quantitative (error analysis of spelling, punctuation, grammar, lexis, and prepositions; independent and dependent clauses using t-unit) and qualitative (sentence structure argument using theme and rheme, and tenor and interpersonal reference). They found that high and low-scoring scripts were differentiated by a range of features and that IELTS raters seemed to attend to test-takers' scripts more holistically than analytically. Generally, however, they stated text length, low formal error rate, sentence complexity, and occasional use of the impersonal pronoun "one" were the strongest predictors of high scored scripts.

In addition to the formal features, Mayor et al. found some functional features of the scripts (thematic structure, argument genre, and interpersonal tenor) to positively correlate with task scores. They also found that the nature of Task 2 prompts (e.g. write for "an educated reader") may have cued test-takers to adopt a "heavily interpersonal and relatively polemical" style (p 250).

As for the influence of candidates' L1, Mayor et al. found that the two different L1 groups made different kinds of errors in low-scoring scripts. Chinese L1 candidates were found to have "made significantly more grammatical errors than Greek L1 at the same level of performance" (p 251). Little difference was found between Chinese and Greek test-takers in terms of argument structure in their performance for expository over discussion argument genres. As for argument genres, Greek candidates were found to strongly favour hortatory, while Chinese showed a slight preference for formal analytic styles.

The current project differs from that of Mayor et al. in three important ways. First, instead of examining high- and low-scoring scripts (band 5, and bands 7–8 respectively), scripts from three specific band scores are studied. Second, the three L1 groups in the current study are distinct from those in Mayor et al.'s study. Third, quantitative measures of a range of features not examined by Mayor et al. are included. At the same time, there are obvious similarities in the two studies. Both Mayor et al.'s study and the current study employ quantitative analysis and systemic functional analysis (particularly genre analysis and interpersonal analysis) of Academic Writing Task 2 scripts. Thus, the current study builds on the knowledge about features of Task 2 scripts across different L1 groups, expanding the research base in this area from Chinese and Greek L1 groups (Mayor et al. 2007) to include Arabic, Hindi, and European-based L1 groups.

Banerjee, Franceschina and Smith (2007) analysed scripts from Chinese and Spanish L1 candidates on Academic Task 1 and 2, from bands 3 to 8. They examined such aspects as cohesive devices (measured by the number and frequency of use of demonstratives), vocabulary richness (measured by type-token ratio, lexical density, and lexical sophistication), syntactic complexity (measured by the number of clauses per t-unit as well as the ratio of dependent clauses to the number of clauses), and grammatical accuracy (measured by the number of demonstratives, copula in the present and past tense and subject-verb agreement). They found that assessed band level, L1, and task could account for differences on some of these measures. But in contrast to the current study, Banerjee et al. did not include discourse analysis to complement their quantitative analysis.

Banerjee et al. suggest that all except the syntactic complexity measures were informative of increasing proficiency level. Scripts rated at higher bands showed an index of higher type-token ratio, and lexical density, and lexical sophistication (low frequency words). They also found that L1 and writing tasks had critical effects on some of the measures, and so they suggested further research on these aspects.

The current study responds to this and similar suggestions by concentrating on three band score levels and three L1 backgrounds, and by analysing the scripts both quantitatively and qualitatively, including discourse analysis.

In the research published to date, a range of variables affecting candidate performance on the IELTS Writing Test (including the variables of task, L1, and proficiency as indicated by band score) have been studied, and both quantitative and discourse-analytic methods have been used in such studies. However, to date, no study of the IELTS Writing Test has compared three L1 groups, and none has combined the specific combination of quantitative and discourse-analytic methods as is done in this current study.

## 1.5 Research questions

The three research questions underpinning this study are as follows.

*Research Question 1*: What systematic differences are there in the linguistic features of scripts produced for IELTS Academic Writing Task 2 at bands 5, 6 and 7?

*Research Question 2*: What systematic differences are there (if any) in the linguistic features of the scripts produced for IELTS Academic Writing Task 2 for European-based, Hindi, and Arabic L1 backgrounds?

*Research Question 3*: To what extent does the impact of L1 on the linguistic features of the scripts differ at different band levels?

The following section reports on the Computational Text Analysis of the scripts. Section 3 reports on the systemic functional analysis of genre and Appraisal. Section 4 presents the conclusions and recommendations; and acknowledgements are given before the list of references.

# 2 QUANTITATIVE ANALYSIS OF SCRIPTS

To answer the research questions of the project, the Coh-Metrix program (McNamara, Louwerse, McCarthy, and Graesser 2010; Graesser, McNamara and Kulikowich 2011) was used to analyse scripts. Coh-Metrix is software that analyses written texts on multiple measures of language and discourse that range from words to discourse genres (Graesser, McNamara and Kulikowich 2011). As Crossley and McNamara (2010) contend, in recent years, researchers in the area of L2 writing have used computational text analysis tools like Coh-Metrix to investigate more sophisticated linguistic indices in second language writers' texts. Accordingly, Coh-Metrix was used to analyse chosen linguistic features of IELTS Writing Task 2 scripts produced by the three L1 groups as they pertain to the three research questions.

## 2.1 Textual features included in the analysis of scripts

The quantitative analyses of textual features of scripts in this project include text length (number of words), readability (Flesch Reading Ease) of the scripts, word frequency (WF), lexical diversity (LD) represented by type/token ratio (TTR), index of all connectives, coreferentiality (stem and argument overlap), and syntactic complexity (number of words before the main verb). The selection of these linguistic features for the analysis of IELTS Academic Task 2 scripts is theoretically based on other empirical studies as we discuss in Sections 1.4 and 2.2, and is practically based on the fact that the scoring system of IELTS Academic uses criteria that overlap with these measures to assess Task 2 of writing section (IELTS 2009, p 2).

The IELTS criteria are:

- Task Response
- Coherence and Cohesion
- Lexical Resource
- Grammatical Range and Accuracy.

The Task Response criterion is not included in the quantitative analysis because there is no corresponding quantitative measure for it, but it is dealt with in the qualitative analysis section of this report. We have used coreferentiality (stem and argument overlap) and index of all connectives to represent Cohesion and, indirectly, Coherence. Word frequency and lexical diversity indices represent Lexical Resource, and syntactic complexity represents Grammatical Range.

Important as the relations are between the measures used in this study and the IELTS grading criteria, it should be noted that the selection of these indices from the Coh-Metrix program do not fully and exactly correspond to the rating criteria used to assess Task 2 in the IELTS Writing Test. Our purpose is to identify the linguistic characteristics of written texts at each of the three band levels (5, 6 and 7), and of each of the three L1 groups at each band level. It is not our purpose to provide an analytical perfect match to the IELTS criteria.

Discussion of genre and Appraisal analysis is presented in Section 3. More information on the other linguistic features and their measures is presented in Sections 2.2 and 2.3. The next section reviews related literature that provides the theoretical context and support for:

- using Coh-Metrix as the textual analysis tool
- using the selected linguistic features in the analysis of the IELTS Academic Writing Task 2.

## 2.2 Literature review

Coh-Metrix has been used extensively to analyse texts from reading-comprehension and writing perspectives. Readers are recommended to see Crossley and McNamara (2009) for a comprehensive overview of how Coh-Metrix linguistic indices are validated. Here, we present a number of recent studies which have used Coh-Metrix to analyse the linguistic features of written texts, and particularly texts written by L2 writers.

Table 2.1 presents a number of studies in which Coh-Metrix has been used to analyse written text features.

| Author and date | Main research focus | Data | Results |
|---|---|---|---|
| McNamara, Louwerse, McCarthy & Graesser (2010) | To check the validity of Coh-Metrix as a measure of cohesion in text using stimuli from published discourse psychology studies as a benchmark | 19 samples of pairs of texts with high- versus low-cohesion versions from 12 published experimental studies | Results showed that Coh-Metrix indices of cohesion (individually and combined) significantly distinguished the high- versus low-cohesion versions of these texts.<br>The five unique variables that captured the differences between the high- and low-cohesion texts included coreferential noun overlap, LSA sentence to sentence, causal ratio, word concreteness, and word frequency. Of these variables, the coreference, LSA, and causal ratio measures are more likely, in terms of face validity, to be considered direct indices of cohesion, whereas word concreteness and word frequency are indices likely related to the side effects of manipulating cohesion. |
| Crossley & McNamara (2010) | To investigate if higher-rated essays contain more cohesive devices than lower-rated essays, and if more proficient writers demonstrate greater linguistic sophistication than lower-proficiency writers, especially in relation to lexical difficulty | Essays written by graduating Hong Kong high school students for the Hong Kong Advanced Level Examination (HKALE). Essays with text lengths between 485 and 555 words were used | Results showed that five variables (lexical diversity, word frequency, word meaningfulness, aspect repetition and word familiarity) significantly predict L2 writing proficiency. Moreover, the results indicated that highly proficient L2 writers did not produce essays that were more cohesive, but instead produced texts that were more linguistically sophisticated. |
| McNamara, Crossley & McCarthy (2010) | To examine linguistic differences related to cohesion and linguistic sophistication between high- and low-proficiency writers, as indicated by their score on an essay | 120 essays from Mississippi<br>State University MSU corpus rated by five writing tutors with at least one year's experience | The three most predictive indices of essay quality were found to be syntactic complexity (as measured by number of words before the main verb), lexical diversity (as measured by the Measure of Textual Lexical Diversity), and word frequency (as measured by Celex, logarithm for all words). |
| Crossley, Salsbury, McNamara & Jarvis (2011) | To examine the potential for lexical indices to predict human evaluations of lexical proficiency based on lexical features related to vocabulary size, depth of knowledge, and access to core lexical items | 60 texts each from beginning, intermediate, and advanced second language (L2) adult English learners. The texts were collected longitudinally from 10 English learners. In addition, 60 texts from native English speakers were collected | Lexical diversity, word hypernymy values and content word frequency explained 44% of the variance of the human evaluations of lexical proficiency in the examined writing samples. The findings represent an important step in the development of a model of lexical proficiency that incorporates both vocabulary size and depth of lexical knowledge features. |
| Crossley, Salsbury & McNamara (2011) | To investigate how second language (L2) texts written by learners at various proficiency levels can be classified using computational indices that characterise lexical competence | 100 writing samples taken from 100 L2 learners | The strongest predictors of an individual's proficiency level were word imageability, word frequency, lexical diversity, and word familiarity. In total, the indices correctly classified 70% of the texts. |
| Crossley & McNamara (2011) | To investigate intergroup homogeneity within high intermediate and advanced L2 writers of English from Czech, Finnish, German, and Spanish first language backgrounds | Texts written by native speakers of English as baseline and essays written by writers from a variety of L1 language backgrounds | The results provided evidence for intergroup homogeneity in the linguistic patterns of L2 writers in that four word-based indices (hypernymy, polysemy, lexical diversity, and stem overlap) demonstrated similar patterns of occurrence in the sample of L2 writers. Significant differences were found for these indices between L1 and L2 writers. It is concluded that some aspects of L2 writing may not be cultural or independent, but rather based on the amount and type of linguistic knowledge available to L2 learners as a result of language experience and learner proficiency level. |

| | | | |
|---|---|---|---|
| Crossley, Weston, McLain Sullivan & McNamara (2011) | To find out if there are any significant differences between the linguistic features [word frequency, word concreteness, syntactic complexity, and cohesion (i.e., word overlap, incidence of connectives)] produced in essays by Grade 9 and Grade 11 students, and college freshmen | Essays produced by Grade 9 and Grade 11 students, and college freshmen | The results indicated that these writers produced more sophisticated words and more complex sentence structures as grade level increases. In contrast, the findings showed these writers produced fewer cohesive features in text as a function of grade level. The authors contend that linguistic development occurs in the later stages of writing development and that this development is primarily related to producing texts that are less cohesive and more elaborate. |

***Table 2.1: Text analysis studies with Coh-Metrix***

The following points can be highlighted from the studies included in the above table.

1. Coh-Metrix indices of cohesion (individually and combined) significantly distinguished the high- versus low-cohesion versions of published texts. The main indices were the coreference, LSA and causal ratio measures.
2. L2 writing proficiency could be significantly predicted by five Coh-Metrix variables (lexical diversity, word frequency, word meaningfulness, aspect repetition and word familiarity).
3. The three most predictive indices of essay quality were found to be syntactic complexity (as measured by number of words before the main verb), lexical diversity, and word frequency (as measured by Celex, logarithm for all words).
4. Lexical diversity, word hypernymy values and content word frequency explained 44% of the variance of the human evaluations of lexical proficiency in the examined writing samples.
5. The strongest predictors of an individual's proficiency level were word imageability, word frequency, lexical diversity, and word familiarity.
6. Some aspects of L2 writing may not be cultural or independent, but rather based on the amount and type of linguistic knowledge available to L2 learners as a result of language experience and learner proficiency level.
7. As grade level increases, writers produce texts that are less cohesive and more elaborate.

Crossley and McNamara (2010) also report findings from previous studies on L2 writing quality which include the following features.

- Lexical diversity: More proficient L2 writers use a more diverse range of words, and thus show greater lexical diversity (c.f. Engber, 1995; Grant and Ginther, 2000; Jarvis, 2002).
- Cohesion: More proficient L2 writers produce texts with a greater variety of lexical and referential cohesive devices (including all connectives) than less proficient writers (c.f. Connor, 1990; Ferris, 1994; Jin, 2001).
- Word frequency: More proficient L2 writers use less frequent words (c.f. Frase, Falletti, Ginther, and Grant, 1997; Grant and Ginther, 2000; Reid, 1986, 1990; Reppen, 1994).
- Linguistic sophistication: More proficient L2 writers produce texts with more syntactic complexity.

Accordingly, we conclude that lexical diversity, cohesive devices, word frequency, and linguistic sophistication are good predictors of L2 writing quality.

The reviewed studies provide the theoretical background for the use of Coh-Metrix and the selected indices to compare IELTS Academic test-takers' writing scripts across the three band scores and three L1 groups explored in the current study. The methodological aspects of the study are presented in the next section.

## 2.3 Methods

### 2.3.1 Materials

Table 2.2 presents the number of scripts included in the current analysis (cf. Table 1.1).

| | L1 category | | | |
|---|---|---|---|---|
| **Band score** | **European -based** | **Hindi** | **Arabic** | **Total** |
| 5 | 30 | 27 | 30 | 87 |
| 6 | 30 | 27 | 29 | 86 |
| 7 | 30 | 30 | 21 | 81 |
| Total | 90 | 84 | 80 | 254 |

***Table 2.2: Number of scripts included in the analyses***

### 2.3.2 Quantitative text analysis procedures

To analyse the linguistic features of the scripts, Coh-Metrix 2.0 software was used (see http://cohmetrix.memphis.edu; McNamara, Louwerse, McCarthy, and Graesser 2010; Graesser, McNamara and Kulikowich 2011).

Crossley and McNamara (2010) explain that:
"The tool was constructed to investigate various measures of text and language comprehension that augment surface components of language by exploring deeper, more global attributes of language. The tool is informed by various disciplines such as discourse psychology, computational linguistics, corpus linguistics, information extraction and information retrieval. As such, Coh-Metrix integrates lexicons, pattern classifiers, part-of speech taggers, syntactic parsers, shallow semantic interpreters and other components common in computational linguistics." (p 4)

Coh-Metrix provides general word and text information such as number of words, number of sentences, number of paragraphs, number of words per sentence, number of sentences per paragraph, and two readability indices—Flesch Reading Ease and Flesch-Kincaid Grade Level. In addition to identifying the word and text information of the scripts, Coh-Metrix was also used to analyse the scripts and provide indices for the following textual features:

1. Word Frequency Level (WFL). The inclusion of this feature represents the fact that the pattern of word use from different frequency levels is supposed to be different for more proficient writers as compared to writers of low proficiency. The word frequency index of test-takers' scripts is worthy of investigation because Lexical Resource is one of the criteria used by IELTS rating scale and raters. Coh-Metrix reports average frequency counts for the majority of the individual words in the text using CELEX (Baayen, Piepenbrock, and Gulikers 1995 cited in Crossley, Salsbury and McNamara 2011). This index provides Celex, logarithm, and the mean for content words on a scale of 0–6. Content words including nouns, adverbs, adjectives, and main verbs are normally considered in word frequency (WF) computations (Graesser, McNamara and Kulikowich 2011). The word with the lowest mean log frequency comes from low-frequency word lists. An average of log frequency for the words in the scripts is computed and included in the analyses. If the log frequency for texts approaches zero, the interpretation is that the text is difficult to understand because the words come from low-frequency lists.

2. Lexical diversity. Another lexical feature related to both Lexical Resource and grammatical complexity and included in the quantitative analysis of the scripts is lexical diversity. This is operationalised by type-token ratio (Templin, 1957 cited in Coh-Metrix documents).

   This is the number of unique words (called types) divided by the number of tokens of these words. In other words, each unique word in a text is a word type and each instance of a particular word is a token. For example, if the word "information" appears in a text nine times, its type and token values will be one and nine respectively. When type-token ratio approaches one, it means each word occurs only once in the text and, as such, comprehension should be comparatively difficult because many unique words are used to form the text.

   On the other hand, "indices of lexical diversity assess a writer's range of vocabulary and are indicative of greater linguistic skills" (e.g., Ransdell and Wengelin, 2003 cited in McNamara et al. 2010, p 70). Accordingly, texts with higher indices of lexical diversity will presumably be rewarded by raters in the IELTS Academic Writing Test because Lexical Resource is one of the rating criteria.

   One challenge confronting computation of the Type-Token Ratio (TTR) index is text length. Accurate measures of TTR need to be calculated for texts of comparable lengths. At the time we ran the analysis, Coh-Metrix version 2 was accessible which used TTR as the index for lexical diversity. More recently (early 2013), Coh-Metrix version 3 has incorporated Measures of Textual Lexical Diversity (MTLD) that control for text length (Graesser et al. 2011). MTLD allows for comparisons between text segments of considerably different lengths (at least 100 to 2000 words). However, given the limited length of the texts produced by IELTS test-takers, we believe that the TTR measure of Coh-Metrix version 2 remains a reliable index of lexical diversity for the IELTS scripts analysed in this study.

3. Grammatical complexity. Since one of the criteria used in the IELTS Academic Writing Test is Grammatical Range and Accuracy, we were interested to find out if grammatical complexity (operationalised as the number of words before the main verb of the main clause in the sentences of a text) in test-takers' scripts differentiates among the scripts of the three band scores and L1 groups. Sentences that have many words before the main verb are believed to put heavier loads on working memory of the readers, thus rendering more complex sentences.

In addition to the above textual features, indices of all connectives and coreferentiality (stem and argument overlap) were also calculated to obtain a quantitative measure of cohesion as a discoursal feature of the scripts. These features are explained below.

4. Incidence of all connectives. According to Halliday and Hasan (1976), connectives are among important classes of devices for particular categories of cohesion relations in text. Coh-Metrix 2.0 provides an index for all connectives including both positive (e.g., and, after, because) and negative (e.g., but, until, although) as well as other connectives

associated with the type of cohesion—additive (e.g., also, moreover), temporal (e.g., before, after, when, until), logical (e.g., if, or), and causal (e.g., because, so, consequently, nevertheless).

5. Argument overlap. This is the proportion of sentence pairs that share one or more arguments (i.e., noun, pronoun, noun-phrase).

6. Stem overlap. This is the proportion of sentence pairs in which a noun in one sentence has a semantic unit in common with any word in any grammatical category in other sentence (e.g., the noun "photograph" and the verb "photographed") (Graesser et al. 2011).

Indices of all connectives and coreferentiality can therefore provide useful information about text cohesion and, indirectly, about text coherence. For all the textual features, mean indices are computed and used in the analyses and results.

## 2.4 Results of the quantitative analysis

Table 2.3 presents the overall mean for a number of the textual features of the scripts in the three band scores.

The following three observations can be made from Table 2.3.

1. As we move from band 5 to band 7 the number of words in test-takers' scripts increases from a mean of 284 to a mean of 331 words, meaning that test-takers with higher band scores tend to produce lengthier texts. The standard deviation (numbers in parenthesis) is also indicative that, as we move from lower band scores (5) to higher band scores (7), there is less variation in test-takers' texts in terms of the length of their scripts. The same observation is true for the number of sentences and number of paragraphs. Results of analysis of variance (ANOVA) showed a significant difference among the three band score texts in terms of the number of words ($F= 8.80$, $df=2$, $p<0.001$). This may imply that text length has been a determining factor in rating the essays, a finding in line with that of Mayor et al. (2007) who also found text length as one of the strongest predictors of high scored scripts.

   Moreover, Crossley and McNamara (2010, p 6) cite Ferris (1994) and Frase, Faletti, Ginther and Grant (1997), arguing that "text length has historically been a strong predictor of essay scoring with most studies reporting that text length explains about 30% of the variance in human scores".

2. Scripts of band score 7 have fewer words per sentence and less variation, compared to scripts at band scores 5 and 6. This may imply that high scorers (band 7) produce more concise sentences.

3. Number of sentences per paragraph does not convey any particular pattern, while the Flesch Reading Ease index, or the readability index, is certainly capable of differentiating among the three groups. The Flesch Reading Ease Readability index uses two key variables in the calculation of the index: the average sentence length (ASL), and the average syllables per word (ASW). An index of 60–70 indicates standard texts, and 50–60 indicates fairly difficult texts (Heydari and Riazi, 2012). The range of the readability index is 20–100, and lower scores are indicative of more difficult texts.

The information in Table 2.3 shows that scripts with lower readability indices have been rated higher. As can be seen from Table 2.3, the mean and standard deviation of the readability of the scripts for band 5, 6, and 7 were 58.34 (SD=12.33), 56.60 (SD=9.57), and 54.01 (SD=8.5) respectively. Among the three groups, scripts within band 7 were found to be more homogenous, as indicated by their lower standard deviation.

Table 2.4 presents the mean and standard deviation (in parenthesis) for more linguistic features of the scripts. The Flesch Reading Ease index is also included in this table as it is used as one of the variables in the statistical analysis.

In addition to the Flesch Reading Ease, lexical diversity (TTR), word frequency (Celex, log, mean for content words), syntactic complexity (mean number of words before the main verb), and indices of cohesion (all connectives and coreferentiality) also show patterns in the data.

| Variable | Band 5 | Band 6 | Band 7 |
|---|---|---|---|
| No. of words | 284.19 (68.35) | 308.23 (64.85) | 330.58 (62.55) |
| No. of sentences | 14.96 (5.38) | 16.17 (4.55) | 16.73 (4.23) |
| No. of paragraphs | 4.65 (1.53) | 4.7 (1.72) | 4.78 (1.28) |
| No. of words per sentence | 20.35 (6.12) | 20.12 (5.9) | 20 (3.98) |
| No. of sentences per paragraph | 3.8 (2.67) | 4.10 (3.15) | 3.83 (1.51) |
| Flesch Reading Ease index (Readability) | 58.34 (12.33) | 56.60 (9.57) | 54.01 (8.5) |

*Table 2.3: Mean and standard deviation of some features of the scripts at the three band scores*

| | Band score | | |
|---|---|---|---|
| | **5 (n=86)** | **6 (n=86)** | **7 (n=82)** |
| Readability (Flesch Reading Ease) | 58.34 (12.33) | 56.59 (9.57) | 54.01 (8.49) |
| Type-Token Ratio (TTR) | 0.68 (0.85) | 0.68 (0.80) | 0.7 (0.67) |
| Word frequency (Celex, log, mean for content words) | 2.53 (0.147) | 2.47 (0.126) | 2.38 (0.109) |
| Syntactic complexity (Mean no. of words before the main verb) | 4.41 (1.82) | 4.38 (1.14) | 4.48 (1.31) |
| Cohesion (Incidence of all connectives) | 88.03 (22.13) | 87.74 (17.63) | 86.62 (16.76) |
| Cohesion (Coreference: Argument overlap) | 0.54 (0.20) | 0.48 (0.19) | 0.48 (0.17) |
| Cohesion (Coreference: Stem overlap) | 0.49 (0.23) | 0.45 (0.20) | 0.45 (0.19) |

***Table 2.4: Descriptive statistics for linguistic features of the scripts across the three band scores***

As shown in Table 2.4, the TTR increases and approaches a value of 1 as we move from band 5 to band 7, indicating test-takers with higher scores used a greater range of lexis in their texts. Moreover, the Celex index (with the scale of 0–6) shows that band 7 scripts use more infrequent words compared to scripts in the other two band groups. This observation is also true with regard to syntactic complexity, with band 7 scripts showing a higher average number of words before the main verb compared particularly with band 5 scripts. However, this observation is not consistent between bands 5 and 6. Interestingly, measures of cohesion decrease as we move from band 5 to 7 for all connectives and between band 5 and the other two band scores (6 and 7) for argument and stem overlap.

These observations point to the fact that scripts which have received higher band scores have shown to represent higher levels of linguistic complexity, but they are not necessarily more cohesive. This finding is in line with previous findings as reported above. Our findings are particularly consistent with those of Mayor et al. (2007) and Banerjee et al. (2007). Mayor et al. found sentence complexity, and Banerjee et al. fond type-token ratio and word frequency (lexical sophistication) among the strong predictors of high scores on IELTS writing tasks. Furthermore, Crossley et al. (2011) found that as grade level increases, writers produce texts that are less cohesive and more elaborate. An implication of this finding is that text complexity has been rewarded more than text cohesion in the ratings of Task 2 of the IELTS Academic Writing Test. Given that some indices of cohesion were the same for bands 6 and 7, this finding is most important for distinguishing between band 5 and band 6 scripts in our data.

To this point, we can see some consistencies in band scores in terms of linguistic features of the scripts. Of course, this observation needs to be verified through inferential statistical analyses if we want to generalise from this sample to the whole population of the three band scores and L1 groups. Table 2.5 presents the same linguistic features across the band scores and L1 categories. The information in Table 2.5 can help us infer how scripts related to the three L1 categories are rated.

| | **European-based** | | | **Hindi** | | | **Arabic** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **5 (n=30 )** | **6 (n=30 )** | **7 (n=30 )** | **5 (n=27 )** | **6 (n=27 )** | **7 (n=30 )** | **5 (n=30 )** | **6 (n=29 )** | **7 (n=21 )** |
| Flesch Reading Ease | 58.64 (13.12) | 52.92 (8.54) | 57.04 (9.05) | 62.34 (11.23) | 58.54 (9.52) | 51.51 (7.17) | 54.57 (11.46) | 53.56 (10.1) | 53.35 (8.51) |
| TTR | 0.7 (0.07) | 0.72 (0.07) | 0.70 (0.06) | 0.64 (0.08) | 0.66 (0.08) | 0.70 (0.06) | 0.69 (0.08) | 0.68 (0.06) | 0.72 (0.07) |
| Word frequency (Celex, log, mean for content words) | 2.55 (0.13) | 2.47 (0.12) | 2.44 (0.09) | 2.52 (0.14) | 2.5 (0.12) | 2.35 (0.12) | 2.52 (0.08) | 2.46 (0.06) | 2.38 (0.07) |
| Syntactic complexity (Mean no. of words before the main verb) | 4.37 (2.1) | 4.39 (1.14) | 4.72 (1.4) | 4.46 (1.93) | 4.10 (0.99) | 4.35 (1.37) | 4.42 (1.41) | 4.64 (1.24) | 4.35 (1.10) |
| Cohesion (Incidence of all connectives) | 83.18 (19.74) | 84.66 (14.6) | 87.04 (16.07) | 93.26 (22.95) | 88.73 (19.3) | 83.9 (18.31) | 88.35 (22.9) | 89.82 (18.91) | 90.05 (15.37) |
| Cohesion (Argument overlap) | 0.53 (0.20) | 0.40 (0.14) | 0.47 (0.14) | 0.51 (0.19) | 0.51 (0.19) | 0.52 (0.20) | 0.58 (0.19) | 0.53 (0.22) | 0.47 (0.18) |
| Cohesion (Stem overlap) | 0.45 (0.26) | 0.37 (0.14) | 0.43 (0.18) | 0.48 (0.21) | 0.46 (0.19) | 0.52 (0.20) | 0.55 (0.20) | 0.52 (0.23) | 0.39 (0.17) |

***Table 2.5: Descriptive statistics for linguistic features of the scripts across the three band scores and L1 categories***

Table 2.6 shows the Pearson correlation among the textual features of the scripts.

| | | Flesch Reading Ease | Mean no. of words before the main verb | TTR | Celex, log, mean for content words | Incidence of all connectives | Argument overlap | Stem overlap |
|---|---|---|---|---|---|---|---|---|
| Flesch Reading Ease | Pearson Correlation | 1 | -.300** | -.308** | .524** | .020 | -.284** | -.390** |
| | Sig. (2-tailed) | | .000 | .000 | .000 | .747 | .000 | .000 |
| | N | 254 | 254 | 254 | 254 | 254 | 254 | 254 |
| Mean no. of words before the main verb | Pearson Correlation | -.300** | 1 | .154* | -.047 | .150* | .184** | .171** |
| | Sig. (2-tailed) | .000 | | .014 | .458 | .017 | .003 | .006 |
| | N | 254 | 254 | 254 | 254 | 254 | 254 | 254 |
| TTR | Pearson Correlation | -.308** | .154* | 1 | -.459** | -.115 | -.343** | -.343** |
| | Sig. (2-tailed) | .000 | .014 | | .000 | .067 | .000 | .000 |
| | N | 254 | 254 | 254 | 254 | 254 | 254 | 254 |
| Celex, log, mean for content words | Pearson Correlation | .524** | -.047 | -.459** | 1 | .165** | .153* | .047 |
| | Sig. (2-tailed) | .000 | .458 | .000 | | .008 | .015 | .453 |
| | N | 254 | 254 | 254 | 254 | 254 | 254 | 254 |
| Incidence of all connectors | Pearson Correlation | .020 | .150* | -.115 | .165** | 1 | .193** | .222** |
| | Sig. (2-tailed) | .747 | .017 | .067 | .008 | | .002 | .000 |
| | N | 254 | 254 | 254 | 254 | 254 | 254 | 254 |
| Coreference (Argument overlap) | Pearson Correlation | -.284** | .184** | -.343** | .153* | .193** | 1 | .869** |
| | Sig. (2-tailed) | .000 | .003 | .000 | .015 | .002 | | .000 |
| | N | 254 | 254 | 254 | 254 | 254 | 254 | 254 |
| Coreference (Stem overlap) | Pearson Correlation | -.390** | .171** | -.343** | .047 | .222** | .869** | 1 |
| | Sig. (2-tailed) | .000 | .006 | .000 | .453 | .000 | .000 | |
| | N | 254 | 254 | 254 | 254 | 254 | 254 | 254 |

***Table 2.6: Relationship between the measures of the linguistic features of the scripts***

Before performing Multivariate Analysis of Variance (MANOVA) with band score and L1 as independent variables and the textual features of the scripts as the dependent variables, we needed to ensure that there are not high correlations among the dependent variables. Table 2.6 presents the results of the Pearson correlation among the seven measures (dependent variables). As can be seen in Table 2.6, there is only a high ($r$= 0.87) and significant ($p<0.01$) correlation between the two measures of coreferentiality (argument overlap and stem overlap). This is indeed natural as the two measures are highly related as measures of text cohesion. We will, therefore, include only one of these two measures (stem overlap) in MANOVA analysis. The choice of stem overlap is based on the fact that, as Table 2.5 indicates, it showed more variation across band scores compared to argument overlap.

Accordingly, a two-way MANOVA was run to find out if there is a significant difference among the six measures of textual features in terms of band scores and three L1 categories. Before running the MANOVA we need to check the following assumptions (Pallant 2007; Stevens 1996) for this parametric test:

1. sample size
2. normality
3. outliers
4. linearity
5. homogeneity of regression
6. multicollinearity and singularity
7. homogeneity of variance-covariance matrices.

In terms of sample size, as Stevens (1996) argues, we should have at least 20 participants for every dependent variable, thus 140 for the seven dependent variables in this study. Our sample size goes well beyond this. Normality of the seven dependent variables was checked through histograms and though they were not perfectly normal, no abnormality was observed. Moreover, as Pallant (2007, p 277) states, "in practice it (MANOVA) is reasonably robust to modest violation of normality".

The outliers were checked using both univariate (through box plots) and multivariate (through Mahalanobis distances) normality. The box plots for univariate normality indicated the following outliers for the designated variables.

| Variable | Case |
|---|---|
| Flesch Reading Ease | 39 |
| Mean number of words before the main verb | 3, 39, 48, 54, 77, 88, 237 |
| TTR | 3, 8, 32, 96, 142, 250 |
| Incidence of all connectives | 72 |

*Table 2.7: Univariate results for outliers*

As relates to the multivariate outliers, the Mahalanobis distance was found to be 32.64 which was higher than the critical value (24.32) with six dependent variables. Using the critical value as our reference, the four multivariate outliers were found to be cases 19, 39, 3, and 88 with Mahl distances of 32.64, 31.8, 26.31, and 24.38 respectively. Accordingly, the decision was made to exclude cases 3, 39, and 88 which were common between the univariate and multivariate outliers and case 19 which indicated the largest Mahl distance (32.64). Moreover, since MANOVA can deal with only a few outliers, more univariate outliers, including cases 8, 77, 96, and 142, were deleted from MANOVA analysis. The deleted cases were five band 5 test-takers (cases 3, 8, 19, 39, 77) and three band 6 cases (88, 96, 142). They were also five European-based L1 cases (3, 19, 39, 88, 96), two Hindi L1 cases (77, 142) and one Arabic L1 test-taker (8).

This left us with n=247 which was still beyond the set sample size criteria for MANOVA. To check the linearity of the dependent variables, a matrix of scatterplots between each pair of the variables, separately for our groups were obtained. These plots did not show any obvious evidence of non-linearity. Therefore, the assumption of linearity was satisfied. The following table presents the ultimate number of scripts included in MANOVA.

| | | Value label | N |
|---|---|---|---|
| **Band group** | 1 | Band 5 | 82 |
| | 2 | Band 6 | 83 |
| | 3 | Band 7 | 82 |
| **L1 category** | 1 | European-based | 84 |
| | 2 | Hindi | 82 |
| | 3 | Arabic | 81 |

*Table 2.8: Number of scripts across band score and L1 categories included in MANOVA*

Table 2.9 shows the correlation between the six dependent variables and the overall mean and standard deviation for each variable.

As can be seen from Table 2.9, the highest significant and direct relationship is between Flesch Reading Ease and word frequency ($r$= 0.524). The highest significant and reverse relationship exists between lexical diversity and word frequency.

Homogeneity of regression was not an issue here because it is only important if stepdown analysis is to be done (Pallant 2007, p 282), which was not the case in this study. Pearson correlation was run between the seven dependent variables to check the multicollinearity (when the dependent variables are highly correlated). As can be seen in Table 2.5 these variables were moderately correlated, with the exception of the two variables related to coreference (argument overlap and stem overlap) which were highly and significantly correlated (r= 0.87, p<0.01). Given the common variance between these two variables, it was therefore decided to include only one of them (stem overlap) in the MANOVA model.

Finally, the test of homogeneity of variance–covariance is generated as part of MANOVA output (Box's M Test of Equality of Covariance Matrices) as presented below. Since the significance value (0.180) is much larger than 0.001, we have not violated the homogeneity of variance–covariance.

| | Flesh Reading Ease | Syntactic complexity | TTR | WF | Connectives | Stem overlap | Mean | SD |
|---|---|---|---|---|---|---|---|---|
| Flesh Reading Ease | 1 | | | | | | 56.66 | 10.15 |
| Syntactic complexity | -.300** | 1 | | | | | 4.36 | 1.31 |
| TTR | -.308** | .154* | 1 | | | | 0.69 | 0.07 |
| WF | .524** | -.047 | -.459** | 1 | | | 2.46 | 0.14 |
| Connectives | .020 | .150* | -.115 | .165** | 1 | | 87.59 | 18.64 |
| Stem overlap | -.390** | .171** | -.343** | .047 | .222** | 1 | 0.46 | 0.20 |

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

***Table 2.9: Correlation matrix for the six dependent variables***

| | |
|---|---|
| **Box's M** | 203.719 |
| F | 1.099 |
| df1 | 168 |
| df2 | 55808.043 |
| Sig. | .180 |

***Table 2.10: Box's test of equality of covariance matrices***

The Levene's test of equality of error variances is presented below. Mean number of words before the main verb and Celex, log, mean for content words, violated equality of variances because the significance values for these two variables are less than 0.05. We, therefore, need to set a more conservative alpha level for determining significance for these variables in the univariate F-test (Pallant 2007). Therefore, as Tabachnick and Fidell (2007) suggest, we use 0.025 rather than 0.05 as the set level of significance for findings.

| | F | df1 | df2 | Sig. |
|---|---|---|---|---|
| Flesch Reading Ease | 1.764 | 8 | 238 | .085 |
| Mean no. of words before the main verb | 2.364 | 8 | 238 | .018 |
| TTR | .641 | 8 | 238 | .743 |
| Celex, log, mean for content words | 2.081 | 8 | 238 | .038 |
| Incidence of all connectives | 1.688 | 8 | 238 | .102 |
| Coreference (Stem overlap) | 1.955 | 8 | 238 | .053 |

a. Design: Intercept + Band group + L1 category + Band group * L1 category

***Table 2.11: Levene's test of equality of error variances[a]***

Results of the two-way MANOVA using the six criterion variables across the three band scores and L1 categories are presented in Table 2.12.

| Effect | | Value | F | Hypothesis df | Error df | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|---|
| Intercept | Pillai's Trace | .999 | 38460.625[a] | 6.000 | 233.000 | .000 | .999 |
| | Wilks' Lambda | .001 | 38460.625[a] | 6.000 | 233.000 | .000 | .999 |
| | Hotelling's Trace | 990.402 | 38460.625[a] | 6.000 | 233.000 | .000 | .999 |
| | Roy's Largest Root | 990.402 | 38460.625[a] | 6.000 | 233.000 | .000 | .999 |
| BandGroup | Pillai's Trace | .192 | 4.132 | 12.000 | 468.000 | .000 | .096 |
| | Wilks' Lambda | .810 | 4.328[a] | 12.000 | 466.000 | .000 | .100 |
| | Hotelling's Trace | .234 | 4.523 | 12.000 | 464.000 | .000 | .105 |
| | Roy's Largest Root | .228 | 8.888[b] | 6.000 | 234.000 | .000 | .186 |
| L1Category | Pillai's Trace | .146 | 3.061 | 12.000 | 468.000 | .000 | .073 |
| | Wilks' Lambda | .859 | 3.059[a] | 12.000 | 466.000 | .000 | .073 |
| | Hotelling's Trace | .158 | 3.056 | 12.000 | 464.000 | .000 | .073 |
| | Roy's Largest Root | .103 | 4.036[b] | 6.000 | 234.000 | .001 | .094 |
| BandGroup * L1Category | Pillai's Trace | .143 | 1.457 | 24.000 | 944.000 | .072 | .036 |
| | Wilks' Lambda | .863 | 1.463 | 24.000 | 814.050 | .071 | .036 |
| | Hotelling's Trace | .152 | 1.466 | 24.000 | 926.000 | .069 | .037 |
| | Roy's Largest Root | .088 | 3.466[b] | 6.000 | 236.000 | .003 | .081 |

a. Exact statistic
b. The statistic is an upper bound on F that yields a lower bound on the significance level
c. Design: Intercept + BandGroup + L1Category + BandGroup * L1Category

***Table 2.12: Multivariate tests[c]***

The two-way MANOVA revealed significant multivariate main effect for band group (Wilks' λ =0.810, $F = 4.33$, p <. 001, partial eta squared =0.10) and L1 category (Wilks' λ =0.859, $F$=3.06, p <. 001, partial eta squared =0.07). The second part of MANOVA results are Tests of Between-Subjects Effects which is presented in Table 2.13.

| Source | Dependent variable | Type III sum of squares | df | Mean square | F | Sig. | Partial Eta squared |
|---|---|---|---|---|---|---|---|
| **Corrected model** | Flesch Reading Ease | 3145.117[a] | 8 | 393.140 | 4.217 | .000 | .124 |
| | Mean no. of words before the main verb | 11.726[b] | 8 | 1.466 | .849 | .560 | .028 |
| | TTR | .129[c] | 8 | .016 | 2.971 | .003 | .091 |
| | Celex, log, mean for content words | .966[d] | 8 | .121 | 7.601 | .000 | .203 |
| | Incidence of all connectives | 1611.422[e] | 8 | 201.428 | .572 | .801 | .019 |
| | Coreference (Stem overlap) | .800[f] | 8 | .100 | 2.450 | .014 | .076 |
| **Intercept** | Flesch Reading Ease | 785938.54 | 1 | 785938.4 | 8429.98 | .000 | .973 |
| | Mean no. of words before the main verb | 4603.178 | 1 | 4603.178 | 2666.95 | .000 | .918 |
| | TTR | 116.343 | 1 | 116.343 | 21405.4 | .000 | .989 |
| | Celex, log, mean for content words | 1481.398 | 1 | 1481.398 | 93287.6 | .000 | .997 |
| | Incidence of all connectors | 1875028.59 | 1 | 1875028.6 | 5323.17 | .000 | .957 |
| | Coreference (Stem overlap) | 51.671 | 1 | 51.671 | 1266.08 | .000 | .842 |

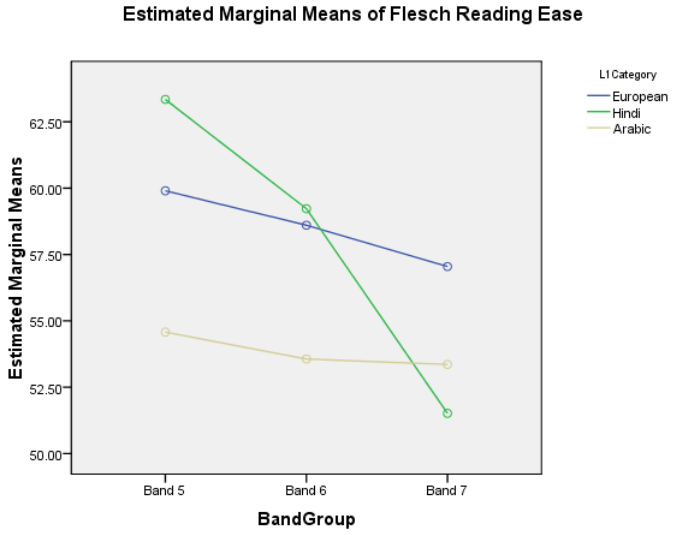
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **BandGroup** | Flesch Reading Ease | 1142.651 | 2 | 571.325 | 6.128 | .003 | .049 |
| | Mean no. of words before the main verb | 2.266 | 2 | 1.133 | .657 | .520 | .005 |
| | TTR | .046 | 2 | .023 | 4.199 | .016 | .034 |
| | Celex, log, mean for content words | .807 | 2 | .403 | 25.407 | .000 | .176 |
| | Incidence of all connectives | 95.021 | 2 | 47.511 | .135 | .874 | .001 |
| | Coreference (Stem overlap) | .068 | 2 | .034 | .838 | .434 | .007 |
| **L1Category** | Flesch Reading Ease | 1063.006 | 2 | 531.503 | 5.701 | .004 | .046 |
| | Mean no. of words before the main verb | 2.362 | 2 | 1.181 | .684 | .506 | .006 |
| | TTR | .062 | 2 | .031 | 5.695 | .004 | .046 |
| | Celex, log, mean for content words | .056 | 2 | .028 | 1.750 | .176 | .014 |
| | Incidence of all connectives | 622.163 | 2 | 311.082 | .883 | .415 | .007 |
| | Coreference (Stem overlap) | .242 | 2 | .121 | 2.968 | .053 | .024 |
| **BandGroup * L1Category** | Flesch Reading Ease | 934.114 | 4 | 233.529 | 2.505 | .043 | .040 |
| | Mean No. of words before the main verb | 6.249 | 4 | 1.562 | .905 | .462 | .015 |
| | TTR | .034 | 4 | .008 | 1.552 | .188 | .025 |
| | Celex, log, mean for content words | .090 | 4 | .023 | 1.423 | .227 | .023 |
| | Incidence of all connectives | 934.477 | 4 | 233.619 | .663 | .618 | .011 |
| | Coreference (Stem overlap) | .441 | 4 | .110 | 2.700 | .031 | .043 |
| **Error** | Flesch Reading Ease | 22189.044 | 238 | 93.231 | | | |
| | Mean no. of words before the main verb | 410.792 | 238 | 1.726 | | | |
| | TTR | 1.294 | 238 | .005 | | | |
| | Celex, log, mean for content words | 3.779 | 238 | .016 | | | |
| | Incidence of all connectives | 83832.835 | 238 | 352.239 | | | |
| | Coreference (Stem overlap) | 9.713 | 238 | .041 | | | |
| **Total** | Flesch Reading Ease | 818418.01 | 247 | | | | |
| | Mean no. of words before the main verb | 5111.866 | 247 | | | | |
| | TTR | 119.354 | 247 | | | | |
| | Celex, log, mean for content words | 1506.283 | 247 | | | | |
| | Incidence of all connectives | 1980443.22 | 247 | | | | |
| | Coreference (Stem overlap) | 63.849 | 247 | | | | |
| **Corrected total** | Flesch Reading Ease | 25334.161 | 246 | | | | |
| | Mean no. of words before the main verb | 422.518 | 246 | | | | |
| | TTR | 1.423 | 246 | | | | |
| | Celex, log, mean for content words | 4.745 | 246 | | | | |
| | Incidence of all connectives | 85444.258 | 246 | | | | |
| | Coreference (Stem overlap) | 10.513 | 246 | | | | |

a. R Squared = .124 (Adjusted R Squared = .095)
b. R Squared = .028 (Adjusted R Squared = -.005)
c. R Squared = .091 (Adjusted R Squared = .060)
d. R Squared = .203 (Adjusted R Squared = .177)
e. R Squared = .019 (Adjusted R Squared = -.014)
f. R Squared = .076 (Adjusted R Squared = .045)

***Table 2.13: Tests of between-subjects effects***

Given the significance of the overall MANOVA test, the univariate main effects were examined through tests of between-subjects effects. Because we look at a number of separate analyses here, we use Bonferroni adjustment (Pallant 2007). Accordingly, we set the level of significance to 0.004 or less for each of the six variables (0.025/6= 0.004).

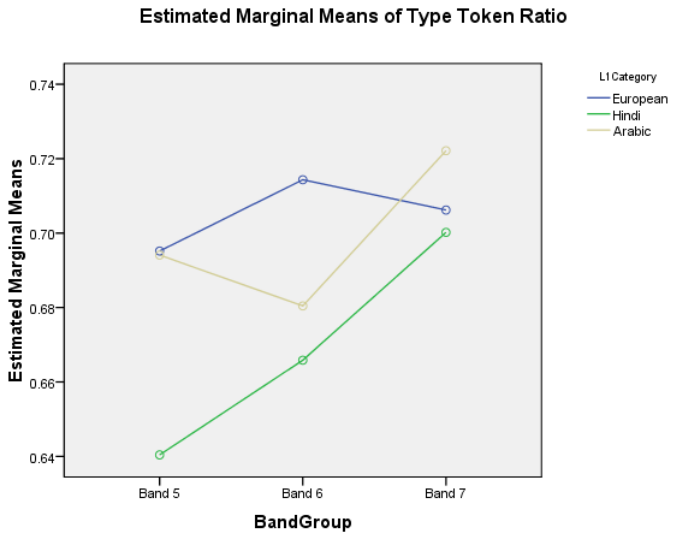
Accordingly, significant univariate main effects for band groups were obtained for Flesch Reading Ease ($p$<0.001, partial eta squared=0.05) and Word Frequency ($p$<0.004, partial eta squared=0.18). Also, significant main effects for L1 category were obtained for Flesch Reading Ease ($p$=0.004, partial eta square =0.047) and TTR (p=0.004, partial eta square=0.047). The other variables did not show either significant difference or if they did, they did not meet the set criteria value of being lower than 0.004. This holds true for the interaction between band score and L1 category. The importance of the impact of these linguistic features of the scripts on band scores and L1 categories can be evaluated using the effect size (partial eta squared) which represents the proportion of the variance in the band score accounted for by the linguistic features of the scripts. The effect size for Flesch Reading Ease and Word Frequency (Celex, log, mean for content words) for band groups were 0.05 and 0.18 respectively. This means that 5% of variance in group differences (band scores) can be accounted for by Flesch Reading Ease and 18% of variance in band group difference by Word Frequency. On the other hand, the effect size of Flesch Reading Ease and Lexical Diversity (TTR) for L1 category were 0.047 which can be rounded up to 0.05, meaning that 5% of variance in L1 category differences could be accounted for by Flesch Reading Ease and 5% by Lexical Diversity.




The following figures present the comparison of the three L1 categories across the three band scores in terms of the significant results of the linguistic features of the scripts.

As can be seen from Figure 2.1, Flesch Reading Ease had the most variation for the scripts written by Hindi L1 test-takers across the three band scores. Moving from band 5 to band 7, Hindi L1 test-takers produced more difficult texts consistently. Scripts written by European-based L1 test-takers showed the next highest variation, and scripts written by Arabic L1 test-takers had the least variation in terms of Flesch Reading Ease across the three band scores. In conclusion, while Flesch Reading Ease could differentiate both among the three band scores and the three L1 categories, this differentiation was more significant for scripts written by Hindi L1 test-takers.

***Figure 2.1: Estimated marginal means of Flesch Reading Ease***




Lexical Diversity did not show a significant difference among the three band scores. However, it did across the three L1 categories. As seen in Figure 2.2, the scripts written by Hindi L1 test-takers once again showed the most consistent pattern. As we move from band score 5 to 7, Hindi L1 test-takers have produced greater lexical diversity in their texts; a finding in line with previous studies as reviewed earlier. This is in line with the observation in Figure 2.1, in which scripts produced by Hindi L1 test-takers were shown to have lower indices of readability at higher band scores. In contrast, at band score 5, scripts produced by European-based L1 and Arabic L1 test-takers show exactly the same lexical diversity; at band score 6 they are diametrically different. That is, scripts written by European-based L1 test-takers represent greater lexical diversity, while scripts written by Arabic L1 test-takers at this band score show lower lexical diversity. At band 7, this pattern is almost reversed.

***Figure 2.2: Estimated marginal means of Lexical Diversity***

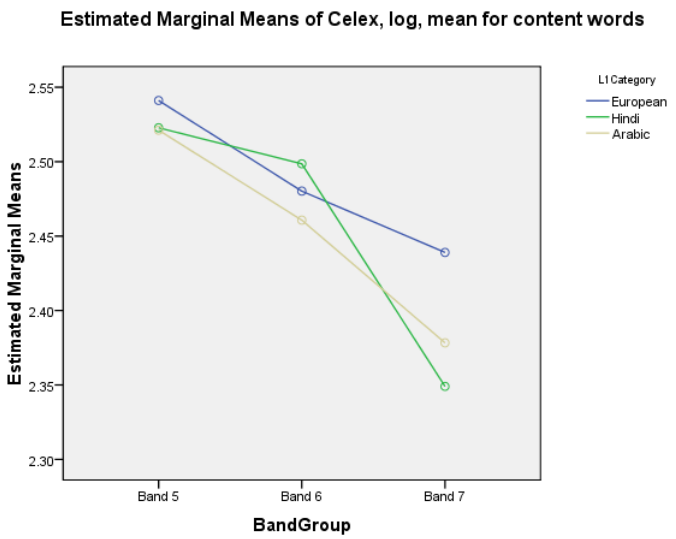



Figure 2.3 shows another interesting observation. Word frequency turns out to be a significant predictor of test-takers' writing performance in IELTS. All the three L1 categories present almost the same pattern. That is, as we move from band 5 to 7, the texts have increasingly used words from lower-frequency lists, regardless of the L1 category. Higher scores are assigned to scripts which included words from lower-frequency lists. Accordingly, results show that the Flesch Reading Ease and Word Frequency (Celex, log, mean for content words) have significantly and consistently differentiated among the scripts of the three band scores.

A follow-up Analysis of Variance (ANOVA) was conducted to find out where the differences in Flesch Reading Ease and Word Frequency indices in the three band score groups lie. The following is the result of ANOVA and Tukey's post-hoc test.

***Figure 2.3: Estimated marginal means of Word Frequency (Celex, log, mean for content words)***

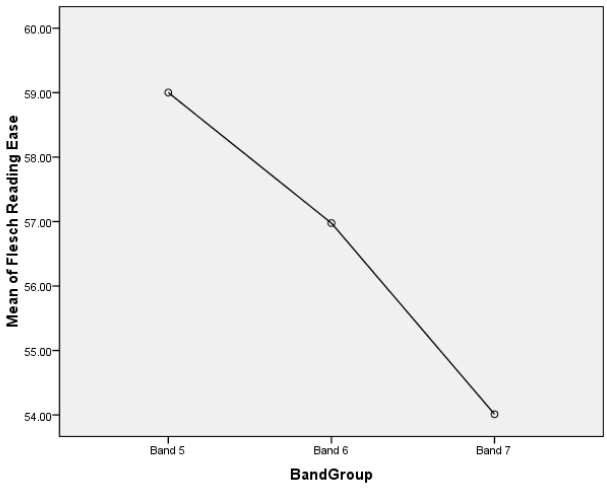
| | | Sum of squares | df | Mean square | F | Sig. |
|---|---|---|---|---|---|---|
| **Flesch Reading Ease** | Between groups | 1033.249 | 2 | 516.625 | 5.187 | .006 |
| | Within groups | 24300.912 | 244 | 99.594 | | |
| | Total | 25334.161 | 246 | | | |
| **Word Frequency (Celex, log, mean for content words)** | Between groups | .812 | 2 | .406 | 25.185 | .000 |
| | Within groups | 3.933 | 244 | .016 | | |
| | Total | 4.745 | 246 | | | |

***Table 2.14: ANOVA results***

| Dependent variable | (I) Band Group | (J) Band Group | Mean difference (I-J) | Std. error | Sig. |
|---|---|---|---|---|---|
| Flesch Reading Ease | Band 5 | Band 6 | 2.02610 | 1.55386 | .394 |
| | | Band 7 | 4.99052* | 1.55856 | .004 |
| | Band 6 | Band 5 | -2.02610 | 1.55386 | .394 |
| | | Band 7 | 2.96443 | 1.55386 | .139 |
| | Band 7 | Band 5 | -4.99052* | 1.55856 | .004 |
| | | Band 6 | -2.96443 | 1.55386 | .139 |
| Celex, log, mean for content words | Band 5 | Band 6 | .04929* | .01977 | .035 |
| | | Band 7 | .13878* | .01983 | .000 |
| | Band 6 | Band 5 | -.04929* | .01977 | .035 |
| | | Band 7 | .08949* | .01977 | .000 |
| | Band 7 | Band 5 | -.13878* | .01983 | .000 |
| | | Band 6 | -.08949* | .01977 | .000 |

*The mean difference is significant at the 0.05 level.

***Table 2.15: Post-hoc multiple comparisons: Tukey HSD***




As the results of the post-hoc Tukey test indicate, band scores 5 and 7 were differentiated in terms of Flesch Reading Ease. Word frequency (Celex, log, content words) was able to differentiate between the three band scores. Figures 2.4 and 2.5 present this information in graph form.

***Figure 2.4: Mean of Flesch Reading Ease over band scores***

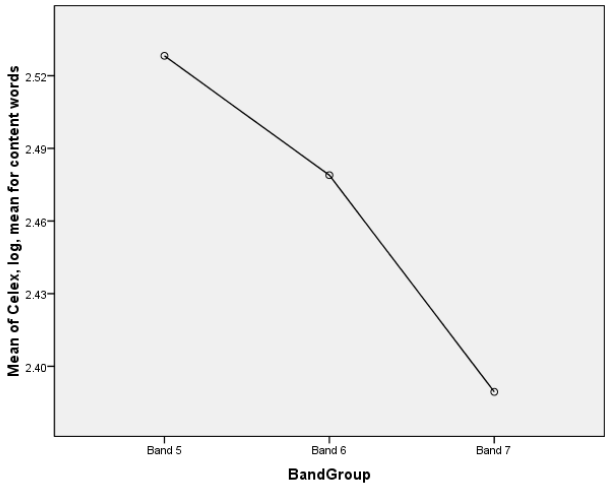



***Figure 2.5: Mean of Celex, log, mean for content words over band scores***

In addition, a follow-up ANOVA was also conducted to find out where the differences lay in terms of L1 categories and the two linguistic features which showed significant results. The following are the results.

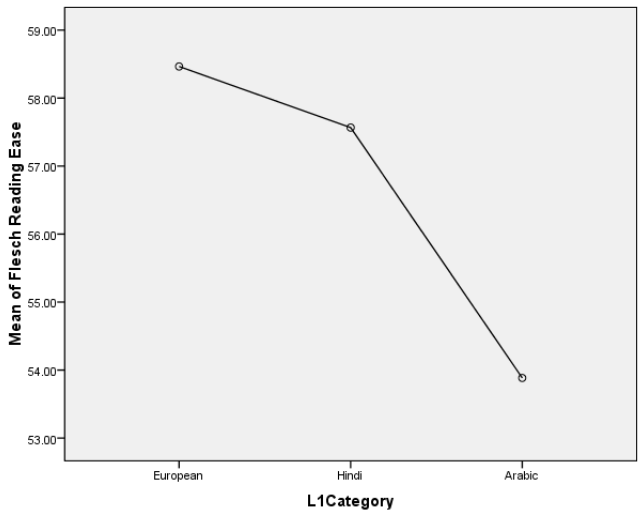
| | | Sum of squares | df | Mean square | F | Sig. |
|---|---|---|---|---|---|---|
| **Flesch Reading Ease** | Between groups | 964.819 | 2 | 482.409 | 4.830 | .009 |
| | Within groups | 24369.342 | 244 | 99.874 | | |
| | Total | 25334.161 | 246 | | | |
| **Lexical Diversity (TTR)** | Between groups | .052 | 2 | .026 | 4.624 | .011 |
| | Within groups | 1.371 | 244 | .006 | | |
| | Total | 1.423 | 246 | | | |

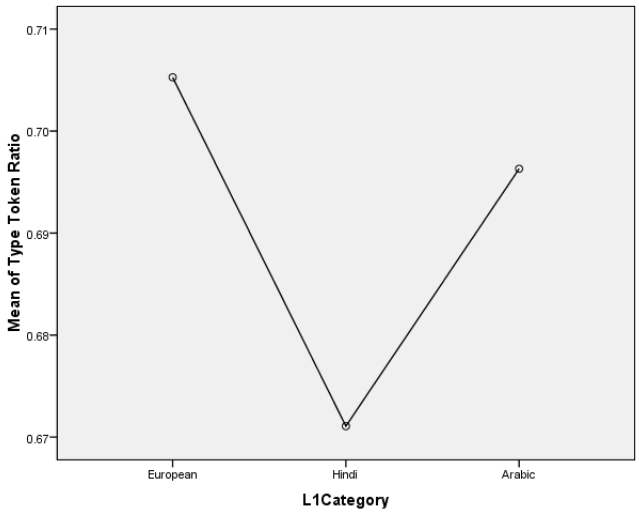
***Table 2.16: ANOVA results for L1 categories***

| Dependent variable | (I) L1Category | (J) L1Category | Mean difference (I-J) | Std. error | Sig. |
|---|---|---|---|---|---|
| **Flesch Reading Ease** | European | Hindi | .89835 | 1.55144 | .831 |
| | | Arabic | 4.57999* | 1.55628 | .010 |
| | Hindi | European | -.89835 | 1.55144 | .831 |
| | | Arabic | 3.68164 | 1.56557 | .051 |
| | Arabic | European | -4.57999* | 1.55628 | .010 |
| | | Hindi | -3.68164 | 1.56557 | .051 |
| **Lexical Diversity (TTR)** | European | Hindi | .03420* | .01164 | .010 |
| | | Arabic | .00897 | .01167 | .723 |
| | Hindi | European | -.03420* | .01164 | .010 |
| | | Arabic | -.02524 | .01174 | .082 |
| | Arabic | European | -.00897 | .01167 | .723 |
| | | Hindi | .02524 | .01174 | .082 |

* The mean difference is significant at the 0.05 level.

***Table 2.17: Multiple comparisons: Tukey HSD***




The following two figures also depict the results of the ANOVA and post-hoc test across the three L1 categories.

As the results of the post-hoc test (Table 2.17) and the two graphs show, European-based L1 scripts are significantly different from Hindi L1 and Arabic L1 scripts in terms of Flesch Reading Ease (for European-based L1 vs. Arabic L1) and Lexical Diversity (for European-based L1 vs. Hindi L1). The overall mean of Flesch Reading Ease was 56.2 and 53.8 for European-based L1 and Arabic L1 scripts, meaning that the scripts produced by Arabic L1 test-takers are more difficult to read.

***Figure 2.6: Mean of Flesch Reading Ease across the three L1 categories***

The overall mean of Lexical Diversity was 0.7 and 0.66 for European-based L1 and Hindi L1 scripts, meaning that scripts produced by European-based L1 test-takers were characterised by greater lexical diversity compared to those produced by Hindi L1 test-takers (despite some finer distinctions in this pattern when broken down by band score, as discussed in relation to Figure 2.2 above). In other words, there was more lexical variation in scripts produced by European-based L1 test-takers compared to those of Hindi L1 test-takers. In terms of simplicity and complexity, the texts produced by European-based L1 test-takers were therefore more complex compared to those produced by Hindi L1 test-takers.

***Figure 2.7: Mean of lexical diversity (TTR) across the three L1 categories***

Furthermore, we compared scripts scored at the same band level across the three L1 categories. The results of this comparison are presented below.

### 2.4.1 Comparison of scripts of the same band score across the three L1 categories

The third research question was concerned with whether consistency could be observed for the scripts scored at the same band level across the three different L1 categories. Accordingly, three ANOVA, together with post-hoc tests, were run for each band score across the three L1 categories with the six linguistic features as the dependent variables. Results are presented below.

As the results of the ANOVA in Table 2.18 show, the only significant difference observed between the band 5 scripts across the three L1 categories is Lexical Diversity (p=0.013). This implies that texts scored at band 5 were consistent in terms of the linguistic features measured across the three L1 categories, except for the measure of Lexical Diversity (TTR). To find out where the difference in the three L1 categories lies, a post-hoc test was run, and the results are presented in Table 2.19.

| | | Sum of squares | df | Mean square | F | Sig. |
|---|---|---|---|---|---|---|
| **Flesch Reading Ease** | Between groups | 859.679 | 2 | 429.840 | 2.987 | .056 |
| | Within groups | 12086.965 | 84 | 143.892 | | |
| | Total | 12946.644 | 86 | | | |
| **Syntactic Complexity (Mean no. of words before the main verb)** | Between groups | .125 | 2 | .062 | .018 | .982 |
| | Within groups | 284.065 | 84 | 3.382 | | |
| | Total | 284.190 | 86 | | | |
| **Lexical Diversity (TTR)** | Between groups | .062 | 2 | .031 | 4.606 | .013 |
| | Within groups | .562 | 84 | .007 | | |
| | Total | .624 | 86 | | | |
| **Word Frequency (Celex, log)** | Between groups | .015 | 2 | .007 | .333 | .718 |
| | Within groups | 1.842 | 84 | .022 | | |
| | Total | 1.857 | 86 | | | |
| **All connectives** | Between groups | 1445.323 | 2 | 722.662 | 1.509 | .227 |
| | Within groups | 40222.443 | 84 | 478.839 | | |
| | Total | 41667.766 | 86 | | | |
| **Stem overlap** | Between groups | .147 | 2 | .073 | 1.385 | .256 |
| | Within groups | 4.448 | 84 | .053 | | |
| | Total | 4.595 | 86 | | | |

*Table 2.18: ANOVA for band score 5 across L1 categories*

| Dependent variable | (I) L1Category | (J) L1Category | Mean difference (I-J) | Std. error | Sig. |
|---|---|---|---|---|---|
| **Flesch Reading Ease** | 1 | 2 | -3.70030 | 3.18210 | .479 |
| | | 3 | 4.06433 | 3.09723 | .392 |
| | 2 | 1 | 3.70030 | 3.18210 | .479 |
| | | 3 | 7.76463* | 3.18210 | .044 |
| | 3 | 1 | -4.06433 | 3.09723 | .392 |
| | | 2 | -7.76463* | 3.18210 | .044 |
| **Lexical Diversity (TTR)** | 1 | 2 | .06287* | .02170 | .013 |
| | | 3 | .01307 | .02113 | .810 |
| | 2 | 1 | -.06287* | .02170 | .013 |
| | | 3 | -.04980 | .02170 | .062 |
| | 3 | 1 | -.01307 | .02113 | .810 |
| | | 2 | .04980 | .02170 | .062 |

* The mean difference is significant at the 0.05 level.

***Table 2.19: Post-hoc multiple comparisons for band score 5 across L1 categories: Tukey HSD***

Table 2.19 indicates that the European-based L1 scripts scored at band 5 were significantly different in terms of Lexical Diversity as compared to the Hindi L1 scripts scored at the same band score, a finding which was also observed for the overall scripts. Lexical Diversity was found to be 0.7 and 0.64 for scripts at band 5 level for European-based L1 and Hindi L1 test-takers respectively (see Table 2.5). The European-based L1 band 5 scripts, therefore, show a greater lexical diversity compared to Hindi L1 band 5 scripts. Greater lexical diversity has been shown to be a feature of texts produced by more proficient L2 writers, both from the data of the present study and in previous studies. Therefore, it is possible that this linguistic feature of the European-based L1 band 5 scripts could possibly have resulted in these scripts being scored higher. The same analysis was run for scripts at band score 6 with the results in the following tables.

| | | Sum of squares | df | Mean square | F | Sig. |
|---|---|---|---|---|---|---|
| **Flesch Reading Ease** | Between groups | 430.044 | 2 | 215.022 | 2.425 | .095 |
| | Within groups | 7359.848 | 83 | 88.673 | | |
| | Total | 7789.892 | 85 | | | |
| **Syntactic complexity (mean no. of words before the main verb)** | Between groups | 4.109 | 2 | 2.055 | 1.590 | .210 |
| | Within groups | 107.237 | 83 | 1.292 | | |
| | Total | 111.346 | 85 | | | |
| **Lexical Diversity (TTR)** | Between groups | .055 | 2 | .028 | 4.644 | .012 |
| | Within groups | .493 | 83 | .006 | | |
| | Total | .548 | 85 | | | |
| **Word Frequency (Celex, log)** | Between groups | .028 | 2 | .014 | .885 | .417 |
| | Within groups | 1.328 | 83 | .016 | | |
| | Total | 1.356 | 85 | | | |
| **All connectives** | Between groups | 430.940 | 2 | 215.470 | .688 | .506 |
| | Within Groups | 26007.937 | 83 | 313.349 | | |
| | Total | 26438.878 | 85 | | | |
| **Stem overlap** | Between groups | .342 | 2 | .171 | 4.492 | .014 |
| | Within groups | 3.158 | 83 | .038 | | |
| | Total | 3.500 | 85 | | | |

***Table 2.20: ANOVA for band score 6 across L1 categories***

As the results of the ANOVA for band score 6 across the three L1 categories shows, Lexical Diversity (p=0.012) and stem overlap (p=0.014) show significant difference among the scripts at this band score. To find out where these differences lie across the three L1 categories, a post-hoc test was run with the results in Table 2.21.

| Dependent variable | (I) L1Category | (J) L1Category | Mean difference (I-J) | Std. error | Sig. |
|---|---|---|---|---|---|
| **Lexical Diversity** | 1 | 2 | .06165* | .02060 | .010 |
| | | 3 | .03984 | .02006 | .122 |
| | 2 | 1 | -.06165* | .02060 | .010 |
| | | 3 | -.02180 | .02044 | .537 |
| | 3 | 1 | -.03984 | .02006 | .122 |
| | | 2 | .02180 | .02044 | .537 |
| **Stem overlap** | 1 | 2 | -.09135 | .05217 | .193 |
| | | 3 | -.15141* | .05080 | .010 |
| | 2 | 1 | .09135 | .05217 | .193 |
| | | 3 | -.06006 | .05175 | .480 |
| | 3 | 1 | .15141* | .05080 | .010 |
| | | 2 | .06006 | .05175 | .480 |

*. The mean difference is significant at the 0.05 level.

***Table 2.21: Post-hoc multiple comparisons for band score 6 across L1 categories: Tukey HSD***

Table 2.21 indicates that European-based L1 scripts scored at band 6 were significantly different in terms of Lexical Diversity as compared to Hindi L1 scripts scored at band 6. Lexical Diversity was shown to be 0.72 and 0.66 for scripts at band score 6 for European-based L1 and Hindi L1 test-takers respectively (see Table 2.5). European-based L1 band 6 scripts, therefore, show a greater lexical diversity compared to Hindi L1 band 6 scripts. Moreover, Table 2.21 shows that the European-based L1 scripts scored at band 6 were significantly different in terms of stem overlap as compared to scripts scored at the same band score from Arabic L1 candidates. The stem overlap as an index of coreferentiality is one of the indices of text cohesion. This index was 0.37 and 0.52 for European-based L1 and Arabic L1 test-takers respectively. The same analysis was run for scripts at band score 7 with the results in the following tables.

| | | Sum of squares | df | Mean square | F | Sig. |
|---|---|---|---|---|---|---|
| **Flesch Reading Ease** | Between groups | 478.696 | 2 | 239.348 | 3.51 | .034 |
| | Within groups | 5370.443 | 79 | 67.980 | | |
| | Total | 5849.139 | 81 | | | |
| **Syntactic Complexity (mean no. of words before the main verb)** | Between groups | 2.522 | 2 | 1.261 | .725 | .487 |
| | Within groups | 137.388 | 79 | 1.739 | | |
| | Total | 139.910 | 81 | | | |
| **Lexical Diversity (TTR)** | Between groups | .006 | 2 | .003 | .666 | .517 |
| | Within groups | .367 | 79 | .005 | | |
| | Total | .373 | 81 | | | |
| **Word Frequency (Celex, log)** | Between groups | .127 | 2 | .063 | 5.88 | .004 |
| | Within groups | .853 | 79 | .011 | | |
| | Total | .980 | 81 | | | |
| **All connectives** | Between groups | 482.412 | 2 | 241.206 | .855 | .429 |
| | Within groups | 22282.474 | 79 | 282.057 | | |
| | Total | 22764.887 | 81 | | | |
| **Stem overlap** | Between groups | .257 | 2 | .128 | 3.55 | .033 |
| | Within groups | 2.838 | 79 | .036 | | |
| | Total | 3.094 | 81 | | | |

***Table 2.22: ANOVA for band score 7 across L1 categories***

As Table 2.22 shows, three linguistic features significantly differentiated among the scripts rated at band 7. To find out where the differences among the three L1 categories lie, a post-hoc test was run. The results are presented in Table 2.23.

| Dependent variable | (I) L1Category | (J) L1Category | Mean difference (I-J) | Std. Error | Sig. |
|---|---|---|---|---|---|
| **Flesch Reading Ease** | 1 | 2 | 5.53230* | 2.11162 | .028 |
| | | 3 | 3.69066 | 2.34588 | .263 |
| | 2 | 1 | -5.53230* | 2.11162 | .028 |
| | | 3 | -1.84164 | 2.33025 | .710 |
| | 3 | 1 | -3.69066 | 2.34588 | .263 |
| | | 2 | 1.84164 | 2.33025 | .710 |
| **Word Frequency (Celex, Log.)** | 1 | 2 | .08997* | .02661 | .003 |
| | | 3 | .06071 | .02956 | .106 |
| | 2 | 1 | -.08997* | .02661 | .003 |
| | | 3 | -.02925 | .02937 | .581 |
| | 3 | 1 | -.06071 | .02956 | .106 |
| | | 2 | .02925 | .02937 | .581 |
| **Stem overlap** | 1 | 2 | -.09691 | .04854 | .120 |
| | | 3 | .03604 | .05392 | .783 |
| | 2 | 1 | .09691 | .04854 | .120 |
| | | 3 | .13295* | .05356 | .040 |
| | 3 | 1 | -.03604 | .05392 | .783 |
| | | 2 | -.13295* | .05356 | .040 |

* The mean difference is significant at the 0.05 level.

***Table 2.23: Post-hoc multiple comparisons for band 7 across L1 categories: Tukey HSD***

Table 2.23 indicates that European-based L1 scripts scored at band 7 were significantly different in terms of Flesch Reading Ease, as compared to Hindi L1 scripts scored at the same band score. Flesch Reading Ease was shown to be 57.04 and 51.51 for scripts at band 7 for European L1 and Hindi L1 test-takers respectively (see Table 2.5). Scripts from European-based L1 candidates, therefore, appear to be easier to read compared to scripts from Hindi L1 candidates at band 7. If Flesch Reading Ease were used as a criterion for scoring, then the Hindi L1 scripts would have been marked at a higher score compared to European-based L1 scripts at band 7.

Moreover, Table 2.23 shows that the band 7 European-based L1 scripts were significantly different in terms of Word Frequency index compared to Hindi L1 scripts scored at the same band. The word frequency index was 2.44 and 2.35 for European-based L1 and Hindi L1 band 7 scripts respectively. This means Hindi L1 test-takers used words from low frequency levels compared to European-based L1 test-takers; however, this difference was not recognised in the IELTS examiners' ratings of scripts at this band.

Another finding from Table 2.23 is that Hindi L1 and Arabic L1 scripts at band 7 were significantly different in terms of stem overlap as an index of coreferentiality and, therefore, text cohesion. The stem overlap at band 7 was 0.52 and 0.39 for Hindi L1 and Arabic L1 test-takers respectively (see Table 2.5).

## 2.5 Discussion

Seven linguistic features of the IELTS scripts scored at band levels of 5, 6, and 7 were measured quantitatively using the Coh-Metrix program. The seven features were:

- text length
- readability (Flesch Reading Ease)
- syntactic complexity (number of words before the main verb)
- Lexical Diversity (TTR)
- Word frequency (Celex, log, mean of content words)
- cohesion (all connectives)
- cohesion (stem overlap).

In this section, each research question will be addressed on the basis of the quantitative analysis of the above linguistic features of the scripts.

***Research Question 1*: What systematic differences are there in the linguistic features of scripts produced for IELTS Academic Writing Task 2 at bands 5, 6 and 7?**

Based on Table 2.3, text length was able to systematically and significantly differentiate among the three band scores. This finding is in line with that of Mayor et al. (2007) who found text length as one of the strongest predictors of high scored scripts. Moreover, Crossley and McNamara (2010) cite Ferris (1994) and Frase, Faletti, Ginther and Grant (1997) in saying that "text length has historically been a strong predictor of essay scoring with most studies reporting that text length explains about 30% of the variance in human scores" (p 6).

Descriptive statistics (see Table 2.4) of other textual features also indicate that scripts rated at higher band scores (6 and 7) were found to be more complex (using less frequent words, and having greater lexical diversity, and more syntactic complexity) than cohesive. The readability index, for example, showed that as we move from band 5 to band 7, scripts become more difficult to read.

These observations point to the fact that scripts which have received higher band scores have higher levels of linguistic complexity, but they are not necessarily more cohesive. This finding is in line with previous findings as reported in earlier sections. Our findings are particularly in line with those of Mayor et al. (2007) and Banerjee et al. (2007). Mayor et al. found sentence complexity, and Banerjee et al. found type-token ratio (lexical diversity) and word frequency (lexical sophistication) among the strongest predictors of high scores on IELTS writing tasks.

Inferential statistical analysis (based on the MANOVA and follow-up ANOVA results) showed that only two indices (readability and word frequency) were able to systematically differentiate among the scripts at band 5, 6, and 7. This finding is in line with the descriptive findings, meaning that texts rated at higher band levels show higher levels of complexity. Readability (Flesch Reading Ease) was found to be a distinctive feature of scripts rated at band scores 5 (FRE=58.34) and 7 (FRE=54.01), but not as distinctive for scripts rated at band score 6 (FRE=56.6).

The second differentiating and, perhaps, more powerful linguistic feature which was able to differentiate among the scripts at the three band scores was word frequency. Word frequency for scripts rated at band scores of 5, 6 and 7 were 2.53, 2.47, and 2.38 respectively, and the difference turned out to be significant among the three band scores. Given the range of word frequency, 0–6, and the fact that the lower the index the less frequently words are used, we can infer that the text difficulty (readability) and word frequency level of scripts are more distinctive features of scripts at these score levels than discoursal features such as index of all connectives or stem overlap.

This may be due to the fact that linguistic features, such as text complexity, are easier to assess than discoursal features, such as cohesion and coherence. Cotton and Wilson (2008), for example, investigated whether IELTS examiners find the rating of Coherence and Cohesion more difficult than the rating of the other assessment criteria for IELTS Academic Writing Task 2. Cotton and Wilson's data, from think-aloud protocols, interviews, and surveys, indicated that the majority of examiners in their study found the assessment of Coherence and Cohesion (CC) more difficult than the marking of the other three criteria of Task Response (TR), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA), and that they were less confident when marking CC. The think-aloud data in Cotton and Wilson's study showed that examiners spent more time on the assessment of CC and TR than on LR and GRA.

Moreover, it took examiners longer to read the CC band descriptors and they hesitated slightly more when assessing CC as compared to the other criteria. Moreover, variability was found among examiners in their attention to different features of the CC band descriptors, which could be attributed to the finding that a number of examiners appeared to have an incomplete understanding of some of the linguistic terms used in them. Cotton and Wilson cite Shaw and Falvey (2008) who came to the same conclusions. Although Coh-Metrix indices included in this study only partially capture the assessment criteria set by IELTS rating scales, these findings may warrant further attention, particularly given consistent findings in previous studies.

Overall, we conclude that text length, text difficulty as measured by Flesch Reading Ease, and Word Frequency as measured by Celex, log and content words significantly differentiate scripts rated at bands 5, 6, and 7.

***Research Question 2*: What systematic differences are there (if any) in the linguistic features of the scripts produced for IELTS Academic Writing Task 2 for European-based, Hindi and Arabic L1 backgrounds?**

Based on the results of the MANOVA and the follow-up ANOVA analyses, it was found that test-takers from the three L1 backgrounds (European-based, Hindi, and Arabic) produced scripts which were different in terms of text difficulty as measured by Flesch Reading Ease and Lexical Diversity (TTR). Overall, it was found that Hindi L1 IELTS test-takers produced the most consistent scripts in terms of text difficulty and lexical diversity across the three band scores (see Figures 2.1 and 2.2). Text difficulty was also found to be a feature of European-based L1 and Arabic L1 test-takers' scripts across the three band scores though with lower variations respectively.

In regard to Lexical Diversity, findings showed that this index was the same at band score 5 for European-based L1 and Arabic L1 test-takers. However, it changed diametrically for the two language groups at bands 6 and 7. Scripts of European-based L1 test-takers showed a greater lexical diversity at band 6, but a lower diversity at band score 7. This pattern was completely reversed for Arabic L1 test-takers (see Figure 2.2).

***Research Question 3*: To what extent does the impact of L1 on the linguistic features of the scripts differ at different band levels?**

To answer this research question, we rely on the final ANOVA analyses and post-hoc tests in which L1 category was used as the independent variable and six linguistic variables were used as the dependent variables for each band score. Scripts scored at band 5 were found to be significantly different for European-based L1 and Hindi L1 test-takers in terms of lexical diversity (TTR). While the mean of Lexical Diversity for European-based L1 test-takers was 0.70, this mean was 0.64 for Hindi L1 test-takers at the same band level. This means, European-based L1 test-takers produced scripts with greater lexical diversity compared to Hindi L1 test-takers at band 5. Since Lexical Resource is one of the criteria in scoring IELTS Academic Writing Task 2, and lexical diversity was found to be a distinctive feature of the three band scores overall, the significant difference between scripts rated at band 5 produced by European-based and Hindi L1 may need further attention.

On the other hand, scripts scored at band 6 were found to be different in terms of lexical diversity and cohesion (coreferentiality: stem overlap) across the three L1 categories. European-based L1 and Hindi L1 scripts at band 6 differed in terms of lexical diversity, as was also observed at band 5. The mean for lexical diversity for European-based L1 scripts at band 6 was 0.72, while it was 0.66 for Hindi L1 scripts at this band score.

Coreferentiality (stem overlap) as a representation of text cohesion was another measure which differentiated scripts at band 6 across the three L1 categories. The mean of stem overlap was 0.37 for European-based L1 scripts and 0.52 for Arabic L1 scripts.

Finally, scripts at band score 7 were found to be significantly different in terms of word frequency for European-based L1 and Hindi L1 test-takers. The mean of word frequency was 2.44 and 2.35 for European-based L1 and Hindi L1 scripts respectively, meaning that Hindi L1 test-takers used more words from low-frequency lists compared to European-based L1 test-takers at this band score. Again, since Lexical Resource is one of the scoring criteria for IELTS Writing Task 2, this finding may need further attention. Also, Hindi L1 and Arabic L1 scripts at band 7 were significantly different in terms of the cohesion index as measured by coreferentiality (stem overlap). This index was found to be 0.52 for Hindi L1 test-takers and 0.39 for Arabic L1 test-takers at band 7. Thus, on this measure, Hindi L1 test-takers produced more cohesive texts at this band score than Arabic L1 test-takers.

Table 2.24 summarises the results for RQ3.

| | Band scores | | |
|---|---|---|---|
| **Text features** | **5** | **6** | **7** |
| Flesch Reading Ease | | | Hindi (51.51) vs. European-based L1 scripts (57.04) |
| Lexical Diversity (TTR) | European-based (0.70) vs. Hindi (0.64) L1 scripts | European-based (0.72) vs. Hindi (0.66) L1 scripts | |
| Cohesion (stem overlap) | | European-based (0.37) vs. Arabic (0.52) L1 scripts | Hindi (0.52) vs. Arabic (0.39) L1 scripts |
| Word frequency | | | Hindi (2.35) vs. European-based (2.44) L1 scripts |

***Table 2.24: Summary of results for Research Question 3***

These findings may have implications for the use and interpretation of band descriptors by raters, though these indices could not completely capture the assessment criteria as defined in IELTS rating scales and as used by raters. Since lexical diversity and word frequency together constitute lexical resources, and Lexical Resource is one of the scoring criteria in IELTS Academic Writing Task 2 scoring rubric, more attention to these features when rating scripts from different L1 categories may be warranted.

As can be seen in Table 2.24, significant differences exist in lexical diversity among European-based L1 and Hindi L1 scripts at band scores 5 and 6. Additionally, significant difference in word frequency at band score 7 between Hindi L1 and European-based L1 exist. As Table 2.24 shows, significant differences were found in one of the text cohesion indices (stem overlap) at band 6 between European-based L1 and Arabic L1 scripts, and at band 7 between Hindi L1 and Arabic L1 scripts too.

The findings above are discussed further in the final section of this report where they are also considered in relation to the findings from the qualitative analysis. Conclusions and recommendations are made there.

## 3 DISCOURSE ANALYSIS OF SCRIPTS

In addition to the Computational Text Analysis of 254 scripts, a discourse analysis of a subset of 54 texts (six from each block as shown in Table 1.1) was also conducted. Texts were chosen at random from a subset of the 254 texts that conformed most closely to the 250 minimum word limit set for the IELTS Academic Writing Task 2, in order to work as far as possible with texts of approximately the same length. The discourse analysis used the analytical tools of Systemic Functional Linguistics (SFL).

SFL is a social theory of language. The basic unit of meaning is the text. Analysis at different levels of language (e.g. lexis and grammar; discourse) are conducted in order to identify patterns across whole texts, or groups of texts.

Language is understood as being systematically related to context in SFL. Context is a level of meaning, and is expressed semiotically in our material environment. That is, context is not the material environment itself, but the shared system of meanings that social groups attribute to it. One aspect of context, 'the context of culture', is theorised by many scholars working in SFL as a system of genres, or conventional patterns of social behaviour related to social purpose (Martin and Rose 2008).

For instance, the genre of 'wedding ceremony' in Western, English-speaking cultures involves conventional patterns of dress (typically but not exclusively including a white dress for the bride and (relatively) formal dress for guests), location (traditionally a church, but also outdoor locations or other significant buildings), actors (bride, groom, guests), behaviours (walking down 'the aisle', the playing of music on entrance and exit, an exchange of rings), and language (some of which is legally binding).

Applied linguists have used the notion of genre to explore patterns of meaning required for success in educational contexts, until recently focusing on language to the exclusion of other systems of meaning implicated in genres (see Bateman 2010; Kress and van Leeuwen 2001). Christie (1997) has described primary school curriculum macro-genres, or the patterns of meaning that span an entire curriculum. 'Within' these curriculum macro-genres, there are many 'smaller' genres.

More widely known work from SFL is the description of 'elemental genres' (e.g. narrative, recount, information report, discussion, exposition) which primary students are required to control in order to succeed in primary school (Martin and Rose 2012). Other SFL work has explored the genres of secondary (e.g. Coffin 2006; Veel 1997) and tertiary education (e.g. Hood 2004; Woodward-Kron 2005) which, in general terms, become more complex and more diverse in higher levels of education as might be expected.

The elemental genres common in primary schools are also found in other social spheres, because one of the main functions of primary education is to socialise children into patterns of behaviour typical of the culture. Elemental genres also often form part of longer, more complex texts found in other institutional environments, including those of tertiary education. Two of the elemental genres listed above (and sub-types of them) are common in candidate responses to Task 2 of the IELTS Academic Writing Test (Mayor et al. 2007). This is discussed at length below. The analysis of genre is discussed in Section 3.1 below.

Genre constitutes one level (or stratum) of analysis in SFL theory. Another stratum is that of discourse-semantics, or the patterns of meaning found across stretches of discourse. One area at the level of discourse semantics is the system of Appraisal, which theorises the ways in which speakers and writers evaluate the subject matter of their talk, and position themselves in relation to it, and to their audience (Martin and White 2005). This is clearly important for academic writing, and Appraisal has been applied to the study of academic writing in a range of contexts including secondary history (e.g. Coffin 2006), undergraduate essays (e.g. Woodward-Kron 2005), postgraduate research papers (Hood 2004), and Task 2 of the IELTS Academic Writing Test (Coffin and Hewings 2005). Appraisal theory is discussed and exemplified in detail in Section 3.2 below, where the Appraisal analysis of the texts is also presented.

The research questions, as discussed in earlier sections, guided the approach to analysis, which focused on the similarities and differences between the discursive resources employed in scripts in the three L1 groups (Arabic L1, Hindi L1, and European-based L1), and the three band scores (band 5, 6, and 7). Due to the small number of scripts (six from each 'block' – see Table 1.1) subject to discourse analysis in this part of the project, the first two research questions were the focus of this section of the research, and these are presented again below.

*Research Question 1*: What systematic differences are there in the linguistic features of scripts produced for IELTS Academic Writing Task 2 at bands 5, 6 and 7?

*Research Question 2*: What systematic differences are there (if any) in the linguistic features of the scripts produced for IELTS Academic Writing Task 2 for European-based, Hindi, and Arabic L1 backgrounds?

As stated above, six scripts from each block in Table 1.1 (i.e. six from the Arabic L1 Band 5 block, six from the Arabic L1 Band 6 block, and so on through all nine blocks combining L1 and band score) were analysed using the tools of SFL. A grammatical analysis of the transitivity patterns in each text was conducted. Such an analysis (of grammatical Participants, Processes, and Circumstances) forms the basis on which other clause- and discourse-level phenomena, and other broader discursive patterns can be identified (e.g. in the genre analysis).

Text structures were analysed using SFL genre theory (e.g. Martin and Rose 2008), and this is detailed below. Findings are presented in table form, and discussed group by group (e.g. Arabic L1 Band 5; Arabic L1 Band 6; Arabic L1 Band 7; Hindi L1 Band 5; and so forth). Similarities and differences between the groups according to band level (5, 6 or 7) and L1 (Arabic, Hindi or European-based) are then considered.

The use of the interpersonal resources of Appraisal (e.g. Martin and White 2005) was analysed in each of the 54 texts, and this is detailed in Section 3.2 below. Different aspects of Appraisal theory are presented in turn, and similarities and differences between the groups according to band level (5, 6 or 7) and L1 (Arabic, Hindi, or European-based) are then considered for each area of the theory.

Other areas of qualitative analysis which had been considered for inclusion in the report are not reported below due to the resources required to properly conduct and report on the genre and Appraisal analyses. The findings of the genre analysis (Section 3.1) suggest that, generally speaking, L1 is relatively unimportant as a discursive variable in the corpus, but that differences in genre at different bands are consistent with what might be expected of a valid and reliable test of writing. The findings of the Appraisal analysis (Section 3.2) suggest that, generally speaking, differences in the use of Appraisal resources between the different L1 groups appear to be relatively unimportant. There are important differences between the scripts of candidates who scored band 5 compared to those of candidates who scored band 6, and further research is warranted to explore the extent to which band score is responsible for these differences. In general, tasks (both in terms of topics and rubrics) are an important factor for the frequency and distribution of Appraisal resources in individual scripts, and there are issues worthy of further research in relation to the content validity of Task 2 of the IELTS Academic Writing Test.

## 3.1 Analysis of genre

### 3.1.1 IELTS Academic Writing Task 2 and genres

In Task 2 of the IELTS Academic Writing Test, many of the tasks take one of the following two formats:

- a statement or proposition of some kind, followed by a direction for candidates to indicate the extent to which they agree or disagree
- a statement or proposition which presents two perspectives or two opinions on a (typically social) phenomenon or situation, followed by a direction for candidates to discuss both sides and give their own opinion.

A variation on these is as follows:

- a statement or proposition of some kind, followed by a direction for candidates to consider the reasons, causes, or effects related to the statement / proposition.

This difference in task type can be expected to generate texts following (variations of) two different, but related, generic patterns (for a more detailed treatment of the genres discussed below, see Gerot and Wignell 1994; Martin and Rose 2008; for a study of IELTS Academic Writing Task 2 identifying these genres see Mayor et al. 2007). The first, known in SFL genre theory as an **exposition**, is a text pattern in which an argument or case is presented, essentially from one 'side' or perspective. Expositions typically have a structure of:

- thesis
- (preview of arguments)
- arguments
- reiteration of thesis or recommendation.

The second, known in SFL genre theory as a **discussion**, is a text pattern in which an argument or case is presented, from two or more 'sides' or perspectives. Written discussions typically have a structure of:

- issue
- (preview of arguments)
- arguments for
- arguments against
- conclusion or recommendation.

In each of these genres, a Preview is an optional stage, hence the parentheses above. These two generic patterns are compared in Table 3.1.

| Exposition | Discussion |
|---|---|
| • Thesis<br>• (Preview of arguments)<br>• Arguments<br>• Reiteration of thesis or recommendation | • Issue<br>• (Preview of arguments)<br>• Arguments for<br>• Arguments against<br>• Conclusion or recommendation |

***Table 3.1: Comparison of exposition and discussion generic patterns***

These genres differ in their social purpose, and this is realised by a different typical textual structure, or generic pattern. The main distinction, and the one we are interested in at this point, is **perspective** – whether the case presented is one-sided (as is typical of an exposition, which argues a single point of view), or multi-sided (as is typical of a discussion, which considers more than one point of view). This distinction in the social purpose of these genres is reflected in their similar, but different structures, as shown above.

Cause–Effect (and also Problem–Solution) structures fall outside this taxonomy in some respects, but for the current purpose, because they involve the author in presenting a position with argumentation, they can be included under either 'exposition' or 'discussion' according to whether the task requires the candidate to present a one-sided or multi-sided perspective on the statement/proposition in the task. So the first variable is one of perspective: single (exposition) or multiple (discussion).

Another distinction in the IELTS task types under consideration is whether they ask candidates to present an argument about whether something *is*, *is not*, or *might be* the case (termed here **analytical** – cf. Moore and Moreton's 1999 'epistemic' category of rhetorical function); or whether they ask candidates to present an argument about whether something *should* or *should not be* the case (termed here **hortatory** – see Gerot and Wignell 1994, and c.f. Moore and Moreton's 1999 'deontic' category of rhetorical function). Analytical expositions and discussions typically end with a Reiteration or Conclusion (arguing *what is*), whereas hortatory expositions and discussions typically end with a Recommendation (arguing *what should be*).

In IELTS Academic Writing Task 2, the analytical / hortatory distinction can come about in response to two factors in the task: (1) the directions to the candidate, or (2) the nature of the statement/proposition under consideration.

We first consider directions to candidates. The directions may ask a candidate, for example, whether something is a *positive or negative* development, to consider *advantages and disadvantages*, to say whether they *agree or disagree*, or to consider *reasons*, *causes or effects*.

Successful responses to these directions can expected to be **analytical** – to argue whether something *is or is not* the case and then evaluate that. In contrast, directions sometimes ask candidates to address, for example, *what should or can be done*. Successful responses to these directions can expected to be **hortatory** – to argue that something should or should not be the case and justify that.

Second, we consider the statement/proposition in the task. These typically take one of two forms:

A. a social phenomenon or issue exists (e.g. *migration is changing*; *an aspect of education is problematic*)
B. a social group should or should not do something (e.g. *governments should ...*; *individuals should not ...*).

With type A, the kind of response required (analytical or hortatory) will depend on the directions to the candidate, because the statement/proposition itself is presented as factual. But type B will usually require a hortatory response regardless of the directions, because even if the candidate is asked to agree or disagree, they are still required to argue that something should or should not be the case (rather than something is or is not the case).

Thus, we identify two clines which can be mapped together, providing a topology of task types as shown in Figure 3.1 (cf. Martin and Rose 2008, p 137).

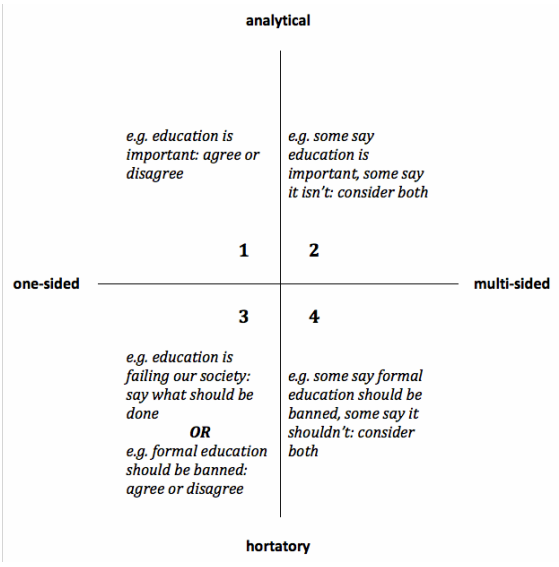



***Figure 3.1: A topology of task types in IELTS Academic Writing Task 2***

On the basis of the topology above, after analysing the generic structure of each script, we can consider the extent to which the structure is consistent with the expectations of the task. This is done by assigning numbers to each space in the topology (see Figure 3.1 above, and Figure 3.2 following).

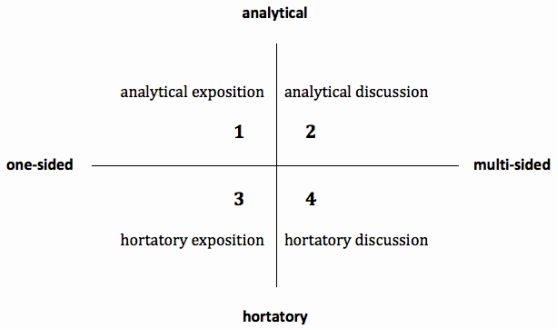



***Figure 3.2: A topology of genres relevant IELTS Academic Writing Task 2***

In some tasks, the distinction between analytical/ hortatory is unclear in the task requirements, due to the wording of the task. For example, modality of obligation (e.g. *should*, *must*) is sometimes not expressed directly in a modal auxiliary, but indirectly (using, in SFL terms, interpersonal grammatical metaphor). To illustrate, the statement/proposition *group A is not suitable for position X* can lead to responses arguing *group A is not suitable*, or *group A should not be in position X*. In such cases, either hortatory or analytical responses would match the requirements of the task.

This way of conceptualising genres draws on established theoretical work in SFL. Martin and Matthiessen (1991), and later Martin and Rose (2008) draw on work by Lemke (e.g. 1999) to oppose genre typologies and topologies. Genre typologies (which are a means of categorisation), provide distinctive 'types' of genres into which texts 'fit' (or doesn't fit, as the case may be). In contrast, a topology 'maps' the genres, and provides a way to conceptualise how some texts clearly fit into one category or another, while others may sit somewhere near or even across the boundary of two genres: so-called 'mixed texts'.

Table 3.2 on the following page lists the 54 texts analysed using SFL, the expected genre based on the topologies in Figures 3.1 and 3.2 above, and the actual generic structure of each text as identified in the analysis. So-called 'mixed texts' are identified in Table 3.2 and are shown by giving more than one number (and, where applicable, with the 'less influential' category number in parentheses). For instance, text A6-9 in Table 3.2 is numbered 2(4), meaning it mostly has the structure of an analytical discussion, but with some features of a hortatory discussion. Similarly, Text A6-110 is numbered 1(2), meaning it mostly has the structure of an analytical exposition, but with some features of an analytical discussion. The analysis conducted for this research has not gone beyond these relatively 'indelicate' topological analyses.

This approach to analysing the genre of each text allows us to compare the texts in terms of the extent to which they match the expectations of the task, and the extent to which they are conventional in their text structure. The approach taken here is that, in terms of their generic structure, the texts are categorised according to **match to task** and **typicality of generic structure**. Texts are identified as having a generic structure which is:

- in their match to task:
  - matched to task
  - partly matched to task
  - not matched to task
- in their typicality, a:
  - typical generic structure
  - variation on a typical structure
  - atypical generic structure.

This allows us to compare the texts on the basis of band score, and on the basis of candidates' L1. Before examining the data 'block by block', we illustrate the classification scheme (i.e. 'match to task' and 'typicality') with extracts from texts that fall into different areas of the scheme.

Complete texts from the data set could not be used in the final version of this report due to issues of test security, so extracts are used, and some extracts have potentially identifying sections removed (indicated by the use of ellipses). This is the case in the reporting of the genre analysis and Appraisal analysis, but in both cases, complete texts were included in the earlier version of this report which was peer reviewed.

The first text shown is Text A6-496, a response to a task requiring a hortatory discussion (Table 3.2). This text does have the typical structure of a hortatory discussion (see Table 3.1). It begins with an Issue, provides Arguments for and against which are clearly indicated in the text structure, and finally gives a Conclusion/ Recommendation which states what should be done. It is therefore analysed as **matched to task** and as having a **typical generic structure**. Extracts from this text, and its generic structure are shown in Table 3.3.

| Group | Script | Task type - Expected Genre | Actual Genre | Topic (total = 26) |
|---|---|---|---|---|
| **Arabic Band 5** | A5-498 | 4/2 | 4 | government and lifestyle |
| | A5-502 | 4/2 | 4 | government and lifestyle |
| | A5-2861 | 3 | 1 | youth and school |
| | A5-4083 | 3 | 3 | travel and prejudice |
| | A5-16163 | 1 | 3 | 'disposable' consumerism |
| | A5-16167 | 1 | 3 | 'disposable' consumerism |
| **Arabic Band 6** | A6-9 | 2 | 2(4) | air travel and environment |
| | A6-110 | 1 | 1(2) | reading versus television |
| | A6-295 | 3 | 3 | environmental responsibility |
| | A6-496 | 4/2 | 4 | government and lifestyle |
| | A6-892 | 3 | 3 | payment and professions |
| | A6-1287 | 4 | 2 | role of museums |
| **Arabic Band 7** | A7-9 | 2/1 | 1 | consumer goods |
| | A7-116 | 1 | 1(3) | reading versus television |
| | A7-2276 | 1 | 1(4) | alternative energy |
| | A7-7726 | 1 | 1 | boarding schools |
| | A7-9459 | 1/2 | 2(1) | international homogeneity |
| | A7-9464 | 1/2 | 1(2) | international homogeneity |
| **Hindi Band 5** | H5-472 | 4/2 | 3 | government and lifestyle |
| | H5-512 | 4/2 | 4 | government and lifestyle |
| | H5-2751 | 3 | 3(1) | youth and school |
| | H5-2783 | 3 | 3 | youth and school |
| | H5-2829 | 3 | 3 | youth and school |
| | H5-4675 | 2/4 | 3/1 | youth in government |
| **Hindi Band 6** | H6-544 | 4/2 | 2(4) | government and lifestyle |
| | H6-1323 | 3 | 1(3) | school subjects |
| | H6-1974 | 1 | (1) | distance education |
| | H6-2097 | 1 | 1 | distance education |
| | H6-2738 | 3 | 3 | youth and school |
| | H6-22626 | 1 | 1 | technology and progress |
| **Hindi Band 7** | H7-3884 | 1 | 2 | career mobility |
| | H7-4643 | 2/4 | 1/3 | youth in government |
| | H7-4733 | 2/4 | 4 | youth in government |
| | H7-7810 | 1 | 2/4 | boarding schools |
| | H7-18005 | 4 | 3(1/4) | charity and localism |
| | H7-18706 | 4 | 4 | charity and localism |
| **Euro Band 5** | E5-340 | 3 | 3 | environmental responsibility |
| | E5-826 | 1 | (1) | 'disposable' consumerism |
| | E5-1004 | 3/1 | 3 | teenage crime |
| | E5-1199 | 1 | 1 | 'disposable' consumerism |
| | E5-1564 | 4/(2) | 3(1) | technology and environment |
| | E5-1792 | 4 | (4) | teenagers and school |
| **Euro Band 6** | E6-99 | 2/1 | 4 | consumer goods |
| | E6-454 | 4 | 4 | teenagers and school |
| | E6-698 | 4/2 | 2 | prison versus alternatives |
| | E6-979 | 3 | 3(1) | environmental responsibility |
| | E6-1002 | 4/2 | 3(4) | technology versus education |
| | E6-1189 | 1 | 3 | 'disposable' consumerism |
| **Euro Band 7** | E7-7 | 2/1 | 2 | consumer goods |
| | E7-100 | 2/1 | 1 | consumer goods |
| | E7-440 | 3 | 3 | environmental responsibility |
| | E7-806 | 1 | 1 | reading versus television |
| | E7-1159 | 1 | 1 | reading versus television |
| | E7-1161 | 1 | 2 | temporary work |

***Table 3.2: Expected and actual genres***

| Text A6-496<br>(Typical hortatory discussion) | Stages |
|---|---|
| In order to provide ... for every person in the society some governments are ... . While, some people are against this because they want to live their lives as they want with out somone telling them what to do. | Issue |
| In this essay both sides will be discussed to determine which one is right.<br>[PARAGRAPH] | Preview |
| Every government's goal is to provide ... for it's people, even if it is against their will. Controlling ... may result in ... . For example, ... . Also, limiting the speed on the roads may ... .<br>[PARAGRAPH] | Argument for |
| On the other hand, changing ... may grauntee ..., but that doesn't mean that .... People would rather ... . Not to mention, ... . This may also ... , if they are set to one lifestyle.<br>[PARAGRAPH] | Argument against |
| In conclusion, in my opinion governments should change ... , but also allowing ... . For example, applying the rules ... . | Conclusion / Recommendation |

***Table 3.3: Extracts from a hortatory discussion which is matched to task and has a typical generic structure***

The next text to be shown has an atypical generic structure. Text A5-2861 is an analytical exposition, but the final stage of the text does not provide a Reiteration of the Thesis, but a Summary of the Arguments. Further, the task to which this text was a response required a hortatory exposition which means that this text has an **atypical generic structure**, and is **not matched to task**. Extracts from the text are shown in Table 3.4.

| Text A5-2861<br>(Atypical analytical exposition) | Stages |
|---|---|
| Young people are the future. Then people must ... . People believe that young people should ... . They do not found any ... . This essay will discuss how we can let thim to do better than they are.<br>[PARAGRAPH] | Thesis |
| Firstly, young people need to have ... . They like use ... . Teachers must take this point. For example, they can ... . Then they will ... . Because they like use intresting technology.<br>[PARAGRAPH] | Argument |
| Secondly, teachers must ... . For example, they can have a good ... , to refresh their ... . For example, ... . They can play and enjoy with other students.<br>[PARAGRAPH] | Argument |
| therdly, a lot of students can learn esialy by doing, then teachers can ... , or to find good job after high school.<br>[PARAGRAPH] | Argument |
| In conclosion, young people like technology, and use it a lot, then they can ... if teachers .... Also, they want to ... . They also, like ... . | Summary |

***Table 3.4: Extracts from an analytical exposition which is not matched to task and which has an atypical generic structure***

The next text to be shown is an analytical exposition which, topologically, is close to an analytical discussion. Text A7-9464 blends aspects of an analytical exposition (the first three Arguments which set out causes) and a discussion (with Arguments that set out 'for' and 'against' positions). This structure is, in fact, well suited to the demands of the task (see further discussion under Arabic L1 Band 7 below) and so this text is analysed as being **matched to task** and having a **variation on a typical generic structure**.

| Text A7-9464<br>(Variation on analytical exposition) | Stages |
|---|---|
| People in the past used to have ... . This features may incloude ... . This features used to be notesable when people ... . Nowadays, ... more similarities are found. | Thesis |
| In my openion, there are many causes of this ... and it incloude ... as well as ... .<br>[PARAGRAPH] | Preview |
| Firstly, globalisation plays big role in creating ... . Globalisation aims to make ... as well as .... This is the great reason that made ... .<br>[PARAGRAPH] | Argument 1 |
| Secondly, ... is also a reason to have ... . Australia is a good example to show the effect of ... . People who ..., practise similar life-style in ... .<br>[PARAGRAPH] | Argument 2 |
| Moreover, turisim make the country provide ... . For example, Dubai provides these things, that why its one of the first countries that attract turist.<br>[PARAGRAPH] | Argument 3 |
| There are many advantages for having ... . First, people will feel ... and the will not feel that they are ... . People will be able to practis their life-style in ... .<br>[PARAGRAPH] | Argument for |
| On the other hand, there are also some disadvantages for this issue. As each ... will lose .... Furthermore, new generations will not know ... . It may also creat crimes and problems.<br>[PARAGRAPH] | Argument against |
| Having a ... may be a good thing but many other thing as the disadvantegs should be counted to avoid the bad secomostances. | Conclusion |

***Table 3.5: Extracts from an analytical exposition which is matched to task and which has a variation on the typical generic structure***

The final full text to be considered here, Text E6-1189, has features of a hortatory exposition, but has an atypical structure due to the Thesis stage being insufficiently developed, and to a lack of clarity in the overall argument structure of the text (contributed to by the very poor application of paragraphing conventions). The text partly meets the requirements of the task because the Arguments and Recommendation do address the task, so this text is analysed as being **partly matched to task** and as having an **atypical generic structure**.

| Text E6-1189<br>(Atypical hortatory exposition) | Stages |
|---|---|
| The ... is a part of everyday live of people all over the world. | Thesis |
| Some evidence is to be found in the way ... in many different countries. This has been leading the ... to try to ... . The ... can produce than a higher amount of ... and with less qualified ... . People can purchase ... in every country now and at a affordable, even cheap, prize. Maybe that is one reason that people always are able to get a ... . They don't have to travel for ... and don't need to ... .<br>[PARAGRAPH] | Argument 1 |
| Another reason that many people ..., is the change of the ... . In times before industrialisation people had sometimes not even enough ..., so having something else, like ..., was very special. In present times nearly everybody can ... .<br>[PARAGRAPH] | Argument 2 |
| Additionally, most people don't have ... . Most persons are ... and don't bother ... .<br>[PARAGRAPH]<br>Sometimes the ... are even more than the ... . | Argument 3 |
| This case may cause a lot of problems now and in the future. First the ... grows bigger and bigger. For example all parts of ... are brought to ..., where people without security equipment ... . Also many of the ... that are produced ... are causing lot of damage in ... . A side effect is also that we will ... .<br>[PARAGRAPH] | Argument 4 |
| In conclusion it would be very beneficiant to ..., if they would ... and look more after ... . | Recommendation |

***Table 3.6: Extracts from a hortatory exposition which is partly matched to task and which has an atypical generic structure***

Based on the topologies discussed above, the data in Table 3.2, and the analyses on which these data draw as exemplified in Tables 3.3 to 3.6, we now explore the candidate responses block by block in more detail.

Other aspects of the candidate's writing (e.g. their control of grammar, their lexical range, their spelling and punctuation) are not considered in the following discussion. The implication is **not** that these other aspects of writing are not relevant and important, nor that genre is more important than these other aspects. It is simply that the focus of the analysis in the following sub-sections is on genre and text structure.

### 3.1.2 Genres: Arabic L1 Band 5

Turning first to the six Arabic L1 Band 5 texts, three texts in this 'block' are structured in a way that aligns with the demands of the task, and three have a structure that does not directly align with the demands of the task (as defined in terms of the discussion above).

Text A5-498, for instance, is required to respond with a hortatory discussion, and provides a text with the typical structure of this genre:

- issue
- preview
- argument for
- argument against
- position / recommendation.

Similarly, Text A5-502 is required to provide a hortatory discussion and does so. Text A5-4083 is required to provide a hortatory exposition and does so, but uses a Problem-Solution structure to form the Arguments. Nonetheless, this candidate ends the text with a Recommendation, and therefore illustrates how an atypical generic structure for a particular task type can still meet the requirements of the task.

In contrast, Text A5-2861 is required to produce a hortatory exposition and provides an analytical exposition (see Table 3.4 above for full text). In what should be a Recommendation (saying what should happen), the candidate provides a summary of the arguments in the paper, thus missing a vital part of the requirements of the task:

> *In conclosion, young people like ..., and use it a lot, then they can learn ... if teachers .... Also, they want to refresh their .... They also, like doing things and ... .*

Both Text A5-16163 and Text A5-16167 are required by the task to provide analytical expositions, and discuss the causes and effects of a 'throw-away society'. Each provides a hortatory exposition, and ends with a Recommendation. The choice to include a hortatory element in these texts is not (in itself) a problem for addressing the task, as long as the demand to address causes and effects is also met. So in this case, we have two texts with a typical genre pattern, which at first glance do not meet the demands of the task, but on closer inspection do meet them as a result of the relation between analytical and hortatory texts (i.e. a hortatory text will generally also deal with facts as required in an analytical text, but an analytical text will not necessarily include arguments about what should be). This is discussed further in following sections.

Table 3.7 shows the structure of each text in this block side-by-side, with atypical generic stages underlined.

We can map the texts according to how 'typical' they are of the identified genres discussed in Section 3.1 above (analytical exposition, analytical discussion, hortatory exposition, hortatory discussion): having a **typical generic structure**, **variation on a generic structure**, or an **atypical generic structure**. At the same time, we can map the texts according to how 'matched' they are in their overall structure to the requirements of the task: being **matched to task**, **partly matched to task**, or **not matched to task**. This allows us to visualise the data as shown in Figure 3.3.

| A5-498 | A5-502 | A5-2861 | A5-4083 | A5-16163 | A5-16167 |
|---|---|---|---|---|---|
| **Expected:** Hortatory discussion | **Expected:** Hortatory discussion | **Expected:** Hortatory exposition | **Expected:** Hortatory exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition |
| **Actual:** Hortatory discussion | **Actual:** Hortatory discussion | **Actual:** Analytical exposition | **Actual:** Hortatory exposition: problem-solution | **Actual:** Hortatory exposition | **Actual:** Hortatory exposition |
| • Issue<br>• Preview<br>• Argument for<br>• Argument against<br>• Position/ recommendation | • Issue<br>• Argument for<br>• Argument against<br>• Position/ recommendation | • Thesis<br>• Arguments<br>• <u>Summary</u> | • <u>Issue</u><br>• <u>Problem</u><br>• <u>Solution</u><br>• Recommenda-tion | • Thesis<br>• Preview<br>• Argument<br>• Recommenda-tion | • Thesis<br>• Arguments<br>• Recommenda-tion |

***Table 3.7: A comparison of the Arabic L1 Band 5 scripts in terms of generic structure (atypical generic stages are underlined)***

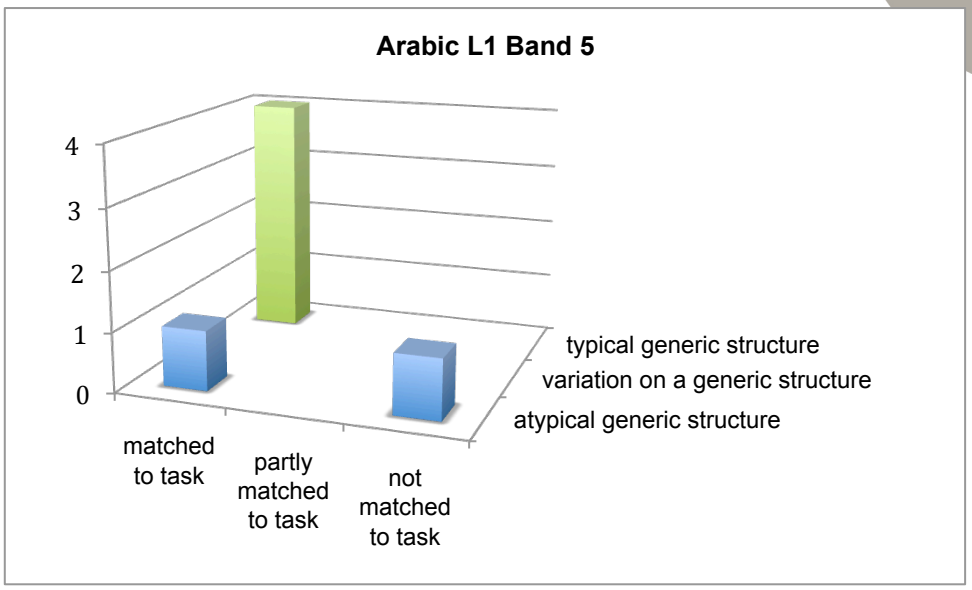



***Figure 3.3: Mapping texts according to generic structure and match to task: Arabic L1 Band 5***

### 3.1.3 Genres: Arabic L1 Band 6

The Arabic L1 Band 6 texts also vary according to the extent to which they meet the generic demands of the question. A6-496, for instance, is required to provide a hortatory discussion and does so, with the overall text structure being typical of this genre:

- issue
- preview
- argument for
- argument against
- conclusion / recommendation.

In contrast, A6-1287 is also required to write a hortatory discussion, yet produces a text closer to the structure of an analytical discussion, which provides no recommendation or even discussion about what *should* happen, but includes a Personal Response to end the text. As can be seen below, the Personal Response provides no recommendations and does not meet the demands of the task (compare the Personal Response below with the Conclusion/Recommendation from Text A6-496 shown in Table 3.3 above):

> *As for me, I love museums and I take the opportunity while being there to learn more about history and entertain my eyes looking at the magnifecnt treasures which make a link between the past time and the present time so I feel myself in another world.*

Two of the candidates produced texts which met the demands of the task, but also showed elements of a related generic structure. Illustrating with the response of A6-9, this candidate was required to write an analytical discussion and did so, but also included a final Recommendation stage: *The airways companies should reduce that to protect the world resorces.*

This addition to the typical discussion structure 'moves' the text topologically more 'towards' a hortatory discussion, though it is the only hortatory part of the text, so in the main it remains analytical. The overall structure of this text is:

- issue
- arguments for
- arguments against
- argument for
- conclusion
- recommendation.

The structure of each text in this block is shown side-by-side in Table 3.8 below, with atypical generic stages underlined. This illustrates the difference in their demonstrated ability to produce a text structured to meet the demands of the set task, with some candidates producing a text with a typical generic structure that matches the demands of the task (A6-295, A6-496), some candidates providing variation on a typical generic structure which is consistent with the demands of the task (A6-9, A6-110, A6-892), and one candidate producing a text with a generic structure that does not meet the demands of the task (A6-1287).

The table shows that five of the six texts in this 'block' meet the generic demands (or do so closely), while the last text responds to a task asking for a hortatory discussion by providing an analytical discussion that ends with a personal response to the task, rather than arguing a position on what museums should do.

As with the Arabic L1 Band 5 texts, we can map the Arabic L1 Band 6 texts according to how generically 'typical' they are, and according to how 'matched' they are to the requirements of the task as shown in Figure 3.4.

| A6-9 | A6-110 | A6-295 | A6-496 | A6-892 | A6-1287 |
|---|---|---|---|---|---|
| **Expected:** Analytical discussion | **Expected:** Analytical exposition | **Expected:** Hortatory exposition | **Expected:** Hortatory discussion | **Expected:** Hortatory exposition | **Expected:** Hortatory discussion |
| **Actual:** Analytical discussion, partly hortatory | **Actual:** Analytical exposition, partly discussion | **Actual:** Hortatory exposition | **Actual:** Hortatory discussion | **Actual:** Hortatory exposition | **Actual:** Analytical discussion |
| • Issue<br>• Arguments for<br>• Arguments against<br>• Argument for<br>• Conclusion<br>• Recommendation | • Thesis<br>• Arguments<br>• Counter-arguments<br>• Reiteration | • Thesis<br>• Arguments<br>• Recommendation | • Issue<br>• Preview<br>• Argument for<br>• Argument against<br>• Conclusion / recommendation | • Thesis<br>• Argument<br>• Recommendation<br>• Argument<br>• Recommendation<br>• Recommendation | • Issue<br>• Arguments for<br>• Arguments against<br>• Arguments for/against<br>• Personal response |

***Table 3.8: A comparison of the Arabic L1 Band 6 scripts in terms of generic structure (atypical generic stages are underlined)***

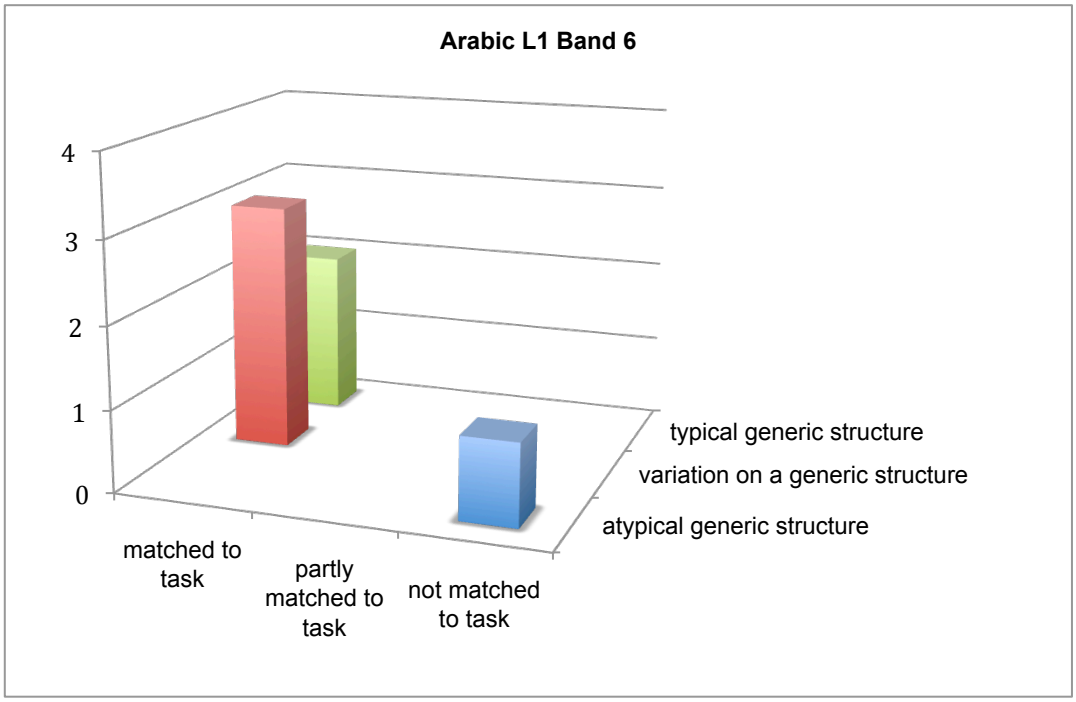



***Figure 3.4: Mapping texts according to generic structure and match to task: Arabic L1 Band 6***

### 3.1.4 Genres: Arabic L1 Band 7

Turning now to the Arabic L1 Band 7 texts, Text A7-9 responds to a task which asks the candidate to assess whether the advantages 'outweigh' the disadvantages. The genre implied here is an analytical discussion, where two perspectives on the question of cheap consumer goods are presented and compared. It would also be possible to produce an exposition structure, where the Thesis sets out a clear position favouring either the advantages or disadvantages, and where counter-arguments are dealt with as concessions to the main arguments. Text A7-9 is an analytical exposition, with a main argument that cheap prices lead to low quality. The Arguments stage also ends with a concession as follows:

> *... even though a product might seem relatively cheap and you may believe that you have made a great bargain by saving a buck, in most likelihood you are receiving the short end of the stick and cheating yourself with a low-standard, sub-par product.*

In Text A7-7762, also an analytical exposition, the Arguments are developed and clearly signposted with paragraphing conventions and the use of conjunction.

Text A7-116 is another example of a mixed text. The final stage of the text mixes a Recommendation with Reiteration of the Thesis, meeting the demands of the task and also urging action:

> *In Conclusion, it is strongly reccommended to read, even for pleasure, to improve and develop imagination and to acquire better language skills. It is well proven through science and experiment that reading is more beneficial in developing language skills and imagination than watching TV.*

Text A7-2276 is an interesting example of how a text can blend aspects of patterns from different genres. The structure of this text is:

- thesis
- arguments
- argument against
- reiteration/recommendation.

Overall, the candidate provides a clear case on a single point (that cleaner alternatives and sources of energy are positive) rather than considering different sides of this debate. In doing so, however, the candidate raises a counter-argument to strengthen their own position:

> *It is as simple as saying if we can have things available and doing the expected job, then why do we need to worry about what is being used to run it? Obviously such a transition from oil and gas to cleaner version would take years to become evident, ...*

and finally provides a Recommendation as to what should happen as part of the Reiteration. Overall, it is an analytical exposition, but also demonstrates features of a hortatory text, and of a discussion, so topologically it would be positioned in the top left quadrant of Figure 3.2, but somewhere close to the middle of the diagram (i.e. blending elements of all four topological spaces).

Text A7-9459 responds to a task requiring an analytical exposition with an analytical discussion. This text is a variation on the typical structure of a discussion, as it includes a stage which outlines the causes of increasing cultural homogeneity, thereby moving a part of the text 'outside' a two-sided argument structure. The three 'Argument' stages (Argument/Causes, Argument for, Argument against) are approximately equal in length, so this text is analysed as a discussion which is topologically close to an exposition.

In comparison, Text A7-9464 (see Table 3.5 for extracts from the full text) is a response to the same task and is topologically similar to Text A7-9459, but is analysed here as an exposition which is topologically close to a discussion, because the Arguments that look at the causes of increasing cultural homogeneity form a greater part of the text than the comparable part of Text A7-9459. Like Text A7-9459, Text A7-9464 has a Conclusion rather than a Reiteration, which also 'moves' it to the topological space with features of both exposition and discussion.

The task to which Text A7-9459 and Text A7-9464 respond asks candidates to discuss causes, and advantages/disadvantages, so texts that are variations on a typical exposition or discussion are perhaps more likely with this task than with others.

Table 3.9 shows the structure of the texts in this block. As with the other Arabic L1 texts, we can now represent the texts visually according to their 'match' with the task, and their typicality according to generic structure. This is shown in Figure 3.5.

| A7-9 | A7-116 | A7-2276 | A7-7762 | A7-9459 | A7-9464 |
|---|---|---|---|---|---|
| **Expected:** Analytical discussion or exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition |
| **Actual:** Analytical exposition | **Actual:** Analytical exposition, partly hortatory | **Actual:** Analytical exposition; partly hortatory, partly discussion | **Actual:** Analytical exposition | **Actual:** Analytical discussion; partly exposition | **Actual:** Analytical exposition; partly discussion |
| • Thesis<br>• Argument<br>• Reiteration | • Thesis<br>• Arguments<br>• <u>Recommendation/reiteration</u> | • Thesis<br>• Arguments<br>• <u>Argument against</u><br>• <u>Reiteration/recommendation</u> | • Thesis<br>• Arguments<br>• Reiteration | • Issue<br>• Preview<br>• <u>Argument / causes</u><br>• Arguments for<br>• Arguments against<br>• Conclusion | • Thesis<br>• Preview<br>• Arguments / causes<br>• Arguments for<br>• Arguments against<br>• Conclusion |

***Table 3.9: A comparison of the Arabic L1 Band 7 scripts in terms of generic structure (atypical generic stages are underlined)***

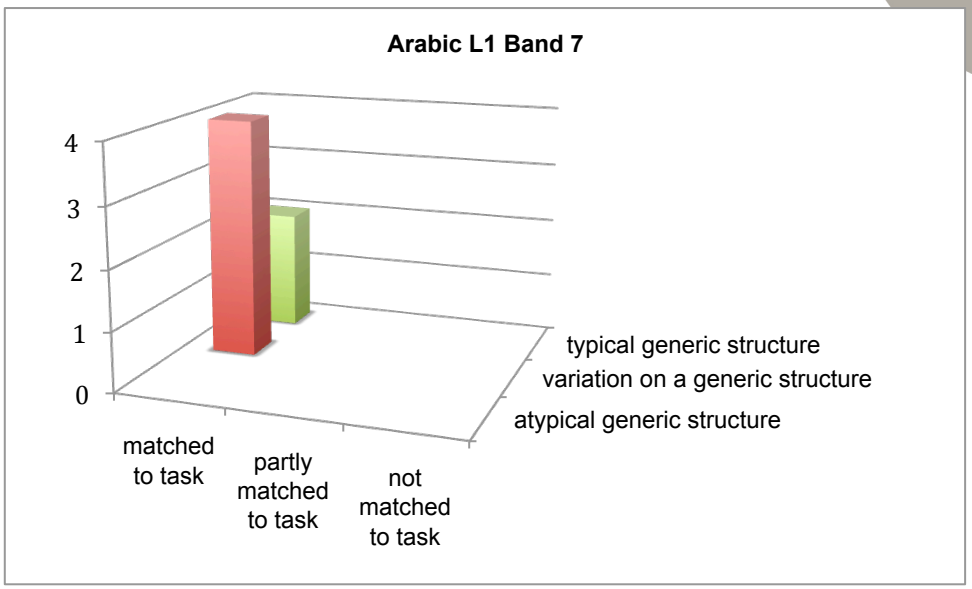



*Figure 3.5: Mapping texts according to generic structure and match to task: Arabic L1 Band 7*

### 3.1.5 Genres: Arabic L1 across the bands

We can now compare the patterns of relation between 'typicality of genre' and 'match to task' among the Arabic L1 texts across the different bands. Figure 3.6 does this visually. As the Arabic L1 texts progress up through the band scales, there is a tendency in this sub-corpus of 18 texts away from texts being not matched to the task, and also away from atypical generic patterns. Both these trends would be expected if the instrument (i.e. IELTS Academic Writing Task 2) is valid and reliable. Another pattern that can be observed is the shift towards texts which are variations on typical generic patterns (or texts which 'mix' genres) that match the task. This may be a reflection of greater expertise in manipulating the resources of language among candidates scoring in the higher bands. Expert language users do not simply 'consume' the typical patterns of a language, but use them as a creative resource. This interpretation is, of course, speculative with such a small number of texts, but does point to possible future research in this area.

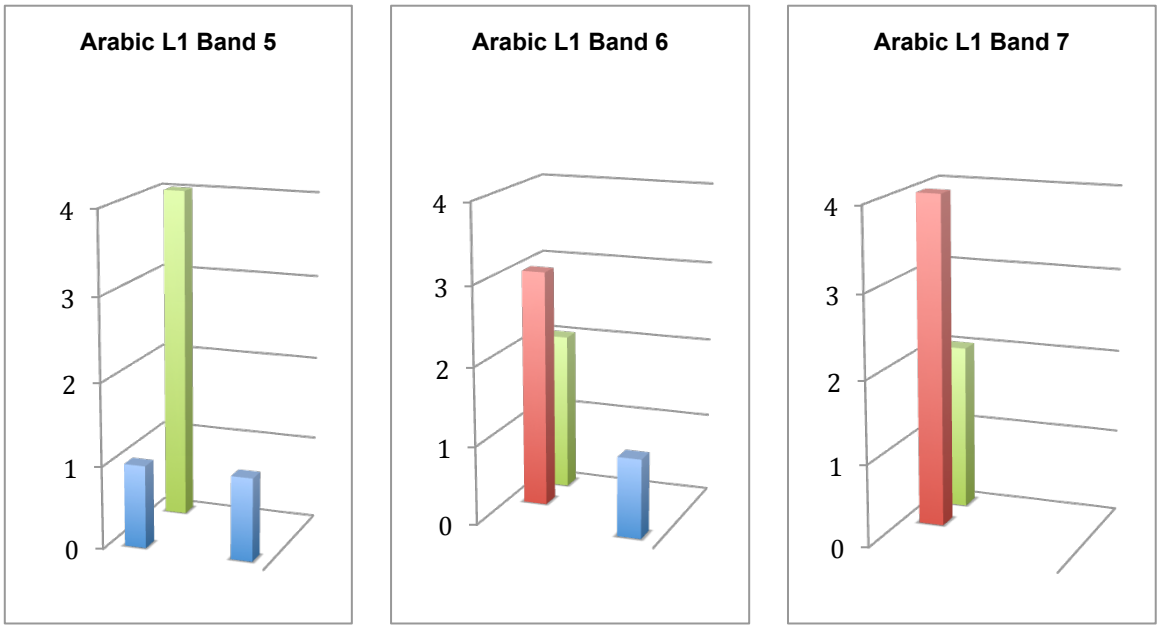



*Figure 3.6: Comparing visual mapping of texts according to generic structure and match to task: Arabic L1 across the bands*

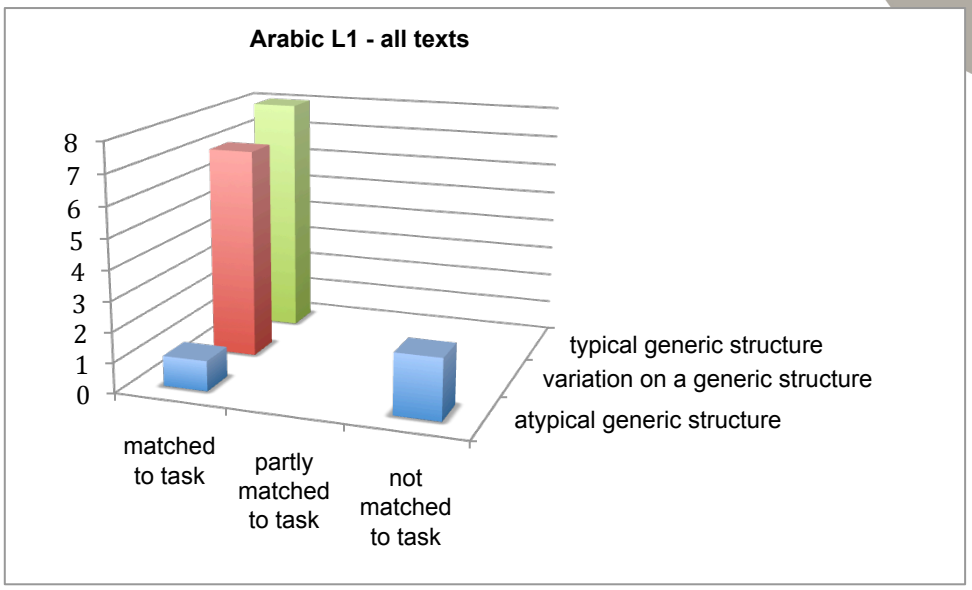



***Figure 3.7: Mapping texts according to generic structure and match to task: all Arabic L1 texts***

At this point, it is possible to look collectively at all the Arabic L1 texts, and to visualise the data according to the extent to which the text structures in this L1 group (regardless of band score) match to task, and are typical of generic patterns. This is shown in Figure 3.7, and will serve as a point of comparison, by which we will be able to compare the influence of L1 background on the use of genre, once we have examined the respective data from the Hindi L1 and European-based L1 candidates.

### 3.1.6 Genres: Hindi L1 Band 5

Turning now to the Hindi L1 group, the same analytical approach was taken to these texts, block-by-block according to band score. In the Hindi L1 Band 5 block, Text H5-472 is required by the task to produce a hortatory discussion. A key element of this task type is to consider more than one side of an argument. This text is, in fact, a straightforward hortatory exposition. At the beginning of the third Argument in the text, the candidate appears to attempt to meet the demands of the genre by offering a contrast:

> *On of the Bad point is Poverty. Government should be decreases the Number of unemployment.*

However, this Argument, like the others, discusses the things the government should do, not an opposing view (e.g. responsibility of individuals, problems with government control over individuals), and so the text structure does not meet the demands of the task.

Text H5-512 is a response to the same task. This text has some elements of an exposition. The initial stage of an exposition or discussion (Thesis or Issue) is very similar and not always easy to distinguish. In this text, the initial stage clearly presents a Thesis with the author's explicit position to be supported, rather than an Issue for investigation:

> *From my own opinion, I believe [that governments should act to ... in order to ensure they have a ... .]*

(Note: Red square brackets are where the examiner has marked part of the text as being verbatim from the task prompt.)

Rhetorically, this positions the early Arguments as part of a one-sided case, rather than a two-sided discussion, and it is not until more than half way through the text when the Arguments Against are introduced that it becomes apparent the text is a discussion and not an exposition. This text also has a Recommendation stage after each set of arguments, which is effective but is a variation on the archetypal pattern of a discussion. Despite these variations on the discussion structure, the text still matches the demands of the task as it presents a two-sided argument and is hortatory.

Text H5-2751 has the structure of a hortatory exposition, with a final Recommendation stage. However, it only partially meets the demands of the task (which does ask for a hortatory exposition) because in the earlier stages of Thesis and Preview, it sets out the response as analytical in nature (talking about what is, rather than what should be):

> *Yes, this is absolutely right that young generation is not much encouraged, they are leaving school with the negative thinking for school study. There is lot of reason that why they thinks negatively.*

This follows through in the Arguments as well. Normally, this would not be a problem (compare Text H5-2829 below), but in this text, the final Recommendation stage, which, does talk about what

should (or, more accurately, can be) is so short that the requirements of the task are barely met:

> *But we can encourage the people by introduce new professional courses.*

Thus, the text is topologically between analytical and hortatory, and because it barely addresses what should be done, it only partially matches the demands of the task.

Text H5-2783 integrates Recommendation stages throughout the text, including some recommendations as part of Arguments. In this way, it meets the demands of the task, which requires a hortatory exposition. Its structure is, however, overall problematic, as Recommendation stages are mixed with Argument stages, one Argument stage is an anecdote serving as an illustration of earlier arguments and recommendations, and the Thesis stage presumes the writer's position by referring to the question (with the definite article *the*) rather than stating the Thesis. So, in terms of the text structure (as opposed to its content), we can say it is partly matched to the task since it does present a single-sided perspective and address what should be, but its atypical structure means that the text is unlikely to meet the expectations of the reader (the examiner).

Text H5-2829 is an archetypal hortatory exposition in structure, with a long Recommendation stage that adequately addresses the demands of the task.

Text H5-4675 is a mixed text, with features of a hortatory exposition and analytical exposition. The Thesis of this text is confused, with contradictions in the polarity of the position set out. The short final stage includes a consideration of what should be done:

> *To my opinion we have to give the chance to the younger people. It is time for technological or pratical not for thinking.*

and also some attempt to reiterate the Thesis, though because of the problems with the Thesis and the fact that the Conclusion makes a different point, it has not been analysed here as a Reiteration:

> *The past time never comes back.*

The task is ambiguous in whether it requires a hortatory or analytical text, but it does require a discussion, and the student text does not provide a discussion with multiple perspectives. For this reason, the text is not matched to the task.

Table 3.10 presents the structure of the six texts in this block. Figure 3.8 represents the Hindi L1 Band 5 texts visually according to their 'match' with the task, and their typicality according to generic structure.

| H5-472 | H5-512 | H5-2751 | H5-2783 | H5-2829 | H5-4675 |
|---|---|---|---|---|---|
| **Expected:** Hortatory discussion | **Expected:** Hortatory discussion | **Expected:** Hortatory exposition | **Expected:** Hortatory exposition | **Expected:** Hortatory exposition | **Expected:** Analytical exposition |
| **Actual:** Hortatory exposition | **Actual:** Hortatory discussion, partly exposition | **Actual:** Hortatory exposition; partly analytical | **Actual:** Hortatory exposition - atypical | **Actual:** Hortatory discussion; partly exposition | **Actual:** Hortatory exposition; partly analytical |
| • Thesis<br>• Arguments<br>• Recommendation | • Thesis<br>• Arguments for<br>• Recommendation<br>• Arguments against<br>• Recommendation | • Thesis<br>• Arguments<br>• Recommendation | • Thesis<br>• Arguments<br>• Recommendations<br>• Argument<br>• Recommendation<br>• Conclusion | • Issue<br>• Preview<br>• Causes<br>• Arguments for<br>• Arguments against<br>• Conclusion | • Thesis<br>• Arguments<br>• Recommendation/conclusion |

***Table 3.10: A comparison of the Hindi L1 Band 5 scripts in terms of generic structure (atypical generic stages are underlined)***

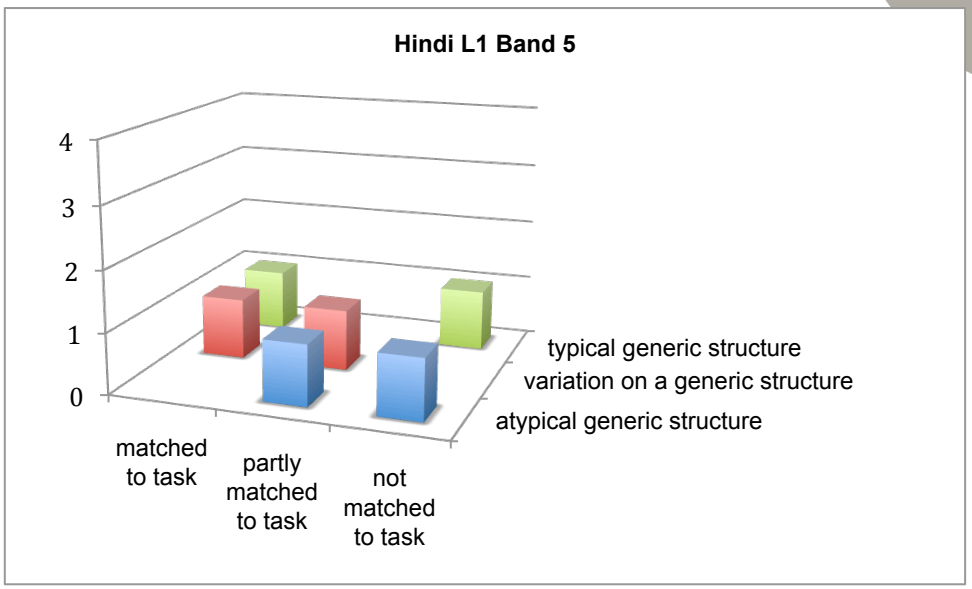



***Figure 3.8: Mapping texts according to generic structure and match to task: Hindi L1 Band 5***

### 3.1.7 Genres: Hindi L1 Band 6

Turning to the Hindi L1 Band 6 block, Text H6-544 addresses a task requiring a hortatory discussion. In this task, the instruction directs candidates to consider both sides of an issue. It is the proposition that has the hortatory elements (whether governments should or should not do something). In the 'Arguments against' stage, the candidate includes recommendations to conclude the two Arguments, both of which are shown below:

> *so it's government's duty to help people by providing free doctor services to the people so the people who are poor and cannot afford the cost of the injection can also be able to have a healthy life.*
>
> *Government are trying their best to give us a healthy life-style, so now it's up to us how we can keep it up.*

This means that the candidate addresses the requirements of the task, even though the final stage is a Reiteration of the Issue, and not a Recommendation.

Similarly, Text H6-1323 addresses a task requiring an exposition, but dealing with a proposition about what schools 'should' do. The overall structure is an analytical exposition, but there are elements of a hortatory exposition in the text, including the final stage

> *At last, I would like to conclude by disagreeing with the point that schools should concentrate on teaching ... and not on ... . I think every subject is necessary to the children in terms of their future careers.*

Here, by explicitly stating disagreement with the topic, and using a recast wording of modal obligation (i.e. an interpersonal grammatical metaphor, in SFL terms, with the wording *is necessary to* meaning *must*), the Reiteration stage also carries an implicit Recommendation, so the requirements of the task are met.

Text H6-1974 is an analytical exposition with no final stage. For this reason, the text reads as 'incomplete' to a reader expecting an argument that meets the conventions of academic contexts where English language is used. Apart from this though, the text is conventional in its structure and addresses the task, so it is analysed here as 'partly matching' the requirements of the task, and also as a 'variation' on the typical generic structure, though it is different in kind from most of the variations in the texts analysed to this point.

Texts H6-2097 and H6-22626 have the classic structure of an analytical exposition in response to a task requiring that. Text H6-2738 has the classic structure of a hortatory exposition in response to a task requiring that. Each of these texts has problematic aspects to do with genre (e.g. the clarity of the Thesis, the strength and internal structure of Arguments), but for the current purpose each can be said to be conventional, and to match the requirements of the task.

Table 3.11 shows the structure of the texts in this block. Figure 3.9 represents the texts visually according to their 'match' with the task, and their typicality according to generic structure.

| H6-544 | H6-1323 | H6-1974 | H6-2097 | H6-2738 | H6-22626 |
|---|---|---|---|---|---|
| **Expected:** Hortatory discussion | **Expected:** Hortatory exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition |
| **Actual:** Analytical discussion, partly hortatory | **Actual:** Analytical exposition, partly hortatory | **Actual:** Analytical exposition (incomplete) | **Actual:** Analytical exposition | **Actual:** Analytical discussion; partly exposition | **Actual:** Analytical exposition; partly discussion |
| • Issue<br>• Arguments for<br>• Arguments against<br>• Reiteration | • Thesis<br>• Arguments<br>• Reiteration/ recommenda-tion | • Thesis<br>• Arguments<br>• - | • Thesis<br>• Arguments<br>• Reiteration | • Thesis<br>• Arguments<br>• Recommenda tion | • Thesis<br>• Preview<br>• Arguments<br>• Arguments for<br>• Arguments against<br>• Conclusion |

***Table 3.11: A comparison of the Hindi L1 Band 6 scripts in terms of generic structure (atypical generic stages are underlined)***

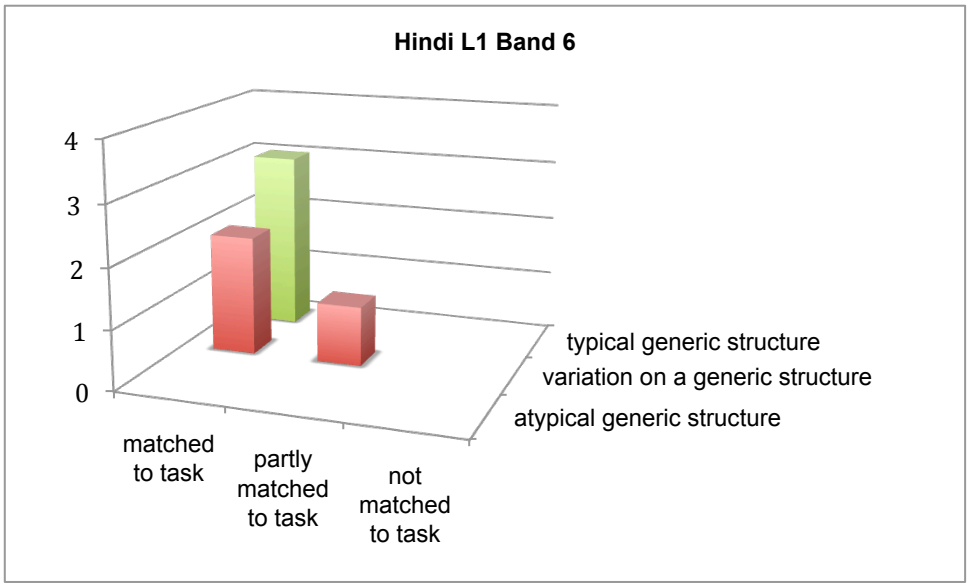



***Figure 3.9: Mapping texts according to generic structure and match to task: Hindi L1 Band 6***

### 3.1.8 Genres: Hindi L1 Band 7

Turning now to the Hindi L1 Band 7 block, Text H7-3884 has the 'classic' structure of an analytical discussion. The task asks for an analytical exposition, but this text matches the requirements because counter arguments can support a case, so a discussion meets the requirements on an exposition (see discussion in Section 3.1.15 below).

Text H7-4643 responds to a task which calls for a discussion of a proposition regarding the suitability of young people to be leaders. Questions of suitability go to whether or not someone should do (or be) something, so the wording of the proposition in this task leaves it open for a hortatory or analytical discussion. The text written by this candidate is an interesting one for (at least) two reasons.

First, it challenges the premise of the proposition in the task by arguing that age is not the deciding factor in choosing leaders, and that different ages provide different strengths. This means that the text (while presenting strengths of younger people and older people) presents a single argument, and is an exposition and not a discussion. Yet it still addresses both viewpoints (i.e. are they suitable, or not?). This text is an example of effective writing which challenges the conventions of an expected genre, using the interplay between the discursive development of ideas on one hand, and the overall structural pattern (or genre) of a text on the other, to address the demands of the task. In contrast, weaker writers typically need to rely more heavily on generic conventions to produce a successful text. Second, it includes both a Recommendation stage and a

Reiteration stage, so addresses both the hortatory and analytical expectations of the task explicitly in the structure of the text, once again by being unconventional and providing a variation on the typical generic structures of an exposition: topologically it 'rides the border' of an analytical and hortatory exposition.

Text H7-4733 is a response to the same task, and is a conventional hortatory discussion in keeping with the demands of the task. Likewise, text H6-18706 is a hortatory discussion as required by the task to which it responds.

Text H7-7810 is a text with features of a hortatory and analytical discussion, in response to a task requiring an analytical exposition. This text meets the demands of the task, however, because it considers the issue (considering more than one side of an issue can be a 'more complete' consideration of the issue), and talks about what 'should be' *in addition to* 'what is'. This is a useful illustration of a point to be taken up in Section 3.1.15 below, that tasks demanding analytical expositions allow for the greatest flexibility in the structure of an acceptable response.

Text H7-18005 is a hortatory exposition, showing some features of an analytical exposition (a final stage that is part Recommendation, part Reiteration) and some of a discussion (Arguments include concessions and the ramifications of not following the text's recommendations). So the text is a variation of the genre of hortatory exposition, but still meets the demand of the task to consider more than one side of the question.

Table 3.12 shows the structure of the texts in this block. Figure 3.10 represents the texts visually according to their 'match' with the task, and their typicality according to generic structure.

| H7-3884 | H7-4643 | H7-4733 | H7-7810 | H7-18005 | H7-18706 |
|---|---|---|---|---|---|
| **Expected:** Analytical exposition | **Expected:** Analytical or hortatory discussion | **Expected:** Analytical or hortatory discussion | **Expected:** Analytical exposition | **Expected:** Hortatory discussion | **Expected:** Hortatory discussion |
| **Actual:** Analytical discussion, | **Actual:** Analytical and hortatory exposition | **Actual:** Hortatory discussion | **Actual:** Analytical discussion, partly hortatory | **Actual:** Hortatory exposition; partly analytical | **Actual:** Hortatory discussion |
| • Issue<br>• Arguments for<br>• Arguments against<br>• Conclusion | • Thesis<br>• Arguments<br>• Recommendation<br>• Reiteration | • Issue<br>• Arguments for<br>• Arguments against<br>• Recommendation | • Issue<br>• Preview<br>• Arguments for<br>• Arguments against<br>• Conclusion/ recommenda-tion | • Thesis<br>• Arguments<br>• Recommendation/reiteration | • Issue<br>• Arguments for<br>• Arguments against<br>• Recommendation |

***Table 3.12: A comparison of the Hindi L1 Band 7 scripts in terms of generic structure (atypical generic stages are underlined)***

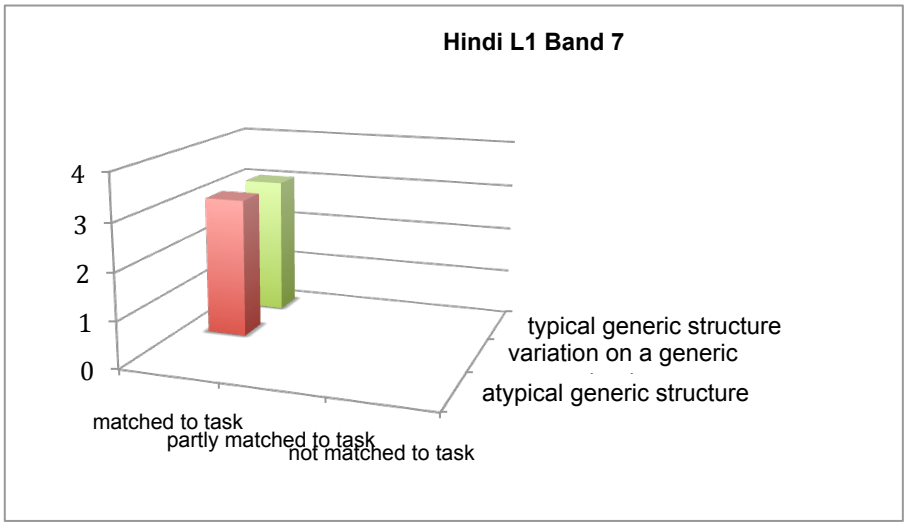



***Figure 3.10: Mapping texts according to generic structure and match to task: Hindi L1 Band 7***

### 3.1.9 Genres: Hindi L1 across the bands

We can now compare the patterns among the Hindi L1 texts across the different bands. Figure 3.11 does this visually.

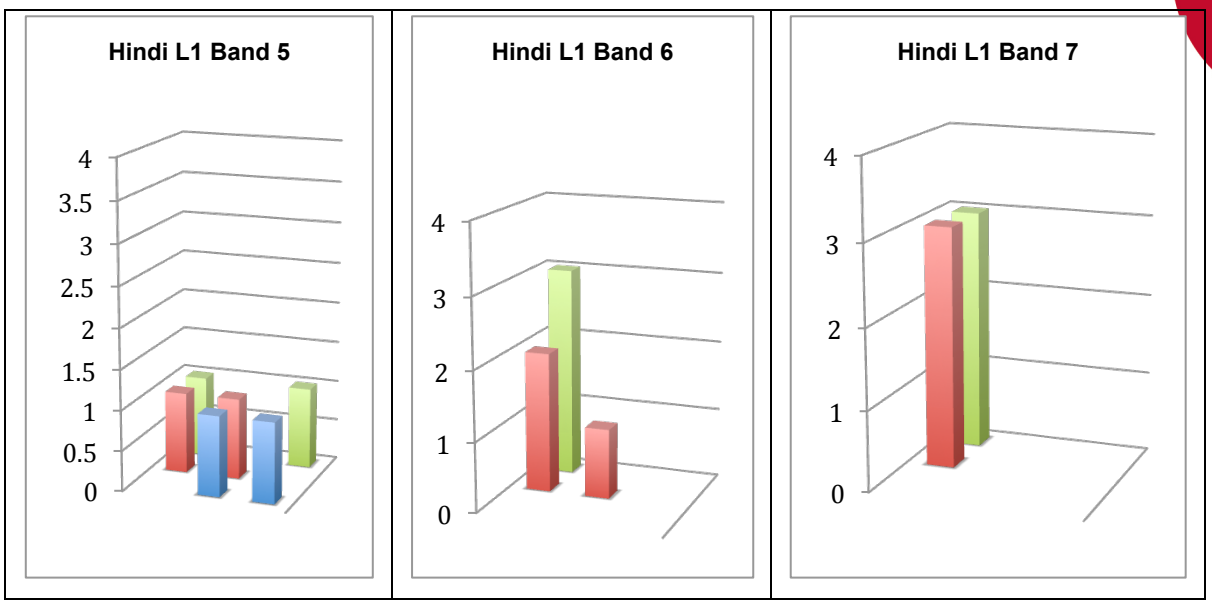



***Figure 3.11: Comparing visual mapping of texts according to generic structure and match to task: Hindi L1 across the bands***

As the Hindi L1 texts progress up through the band scales, there is a similar pattern to that found with the Arabic L1 texts: a tendency away from texts not matched to the task, and also away from atypical generic patterns. The interpretation of this pattern has been discussed above. It need not be repeated here, but is considered again below where the patterns of texts from the three L1 groups and the three bands are compared.

At this point, it is possible to look collectively at the Hindi L1 texts, and to visualise the data according to the extent to which the text structures in this L1 group (regardless of band score) match to task, and are typical of generic patterns. This is shown in Figure 3.12.

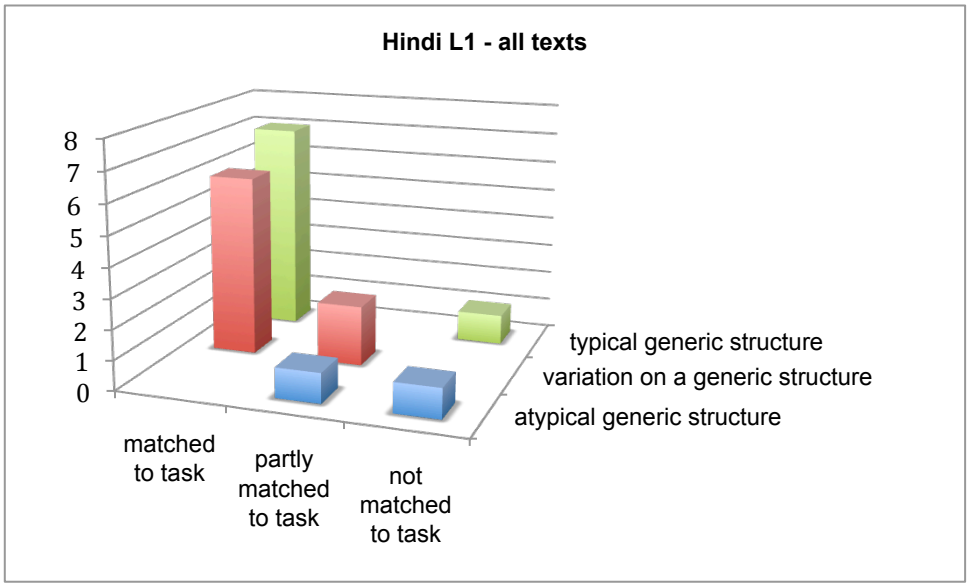



***Figure 3.12: Mapping texts according to generic structure and match to task: all Hindi L1 texts***

### 3.1.10 Genres: European-based L1 Band 5

We turn now to the European-based L1 blocks, beginning with the Band 5 Block. Text E5-340 is structured as a hortatory exposition, which is what the task requires.

Text E5-826 responds to a task requiring a discussion of causes and effects, analysed for our purposes here as requiring an analytical exposition. This text does discuss causes and effects, but there is no Reiteration of the author's Thesis in this text, so the text has an atypical structure, and is not matched to the task because the expectations of a reader in an English-speaking academic culture would not be met.

Text E5-1004 is a hortatory exposition in keeping with the requirements of the task. It is analysed here as only partly matching the demands of the task, however, because the Thesis stage is very brief, and because it has reference outside the text to the task (i.e. *this problem*) which cannot be resolved by the reader, it does not function properly. So, while the structure is a typical hortatory exposition, the initial stage is flawed to the point where the demands of the task are only partially met by the text structure.

Text E5-1199 has a typical generic structure (analytical exposition) which matches the demands of the task. This text is an interesting one. Presumably, part of the reason it is scored as a band 5 is because, it does not properly address the causes and effects of a consumerist, 'throw-away' society, as shown in this extract from the Thesis stage of the text.

> *In my opinion, fixing products is more reutable than throwing them away, because most of the time it is cheaper.*

This paper is analysed here as being 'matched' to the task, not because it addresses the task in terms of its content, but because the overall structure of the text does match the requirements of the task. This provides a useful illustration why we must look beyond genre to determine the effectiveness of a text.

Text E5-1564 is a hortatory exposition, which also includes a Reiteration stage and for that reason is topologically close to an analytical exposition. The task requires the candidate to produce a discussion, and there are indications in the text that the candidate has attempted this, most notably beginning the second Argument with the connector, *On the other hand*. However, the Thesis presents a single position, and the Arguments unfold in a Problem-Solution rather than a For-Against pattern. This text, then, is not matched to the requirements of the task as it provides only a single perspective.

Text E5-1792 is an attempt to produce a hortatory discussion, in keeping with what the task demands. The text is unsuccessful in this regard, however, as it has no Issue stage. Further, the relations between the different perspectives in the text are poorly managed. There are three different perspectives presented (teenagers should concentrate on all school subjects; they should concentrate on the one they are best at; everybody should be free to do what they want), all of which are connected with contrastive conjunctive relations (using 'however' and 'on the other hand'), which makes the logic of the discussion difficult to follow, especially given the lack of an Issue at the outset of the text.

Table 3.13 shows the structure of the texts in this block. Figure 3.13 represents the texts visually according to their 'match' with the task, and their typicality according to generic structure.

| E5-340 | E5-826 | E5-1004 | E5-1199 | E5-1564 | E5-1792 |
|---|---|---|---|---|---|
| **Expected:** Hortatory exposition | **Expected:** Analytical exposition | **Expected:** Hortatory or analytical exposition | **Expected:** Analytical exposition | **Expected:** Hortatory discussion | **Expected:** Hortatory discussion |
| **Actual:** Hortatory exposition | **Actual:** Analytical exposition (incomplete) | **Actual:** Hortatory exposition | **Actual:** Analytical discussion, partly hortatory | **Actual:** Hortatory exposition; partly analytical | **Actual:** Partial hortatory discussion |
| • Thesis<br>• Arguments<br>• Recommenda-tion | • Thesis<br>• Arguments<br>• - | • Thesis<br>• Arguments<br>• Recommenda-tion | • Issue<br>• Preview<br>• Arguments for<br>• Arguments against<br>• <u>Conclusion/ recommenda-tion</u> | • Thesis<br>• Arguments<br>• <u>Reiteration</u><br>• Recommenda-tion | • -<br>• Argument for<br>• Argument against<br>• <u>Argument against</u><br>• Recommenda-tion |

***Table 3.13: A comparison of the European-based L1 Band 5 scripts in terms of generic structure (atypical generic stages are underlined)***

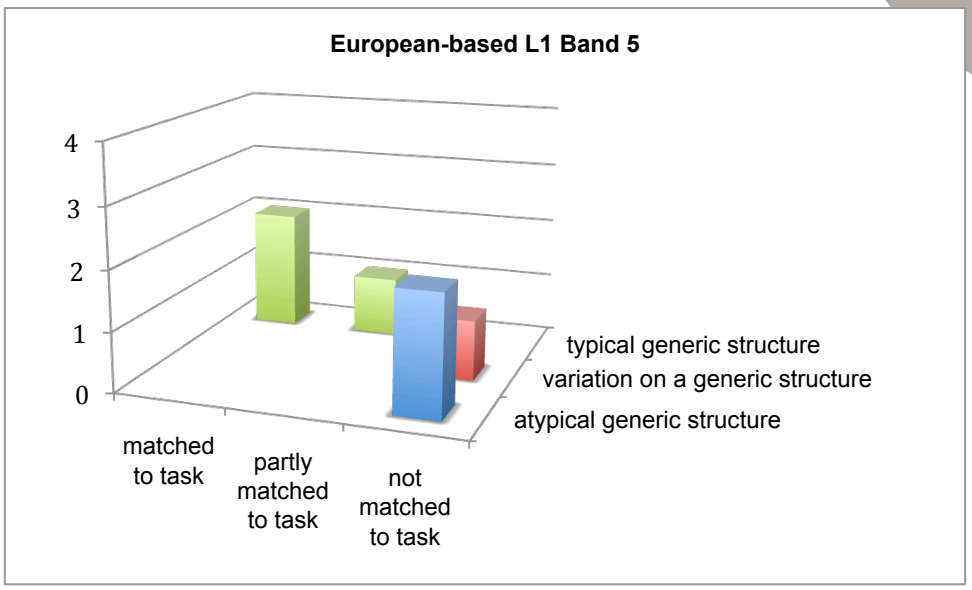



***Figure 3.13: Mapping texts according to generic structure and match to task: European-based L1 Band 5***

### 3.1.11 Genres: European-based L1 Band 6

Turning now to the European-based L1 Band 6 block, Text E6-99 is a hortatory discussion. There are some unusual aspects in the staging of this text. The Arguments Against stage does not clearly present one side of the case, and the final stage has elements of both a Conclusion and Recommendation. But because the task can effectively be met by either a discussion or exposition (see discussion of text A7-9 earlier), and because a hortatory text can often meet the demands of an analytical text (see discussion below), this text matches the demands of the task.

Text E6-454 is a discussion with both hortatory and analytical features, matching the demands of the task it addresses. The Arguments For stage is followed by a Recommendation stage where the modal obligation is expressed indirectly (*it is good for*; *is the suitable time to*).

> *It is good for teenagers to focus on all the subjects because that is the only way to extend the knowledge in all the domains of life Moreover, that age is the suitable time to establish the general basis of knowledge.*

Similar discursive strategies are used in the final Conclusion/Recommendation stage, which makes the text less hortatory than it would be if the Recommendations were expressed directly.

Text E6-698 responds to a task with a modalised proposition about what *should* happen to criminals. This is picked up in the discussion of Arguments For and Arguments Against, but does not translate into a Recommendation stage in the text structure. Rather, the final stage is a Conclusion stage where punishment/ reform practices are discussed in terms of what they *are* rather than what they *should be*. However, the text does address the requirements of the task, as two sides of the case are discussed and a clear position is expressed.

Text E6-979 is a hortatory exposition that talks about what individuals and organisations should do in relation to the environment. The final stage of the text functions both to reiterate the writer's Thesis, and to recommend what should happen, and so has features of an analytical and hortatory exposition.

Text E6-1002 is a hortatory exposition, framed around the Thesis that governments have responsibility to improve quality of life. Two ways to do this are discussed. The first is providing free education. The second Argument is introduced with a contrastive connector, *on the other hand*:

> *On the other hand new technology can also improves quality of life.*

The presence of another connector (of addition) in the same clause, also, demonstrates that even though this Argument looks at the second aspect of the task (education) and therefore meets the demands of the task, the text is structured as an exposition and not as a discussion.

Text E6-1189 is a hortatory exposition. It addresses aspects of the task well: the Arguments are structured around causes and effects of disposable consumerism, in line with the demands of the task, and the final Recommendation stage presents a Thesis based on the discussion. However, paragraphing is poorly managed in this text which makes the structure difficult to follow for the reader. Also, the initial Thesis stage of the text is not clearly relevant to the task, and does not set out a clear position:

> *The globalisation is a part of everyday live of people all over the world.*

For these reasons, the text is atypical (missing an adequate Thesis, with paragraphing that works against the rhetorical structure) and only partly addresses the demands of the task.

Table 3.14 shows the structure of the texts in this block. Figure 3.14 represents the texts visually according to their 'match' with the task, and their typicality according to generic structure.

| E6-99 | E6-454 | E6-698 | E6-979 | E6-1002 | E6-1189 |
|---|---|---|---|---|---|
| **Expected:** Analytical discussion or exposition | **Expected:** Hortatory discussion | **Expected:** Hortatory or analytical discussion | **Expected:** Analytical exposition | **Expected:** Hortatory or analytical discussion | **Expected:** Analytical exposition |
| **Actual:** Hortatory discussion | **Actual:** Hortatory discussion | **Actual:** Analytical discussion | **Actual:** Analytical discussion, partly hortatory | **Actual:** Hortatory exposition | **Actual:** Partial hortatory exposition |
| • Issue<br>• Arguments against<br>• Arguments for<br>• Conclusion/ recommenda-tion | • Issue<br>• Arguments for<br>• Recommendation<br>• Arguments against<br>• Conclusion/ recommenda-tion | • Issue<br>• Preview<br>• Arguments for<br>• Arguments against<br>• Conclusion | • Issue<br>• Preview<br>• Arguments for<br>• Arguments against<br>• Conclusion/ recommenda-tion | • Thesis<br>• Arguments<br>• Recommenda-tion | • (thesis)<br>• Arguments (causes)<br>• Arguments (effects)<br>• Recommenda-tion |

***Table 3.14: A comparison of the European-based L1 Band 6 scripts in terms of generic structure (atypical generic stages are underlined)***

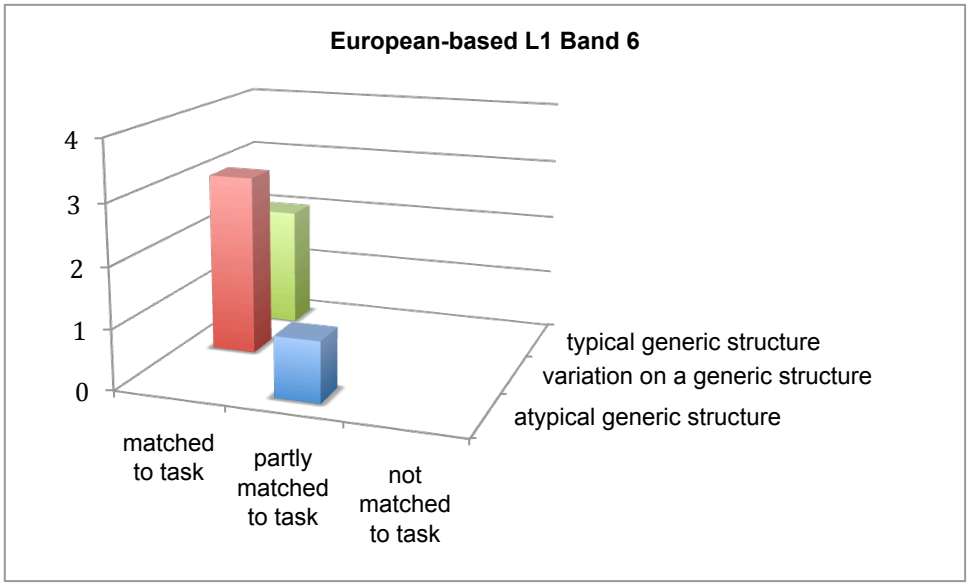



***Figure 3.14: Mapping texts according to generic structure and match to task: European-based L1 Band 6***

### 3.1.12 Genres: European-based L1 Band 7

In the European-based L1 Band 7 block, Text E7-7 and Text E7-100 are responses to the task asking whether the advantages of certain products outweigh the disadvantages (cf. Text A7-9, E6-99 above). As Text E7-100 points out:

> *The question raised, requires insight in what the potential benefits or disadvantages are of … .*

This can be addressed either by an analytical exposition (as in Text E7-100, which argues, in short, that "there are more disadvantages") or an analytical discussion (as in Text E7-7, which argues that the products in question "can have advantages and disadvantages" and explores both these facets).

Text E7-440 is a hortatory exposition in response to a task requiring a hortatory exposition.

Text E7-806 and Text E7-1159 are analytical expositions in response to the same task which requires an analytical exposition.

Text E7-1161 is an analytical discussion, but there are issues in the staging of the text. The Conclusion is not completely consistent with the author's position as stated in the Issue stage, and the structure of the 'for' and 'against' Arguments is not well managed, with very short 'for' and 'against' Arguments coming near the end of the text. For these reasons, the text is analysed here as a variation on a typical discussion, and only partly matching the requirements of the task.

Table 3.15 shows the structure of the texts in this block. Figure 3.15 represents the texts visually according to their 'match' with the task, and their typicality according to generic structure.

| E7-7 | E7-100 | E7-440 | E7-806 | E7-1159 | E7-1161 |
|---|---|---|---|---|---|
| **Expected:** Analytical discussion or exposition | **Expected:** Analytical discussion or exposition | **Expected:** Hortatory exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition | **Expected:** Analytical exposition |
| **Actual:** Analytical discussion | **Actual:** Analytical exposition | **Actual:** Hortatory exposition | **Actual:** Analytical exposition | **Actual:** Analytical exposition | **Actual:** Analytical discussion |
| • Issue<br>• Arguments against<br>• Arguments for<br>• Conclusion | • Thesis<br>• Arguments<br>• Reiteration | • Thesis<br>• Arguments<br>• Recommenda-tion | • Thesis<br>• Arguments<br>• Reiteration | • Thesis<br>• Arguments<br>• Reiteration | • Issue<br>• Arguments for<br>• Arguments against<br>• <u>Arguments for</u><br>• <u>Arguments against</u><br>• Conclusion |

***Table 3.15: A comparison of the European-based L1 Band 7 scripts in terms of generic structure (atypical generic stages are underlined)***

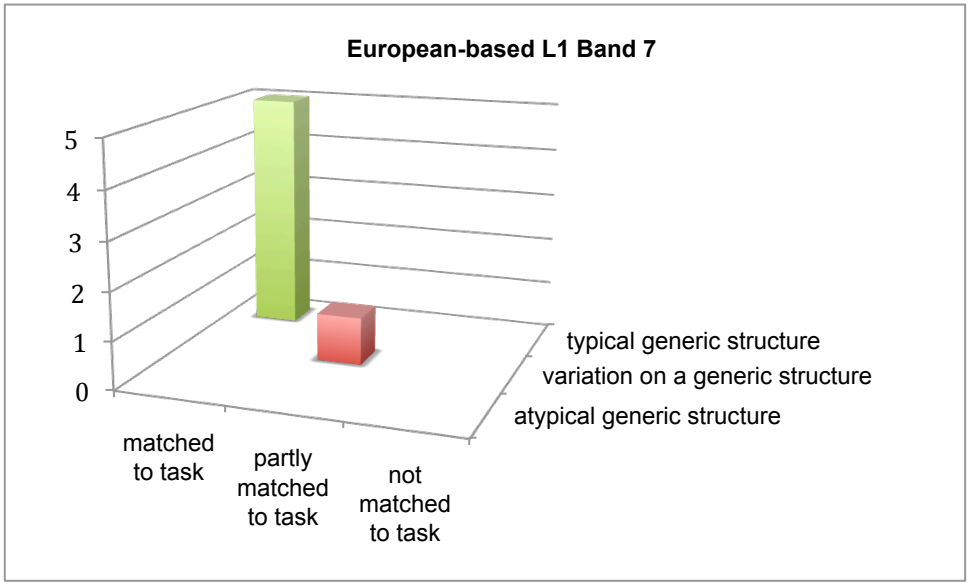



***Figure 3.15: Mapping texts according to generic structure and match to task: European-based L1 Band 7***

### 3.1.13 Genres: European-based L1 across the bands

We can now compare the patterns among the European-based L1 texts across the different bands. Figure 3.16 does this visually.

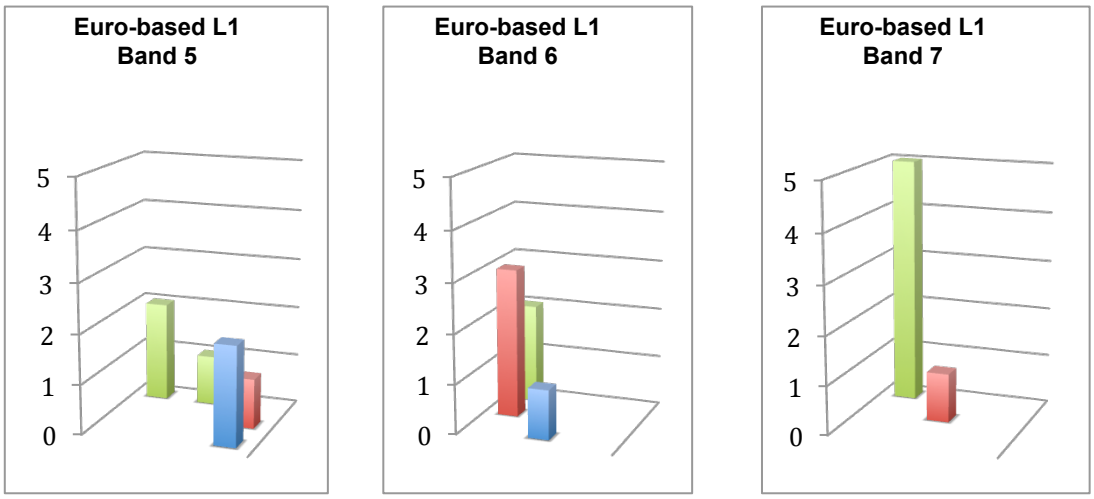



***Figure 3.16: Comparing visual mapping of texts according to generic structure and match to task: European-based L1 across the bands***

As the European-based L1 texts progress up through the band scales, there is a similar pattern to that found with the Arabic L1 and Hindi L1 texts: a tendency away from texts not matched to the task, and also away from atypical generic patterns. The interpretation of this pattern has been discussed above, and is considered again below where the patterns of texts from the three L1 groups and the three bands are compared.

At this point, it is possible to look collectively at the European-based L1 texts, and to visualise the data according to the extent to which the text structures in this L1 group (regardless of band score) match to task, and are typical of generic patterns. This is shown in Figure 3.17.

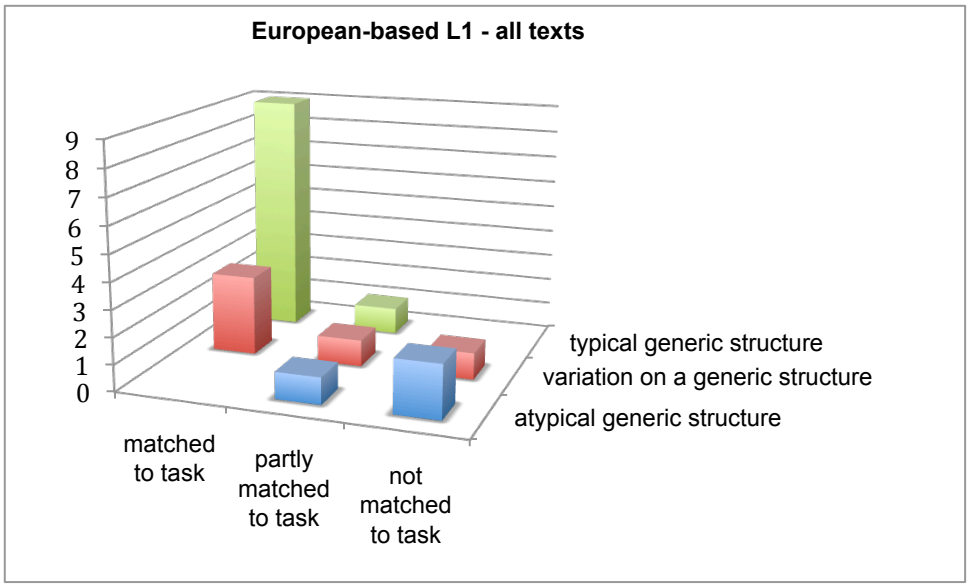



***Figure 3.17: Mapping texts according to generic structure and match to task: all European-based L1 texts***

### 3.1.14 Genres: Comparison across L1 and band score

Figure 3.18 compares the three L1 groups visually, in terms of the 'match to task' and 'typicality of genre' based on the genre analysis discussed above.

Collectively, the texts of each L1 group display a similar tendency towards being matched to the task, and having a typical generic structure, though the European-based L1 texts tend to have more texts of a typical generic structure, and the texts from Arabic L1 and Hindi L1 groups are more likely to show a variation on a typical generic structure. The Arabic L1 texts analysed here are more likely to be matched to task than the texts from the other two L1 groups. This finding suggests that there is some variation in the ways in which candidates of different L1s employ genres in their responses to IELTS Academic Writing Task 2, but the sample here is very small, and there are consistencies as well as differences in the patterns observed. Analysis of a large (statistically significant) number of texts may lead to a finding of consistency across the L1 groups.

Figure 3.19 compares visually all texts analysed above across each band score (regardless of candidate L1), in terms of the 'match to task' and 'typicality of genre'.

Figure 3.19 shows that, in the sample analysed here, the higher band score a candidate response receives, the more likely it is that the generic structure of that text will be both matched to task and of a typical generic structure. This finding is consistent with what would be expected of reliable scoring in a standardised test. Another finding that Figure 3.19 illustrates is that, even among the band 5 and 6 texts, a relatively high proportion of the texts analysed here are both matched to task, and have a typical generic structure. This suggests that there are factors other than genre involved in success on the IELTS Academic Writing Task 2, as would be expected.

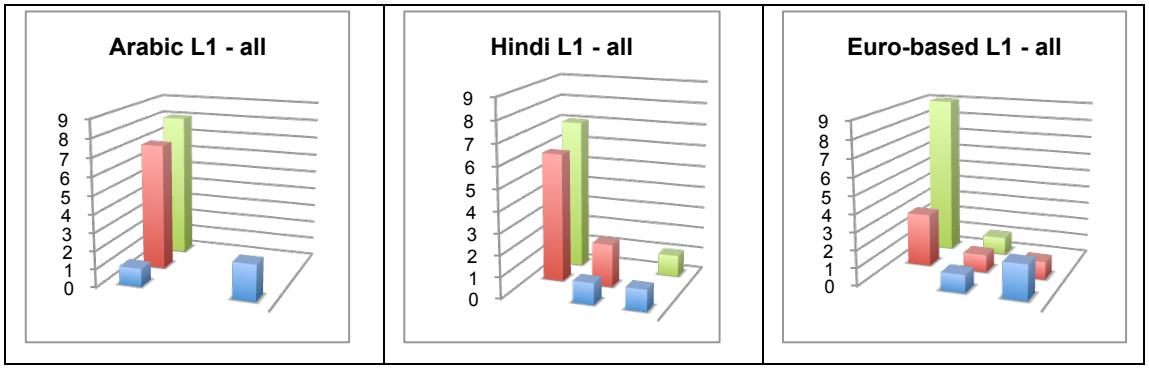



*Figure 3.18: Comparing L1 groups (regardless of band score) according to generic structure and match to task*

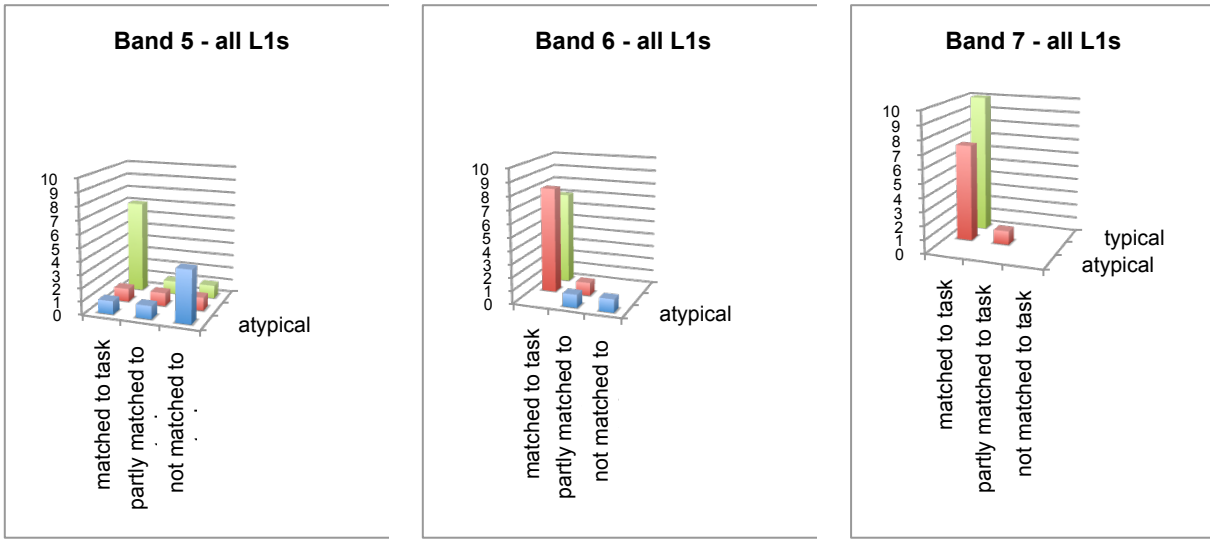



*Figure 3.19: Comparing band scores (regardless of L1 group) according to generic structure and match to task*

The final comparison to be made here is shown in Figure 3.20, which compares all nine blocks, according to both L1 and band score. In the texts from each L1 group, there is a tendency towards greater homogeneity on the variables measured in the analysis above as band score increases, with the possible exception of texts from Arabic L1 candidates which show a relatively high degree of uniformity at band 5. As noted earlier in this sub-section, the sample size in this analysis precludes generalisation. However, as Figure 3.20 illustrates, among the 54 texts analysed for this part of the study, we can conclude:

1. regardless of candidates' L1, in terms of 'match to task' and 'generic typicality' as operationalised above, texts rated as band 5 are more likely to show variability than texts rated band 6, than texts rated band 7
2. texts written by candidates of non-European-based L1s rated as band 7 or higher are more likely to show creative and effective variations of typical genes than those of texts written by candidates of European-based L1.

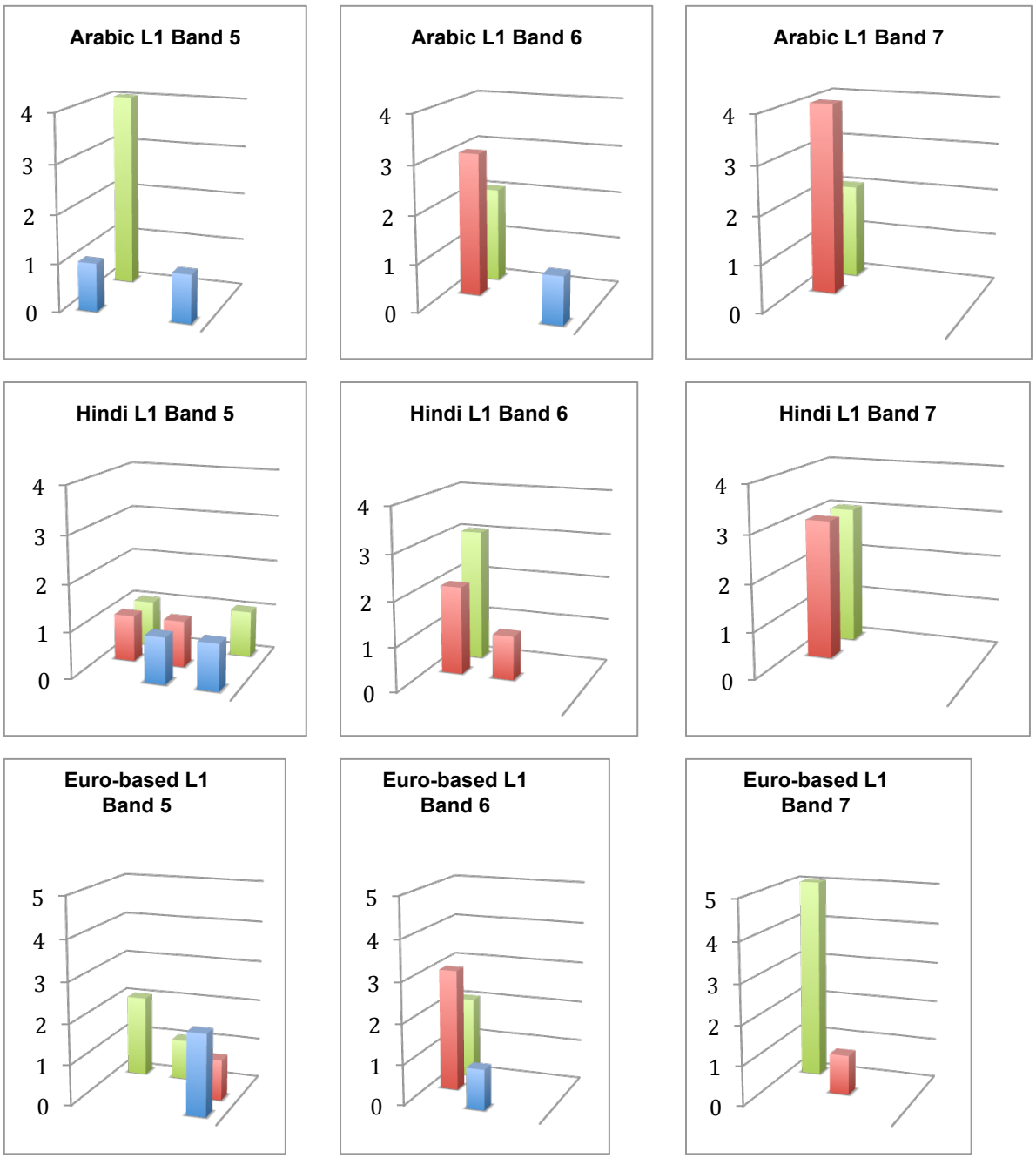



*Figure 3.20: Comparing band scores and L1 according to generic structure and match to task*

### 3.1.15 Genres: Implications and conclusions

In summary, the findings above are that, of the 54 texts analysed here from IELTS Academic Writing Task 2:

- texts written by candidates with a European-based L1 are more likely to have a typical generic structure than those written by candidates with Arabic or Hindi as L1
- texts written by candidates with Arabic as L1 are slightly more likely to be matched to task in terms of genre than candidates with Hindi as L1, which in turn are slightly more likely to be matched to task in terms of genre than candidates with a European-based L1
- the higher the band score, the more likely it is that texts will be matched to task, regardless of L1
- regarding typicality of generic structure:
  - Arabic L1 candidates appear more likely to use a typical generic structure at lower bands, and more likely to use a variation on a generic structure at higher bands
  - Hindi L1 candidates appear to be less likely to use a typical generic structure at band 5 than band 6. At bands 6 and 7, they appear likely to use a typical generic structure or a variation
  - European-based L1 candidates appear to be more likely to use a typical generic structure the higher their band score.

The fact that the higher the band score, the more likely it is that texts will be matched to task suggests that the IELTS Academic Writing Task 2 is valid (candidates' ability to structure texts effectively in relation to the set task – among other things – appears to be being measured) and the texts analysed for this part of the research have been scored reliably (though the numbers are small, the trend is consistent across L1 groups and tasks).

The analysis and discussion presented above also raises issues about the tasks for IELTS Academic Writing Task 2 in relation to genre. If candidates are required to provide an exposition, and actually provide a discussion, they will usually still meet the demands of the task (in terms of genre) because counter-arguments and multiple perspectives on an issue are acceptable even if the task only requires a single perspective. However, if candidates are required to provide a discussion, and actually provide an exposition, they do not meet the demands of the task, as providing a single perspective is insufficient when multiple perspectives are required. Thus, tasks requiring expositions, by their nature, accommodate a greater variety of responses (in terms of text structure), and candidates who are preparing for the IELTS and prepare to write a discussion regardless of the instruction are more likely to succeed on Academic Writing Task 2.

Similarly, if candidates are required to provide an analytical text (whether exposition or discussion) and actually provide a hortatory text, they are likely to meet the requirements of the task as long as they provide a clear position in an initial Thesis or Issue stage. However, if they are required to provide a hortatory response (exposition or discussion) and provide an analytical text, they will not meet the requirements of the task.

Recalling the topology from Section 3.1.1 above, the four quadrants are:

1. analytical exposition
2. analytical discussion
3. hortatory exposition
4. hortatory discussion.

The implication is that:

tasks requiring an **analytical exposition**
*are easier (to prepare for) than*
tasks requiring an **analytical discussion** or **hortatory exposition**
*are easier (to prepare for) than*
tasks requiring a **hortatory discussion**.

In the Mayor et al. (2007) study of the IELTS Academic Writing Task 2, a different type of prompt was used for the two tasks which generated the texts of their study. One task was on space exploration, the other on the pace of modern life, but both had the same generic prompt (Mayor et al. 2007, p 7):

> Present a written argument or case to an educated reader with no specialist knowledge of the following topic.
>
> [Controversial proposition]
>
> To what extent do you agree or disagree with this proposition? ...

Yet they found that candidates tended to answer in one of the four genres identified in this study, and further, that there was a general preference for expository genres over discussion genres, and hortatory texts over analytical texts. While they discuss exam preparation and advice briefly, they conclude that there is no apparent reason for the preferences displayed in their corpus on the basis of task or topic.

In the current study, genre variation can be attributed at least in part to the task, unlike in the study of Mayor et al. (2007), so in this sense the results are not immediately comparable. However, a similar conclusion can be made from both in terms of practicality and task development. It is obviously desirable to give item writers flexibility in developing tasks that have some variation, so, while a strict standardisation might be desirable in terms of reliability, it is not necessarily desirable in terms of practicality. Further, in terms of validity and washback, it is clearly desirable to have tasks requiring a variety of genres.

However, on the basis of the analysis conducted for the current study, there does appear to be an issue with reliability in task design, as a candidate who is asked to produce a hortatory discussion is clearly more constrained than one asked to provide an analytical exposition. The same candidate, on the same day, taking one or other of these two task types could score differently in terms of their response to the task as a result of the task – rather than as a result of their writing ability. In addition, the desirability of having tasks that require a variety of genres (for reasons of validity and washback) is offset by the possibility that candidates could prepare for hortatory discussions which are likely to be a successful response to most or all tasks.

The implication set out above is a hypothesis, but clearly one that is worthy of further investigation and that could be investigated relatively easily. While such an investigation is beyond the scope of the current study, an analysis of tasks, and a large-scale investigation of candidate scores on these tasks could shed light on whether the different types of task are, or are not, in fact, different in difficulty. Such an analysis could include analysis of sub-scores, comparison with scores on Academic Writing Task 1, and comparison of band scores on the IELTS tests of other skills.

Such a project might provide valuable information on relative difficulty, but also on whether there are differences in the extent to which different tasks discriminate between candidates. The kind of information this would generate would be valuable for test validation, and for test development.

## 3.2 Analysis of Appraisal

As discussed earlier, Appraisal is the study of "the interpersonal in language, ... the subjective presence of writers/speakers in texts as they adopt stances towards both the material they present and those with whom they communicate" (Martin and White 2005, p 1), which is an important area of meaning in academic writing.

In this section of the report, we examine the discourse of the 54 texts from IELTS Academic Writing Task 2 discussed above from the perspective of Appraisal theory. We set out Appraisal theory, and exemplify the different categories with examples from the 54 texts under analysis. In doing so, we present the frequencies with which the various discoursal strategies of Appraisal were used in the texts. Counts of the number of instances of each sub-type of Appraisal are given, as well as percentages as a proportion of all instances of Attitude. Due to the small number of texts involved, and the fact that evaluation is a discursive feature that builds across texts (and stages of texts) prosodically, statistical analyses were not conducted. The point of providing counts is not to determine statistical significance, but to enable a broad comparison of Appraisal features across the different groupings (candidate L1, and band score) in order to address the research questions, and to identify possible areas of discursive difference and areas which suggest themselves for further research.

After looking at the Attitude and Engagement categories and their occurrence in the 54 texts, we consider the implications of this for the IELTS Academic Writing Test.

### 3.2.1 Appraisal Theory

Appraisal theory has three basic categories (theorised as systems in SFL). These three systems are listed below, and shown in a system network in Figure 3.21.

- Engagement
- Attitude
- Graduation.

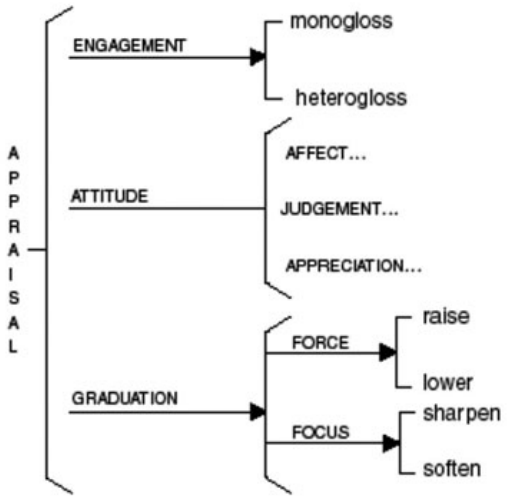



***Figure 3.21: Basic system network of Appraisal theory (source: Martin and White 2005, p 38)***

Figure 3.21 is read from left to right. Upon entering the system, three choices are made (the curved bracket means that 'Engagement' *and* 'Attitude' *and* 'Graduation' are selected). Each sub-system has its own set of more delicate choices (under 'Engagement', the square bracket means that only one of 'monogloss' *or* 'heterogloss' is chosen). In this study, the Graduation sub-system is not considered, as Attitude and Engagement were considered to be of more central importance to candidate achievement on IELTS Academic Writing Task 2 (though cf. Hood and Martin 2007).

In Section 3.2.2, the system of Attitude is discussed and exemplified, and the results of the analysis are presented. In Section 3.2.3, the system of Engagement is discussed and exemplified, and the results of the analysis are presented.

### 3.2.2 Analysis of Attitude

The system of **Attitude** is concerned with "three semantic regions covering what is traditionally referred to as emotion, ethics and aesthetics" (Martin and White 2005, p 42). Emotions are dealt with in the sub-system entitled **Affect**; ethics in the sub-system entitled **Engagement**, aesthetics in the sub-system entitled **Appreciation**.

In examining evaluation in language, explicit evaluation can be usefully distinguished from implicit evaluation (often termed 'connotation'). In Appraisal theory, this distinction is theorised in relation to Attitude, and is operationalised in terms of the presence (**inscribed Attitude**, or less technically, explicit Attitude) or absence (**invoked Attitude**, or less technically, implicit Attitude) of wording with explicit evaluative meaning (see Martin and White 2005, pp 61ff). This can be illustrated with examples that are used again later in this section:

- *sometime people can be lazy*
  (H6-544 - inscribed Attitude)
- *we don't walk enymore. 200-300 m to the shop, people take a car* (E5-340 - invoked Attitude).

The first example above uses the lexical item *lazy* to inscribe an explicit negative appraisal of people. The second example has no explicit lexis but draws on an evaluative position (which, in effective language use is built across a text) where the author and reader are assumed to be aligned around certain values. In this example, driving instead of walking 200-300 metres is seen as lazy or irresponsible. Invoked Attitude can span quite long stretches of text.

One problem for analysing evaluation in text is that invoked Attitude inevitably involves a degree of subjective judgement on the part of the analyst, as the reading position of the analyst is necessarily implicated. However, an analysis of evaluation that does not take invoked Attitude into account is likely to be such an incomplete analysis as to be virtually useless. In the counting of instances of Attitude in this report (see discussion below), no distinction has been made between inscribed or invoked Attitude.

#### 3.2.2.1 Affect

Turning now to the first sub-system of Attitude, **Affect** deals with meanings around human emotion, and has four semantic subdivisions (for more detail, see Martin and White 2005, pp 45ff on which the following discussion is based). The four subdivisions are Inclination, Happiness, Security, and Satisfaction. These are shown as a system in Figure 3.22.

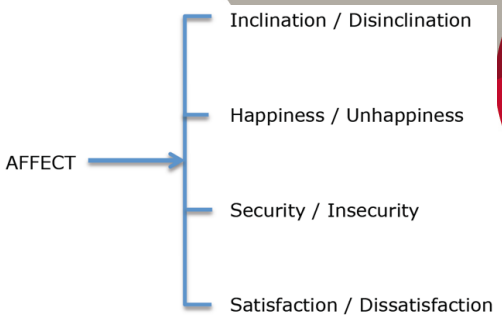



*Figure 3.22: The sub-system of Affect*

**Inclination** (which includes the negative, **Disinclination**) deals with the semantics of fears and desires. This is illustrated in the following examples from texts in the corpus.

- *He don't want to spend money to going Zym or Yoga class.* (H5-512)
- *Individual should live as they like to live.* (H5-472)
- *Some of teenagers re better in Maths, others prefere the Literature, others would like to read history books or to play a football game.* (E6-454)

The first example above illustrates Disinclination, the second Inclination, and the third example shows two instances of Inclination. Overall in the corpus of 54 texts, there are relatively few instances of Inclination. Within this small corpus, band 7 scripts use fewer instances of Inclination, but it is not possible to determine whether this trend would be significant with a larger sample or not. Table 3.16 shows the frequency of instances of Inclination as a percentage of all instances of Attitude according to L1 or band score.

| Group of blocks (L1 or band score) | Number of instances of Inclination | Instances of Inclination as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 19 | 5.2% |
| Hindi L1 | 26 | 5.6% |
| European-based L1 | 17 | 4.6% |
| Band 5 | 27 | 6.9% |
| Band 6 | 22 | 5.3% |
| Band 7 | 13 | 3.3% |

*Table 3.16: Frequency of Inclination*

The next sub-system, **Happiness** (which includes the negative, **Unhappiness**), deals with the semantics of the moods of happiness and sadness, including liking/ disliking and other related emotions. This is illustrated in the following examples from texts in the corpus.

- *Als, they will have a good smile drawn on their faces.* (A6-110)
- *Museums are places where you can look at history enjoy the smell of old civilizations and learn about ancient lives.* (A6-1287)
- *They complain about cyclons, floods and bushfires.* (E7-440)

The first two examples above illustrate expressions of Happiness, and the third Unhappiness. Overall in the corpus of 54 texts, the proportion of instances of Happiness is similar to that of Inclination. Table 3.17 shows the frequency of instances of Happiness as a percentage of all instances of Attitude according to L1 or band score. There is little difference between L1 groups or between band scores in the use of this semantic resource in the 54 texts analysed for this part of the research.

| Group of blocks (L1 or band score) | Number of instances of Happiness | Instances of Happiness as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 25 | 6.8% |
| Hindi L1 | 21 | 4.5% |
| European-based L1 | 23 | 6.3% |
| Band 5 | 25 | 6.4% |
| Band 6 | 22 | 5.3% |
| Band 7 | 22 | 5.6% |

***Table 3.17: Frequency of Happiness***

The third sub-system, **Security** (which includes the negative, **Insecurity**), deals with the semantics of feelings of calm or peacefulness, anxiety or stress, and other related emotions. This is illustrated in the following examples from texts in the corpus.

- *If they get less marks in class then they start to scared with study.* (H5-2751)
- *A young person is not afraid of taking risks and tying out new things.* (E7-4733)

The first example above illustrates Insecurity; the second Security. Overall in the corpus of 54 texts, there are fewer expression of Security than other kinds of Affect. Table 3.18 shows the frequency of instances of Security as a percentage of all instances of Attitude according to L1 or band score.

| Group of blocks (L1 or band score) | Number of instances of Security | Instances of Security as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 3 | 0.8% |
| Hindi L1 | 8 | 1.7% |
| European-based L1 | 4 | 1.1% |
| Band 5 | 5 | 1.3% |
| Band 6 | 3 | 0.7% |
| Band 7 | 7 | 1.8% |

***Table 3.18: Frequency of Security***

The next sub-system, **Satisfaction** (which includes the negative, **Dissatisfaction**), deals with the semantics of feelings of achievement, interest or pleasure, and frustration, boredom or anger, and other related emotions. This is illustrated in the following examples from texts in the corpus.

- *When he got the opportunity all of the world feel happy or proud.* (H5-4675)
- *A task in a field you like to work gives you a great satisfaction ...* (E7-1161)
- *Therefor, those teachr will be more motivated to increase their effort ...* (A6-892)

Overall in the corpus of 54 texts, there are relatively few instances of Satisfaction (more than Security but less than Inclination or Happiness). Table 3.19 shows the frequency of instances of Satisfaction as a percentage of all instances of Attitude according to L1 or band score.

| Group of blocks (L1 or band score) | Number of instances of Satisfaction | Instances of Satisfaction as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 12 | 3.3% |
| Hindi L1 | 10 | 2.1% |
| European-based L1 | 9 | 2.4% |
| Band 5 | 11 | 2.8% |
| Band 6 | 9 | 2.1% |
| Band 7 | 11 | 2.8% |

***Table 3.19: Frequency of Satisfaction***

To this point, we have looked at the sub-system of Affect (dealing with evaluations through the linguistic expression of emotion) and its four sub-divisions.

In the 54 texts analysed, expressions of Affect occurred less commonly than expressions of the other two Attitude sub-systems (i.e. Judgement and Appreciation) in all the following groupings:

- all Arabic L1 scripts, as a block
- all Hindi L1 scripts, as a block
- all European-based L1 scripts, as a block
- all band 5 scripts, as a block
- all band 6 scripts, as a block
- all band 7 scripts, as a block.

Figure 3.23 provides a visual comparison of the percentages of instances of Affect (shown in the Tables 3.16 to 3.19 above) across the three L1 groups. While the numbers are small, when viewed collectively, there is little difference in the use of Affect in the three L1 groups in the 54 texts.

Figure 3.24 provides a visual comparison of the percentages in instances of Affect (shown in the Tables 3.16 to 3.19 above) across the three band scores: band 5, band 6, and band 7. There is a slight tendency for band 5 scripts to use more instances of Affect than band 6 or band 7 scripts in the 54 texts analysed (17.3% of all instances of Attitude in band 5, 13.4% of all instances of Attitude in band 6, 13.6% of all instances of Attitude in band 7).

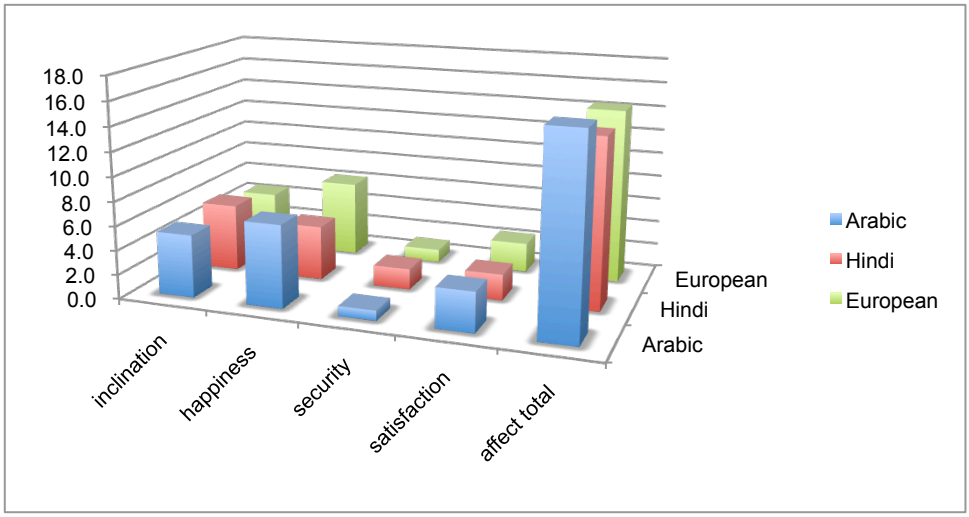



***Figure 3.23: Instances of Affect as a percentage of total instances of Attitude: Comparison across L1 groups***

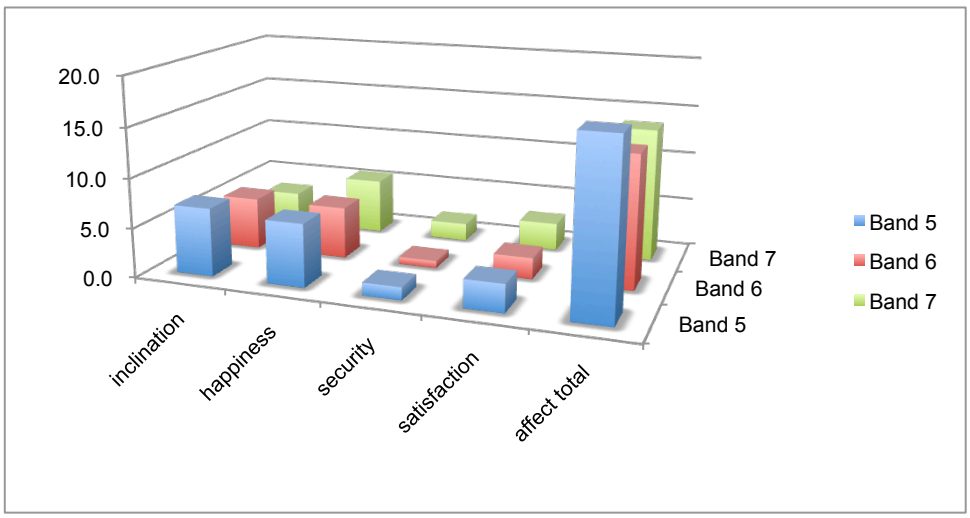



***Figure 3.24: Instances of Affect as a percentage of total instances of Attitude: Comparison across band scores***

Despite the relative homogeneity when viewed collectively, there was, as expected, variation in individual texts (where some tasks required discussion of topics where emotion was potentially implicated). An example is Text H5-2783 which discusses youth and schooling, and the attitude of students is an important part of the argument in this particular text. An extract from the full text is shown below, with the instances of Affect **underlined in bold**, and the type of Affect indicated in square brackets and italics after each instance. (Instances of Judgement and Appreciation are not shown.)

> ... So I would **like** [*Inclination*] to give you example from my own experience one of my friend **don't want** [*Disinclination*] to go to school because he was not a good learner and he always **want** [*Inclination*] stay at home at play game. Second reasion for him not going to school is his teacher was so restricted. After few days his father notice that he **afraid** [*Insecurity*] by his teacher the his father went to his school told to head master about his unand he explain all the situation about his son. So after the headmaster change that teacher and he change his study . Somehow my friend again start to going to school now **he is never late for his class** [*Happiness* - invoked]. So I saw a big change in his life. Now he always **want** [*Inclination*] to learn the new things.
>
> According to me with all these changes all young students **want** [*Inclination*] to learn new things. All young people need good teaching teacher and good invironment and play ground as well.
>
> So in the end I **want** [*Inclination*] to say ...

We can see that apart from the two instances of the author using Affect in the management of different 'voices' in the text (*I would like to give you example ...*, and *I want to say ...*; see discussion of **Engagement** below), much of this text is about students' negative emotion towards learning, and how this can be made positive. In contrast, Text E5-340 (an extract of which is shown below) has no instances of Affect, dealing as it does with the problem of pollution and the responsibilities of different parties for it, human behaviour (more likely to attract **Judgement** – see below) and institutions and abstract entities (more likely to attract **Appreciation** – see below). (Instances of Judgement and Appreciation are not shown.)

> ... In householding life we use to much electricity, gas, coal, air conditioning, we don't walk enymore. 200-300 m to the shop, people take a car. We use big engines on normal cars /3-6l/, in Japan and Europe 1-1.6-2.0l. Australians and Americans are nation wich drive truck not cars. A persson use 2 times more energy like in Europe (petrol, electricity, gas). The houses have a very poore insolations, the tecnologies for savings are not developt enaff. Today, Australia is not enymore a Green South Continent, the drought is geting biger, the forests smaller.
>
> For a better future we must save our planet, cut our uses, live more harmonious with the nature, plant more trees, build more efficient houses and cars, walk more becouse is also halfier for us and planet. ...

As these extracts illustrate, individual texts can vary widely according to the demands of the task. This suggests that the tendency of the band 5 texts to have, collectively, a slightly higher proportion of instances of Affect than band 6 or 7 texts (Figure 3.24 above) is unlikely to be solely attributable to band score. Further, the task to which candidates respond is clearly an important factor in determining which areas of Affect (if any) are likely to be used in candidate responses. The importance of task relative to L1, and in particular relative to band score, is discussed further below.

#### 3.2.2.2 Judgement

Turning away from Affect (or the semantics of human emotion), **Judgement** deals with meanings around the evaluation of human behaviour, and whether it is esteemed or sanctioned behaviour. Broadly, it is about the semantic regions of 'right and wrong'. Judgement has two primary subdivisions (**Social Esteem** and **Social Sanction**), each of which has its own sub-divisions. Here, we will consider the lowest subdivisions one by one (i.e. **Social Esteem: Normality**, **Social Esteem: Capacity**, **Social Esteem: Tenacity**, **Social Sanction: Veracity**, **Social Sanction: Propriety**). Each category includes both positive and negative evaluations within its semantic domain. The sub-system of Judgement is shown in Figure 3.25.

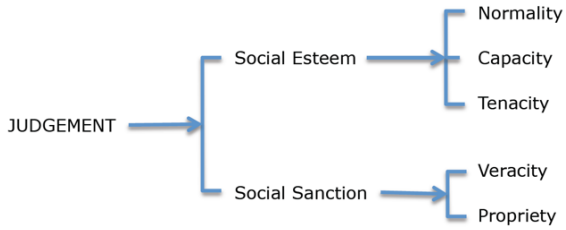



***Figure 3.25: The sub-system of Judgement***

**Social Esteem: Normality** deals with how 'usual' or 'special' someone is, as illustrated in the (respectively) negative and positive examples below.

- *... they are unable to give time to their kids, so they adopt bad habits ...* (H5-2829)
- *Most persons are specialised on some business and some skills ...* (E6-1189)

Overall in the corpus of 54 texts, there are relatively few instances of Social Esteem: Normality. The scripts of Arabic L1 candidates, and band 7 scripts each have a smaller proportion of instances of Social Esteem: Normality than their counterparts, but the numbers are too small to draw any firm conclusions about this.

Table 3.20 shows the frequency of instances of Normality as a percentage of all instances of Attitude in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Normality | Instances of Normality as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 9 | 2.5% |
| Hindi L1 | 29 | 6.2% |
| European-based L1 | 15 | 4.1% |
| Band 5 | 20 | 5.1% |
| Band 6 | 23 | 5.5% |
| Band 7 | 10 | 2.6% |

*Table 3.20: Frequency of Normality*

**Social Esteem: Capacity** deals with how 'capable' or 'skilled' someone is, and is illustrated in the following examples.

- *For example, skilled people are moving from places.* (H7-3884)
- *Younger people are suitable for important position in the government of countries.* (H5-4675)

In the corpus of 54 texts, the scripts from Hindi L1 candidates collectively have a greater proportion of instances of Social Esteem: Capacity than those of other L1 candidates. This may be due to a linguistic tendency in the writing of Hindi L1 candidates, but it is also possible that the tasks were a determining factor in this finding (as discussed above in relation to Affect). There does not appear to be any important difference in the frequency of instances of Capacity between the scripts of different band scores. Table 3.21 shows the frequency of instances of Capacity as a percentage of all instances of Attitude in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Capacity | Instances of Capacity as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 29 | 7.9% |
| Hindi L1 | 79 | 16.9% |
| European-based L1 | 24 | 6.5% |
| Band 5 | 41 | 10.5% |
| Band 6 | 48 | 11.5% |
| Band 7 | 43 | 11% |

*Table 3.21: Frequency of Capacity*

**Social Esteem: Tenacity** deals with how 'determined' or 'persistent' someone is. The two instances below are of negative Tenacity, and the second example is invoked Appraisal.

- *sometime people can be lazy* (H6-544)
- *we don't walk enymore. 200-300 m to the shop, people take a car* (E5-340)

In the corpus of 54 texts, the scripts of European-based L1 candidates collectively have a smaller proportion of instances of Tenacity than those of Arabic L1 and Hindi L1 candidates, but once again, this may be due to a discursive tendency in the writing of European-based L1 candidates, or due to a task effect. Frequency of instances of Tenacity does not appear to be related to band score, though band 7 texts had a slightly lower frequency than the other bands. Table 3.22 shows the frequency of instances of Tenacity as a percentage of all instances of Attitude in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Tenacity | Instances of Tenacity as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 19 | 5.2% |
| Hindi L1 | 34 | 7.3% |
| European-based L1 | 10 | 2.7% |
| Band 5 | 24 | 6.1% |
| Band 6 | 23 | 5.5% |
| Band 7 | 16 | 4.1% |

*Table 3.22: Frequency of Tenacity*

**Social Sanction: Veracity** deals with how 'honest' or 'deceptive' someone is. Examples of explicit lexis denoting judgements of Veracity given by Martin and White (2005, p 53) include *truthful, candid, dishonest, devious*. Surprisingly, there are no instances of Social Sanction: Veracity in the 54 texts analysed for this part of the research, as shown in Table 3.23. It is not clear whether this is a result of the particular tasks to which the candidates responded, but even with the relatively small number of texts (and smaller number of tasks), this result is unexpected. One possible explanatory factor is that potentially controversial or upsetting topics are not suitable for the IELTS test, and tasks may therefore consistently 'lead candidates away' from this semantic domain.

| Group of blocks (L1 or band score) | Number of instances of Veracity | Instances of Veracity as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 0 | 0% |
| Hindi L1 | 0 | 0% |
| European-based L1 | 0 | 0% |
| Band 5 | 0 | 0% |
| Band 6 | 0 | 0% |
| Band 7 | 0 | 0% |

*Table 3.23: Frequency of Veracity*

**Social Sanction: Propriety** deals with how 'ethical' or 'immoral' someone is, and this is illustrated in the examples below.

- *the local people try to be violent with them* (A5-4083)
- *first reason is due to the unfair behaviour of teachers, parents or friends.* (H5-2829)

In the corpus of 54 texts, the scripts of European-based L1 candidates have a higher proportion of instances of Social Sanction: Propriety than those of Arabic L1 and Hindi L1 candidates, though as discussed earlier the numbers are small and the cause of this is not clear. Similarly, band 5 scripts collectively have a higher proportion of instances of Propriety than band 6 or band 7 scripts, but it is not possible to determine whether this is a significant factor in the scoring of scripts. Table 3.24 shows the frequency of instances of Propriety as a percentage of all instances of Attitude in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Propriety | Instances of Propriety as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 52 | 14.2% |
| Hindi L1 | 52 | 11.1% |
| European-based L1 | 62 | 16.8% |
| Band 5 | 79 | 20.2% |
| Band 6 | 47 | 11.2% |
| Band 7 | 40 | 10.2% |

*Table 3.24: Frequency of Propriety*

We have now looked at the sub-system of Judgement (dealing with evaluations through the linguistic expression of Social Esteem and Social Sanction) and its five sub-divisions.

Figure 3.26 provides a visual comparison of the percentages in instances of Judgement (shown in the Tables 3.20 to 3.24 above) across the three L1 groups. The greater occurrence of the three sub-categories of Social Esteem in the Hindi L1 scripts contribute to an overall difference between Hindi L1 scripts compared to the other two groups in terms of the proportion of instances of Judgement in the scripts. It bears repeating that there are only 18 Hindi L1 scripts compared with 18 each from the other two L1 groups here, but the proportional difference (41.5% of total instances of Attitude for Hindi L1 scripts; 29.7% for Arabic L1 and 30.2% for European-based L1 scripts) suggests that Hindi L1 candidates may be more inclined to use Judgement than Arabic L1 or European-based L1 candidates. Further investigation of this hypothesis would need to make some attempt to control or account for task differences.

Figure 3.27 provides a visual comparison of the percentages in instances of Judgement (shown in the Tables 3.20 to 3.24 above) across the three band scores: band 5, band 6, and band 7. The higher number of instances of Propriety in the band 5 scripts (mentioned above) leads to an overall higher proportion of instances of Judgement than that found in band 6 or band 7 scripts (41.8% of all instances of Attitude for band 5 scripts; 33.7% for band 6 and 27.9% for band 7 scripts). This may warrant further exploration in future research: if the use of Judgement is indeed a factor in distinguishing between band 5 scripts and scripts of higher bands, implications for IELTS preparation would follow. However, the caveats expressed in the previous paragraph (regarding the number of scripts) apply also here. The higher observed proportion of Affect and Judgement in the band 5 scripts is discussed further in following sections.

In the 54 texts analysed, expressions of Judgement occurred more commonly than expressions of Affect in all L1 groups and all band scores (see Section 3.2.2.1 above).

Judgment, occurred less commonly than expressions of Appreciation (see below) in all the following groupings:

- all Arabic L1 scripts, as a block
- all Hindi L1 scripts, as a block (by a narrow margin)
- all European-based L1 scripts, as a block
- all band 6 scripts, as a block
- all band 7 scripts, as a block.

However, Judgement occurred more frequently than Appreciation in the following grouping:

- all band 5 scripts, as a block.

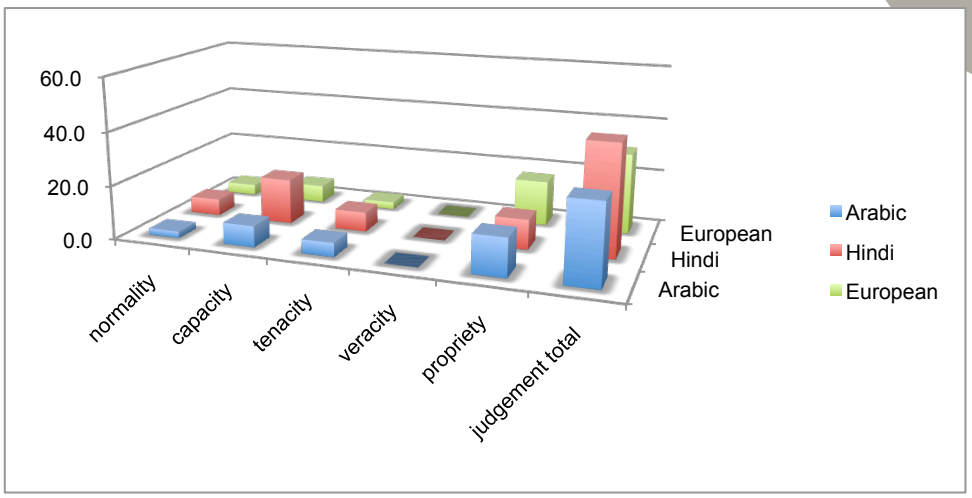



***Figure 3.26: Instances of Judgement as a percentage of total instances of Attitude: Comparison across L1 groups***

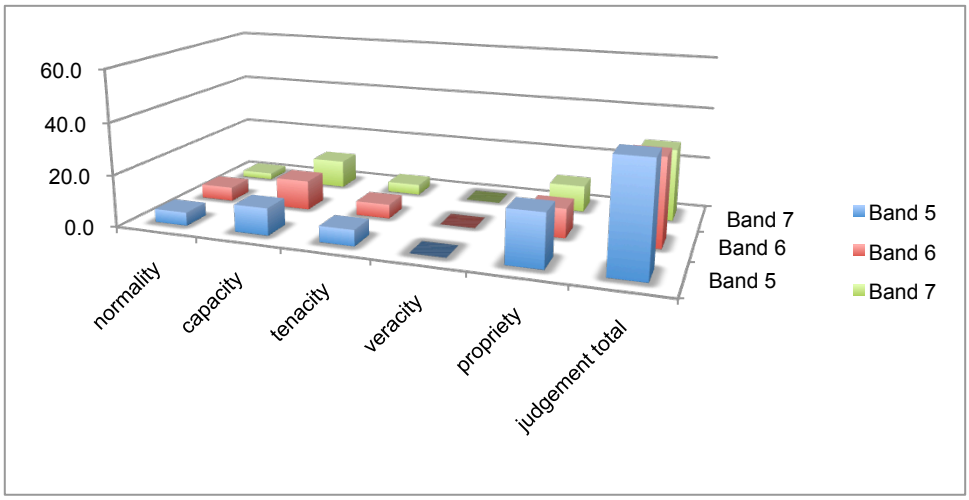



***Figure 3.27: Instances of Judgement as a percentage of total instances of Attitude: Comparison across band scores***

As with Affect, a closer examination of the patterns of Judgement in a number of individual texts is a worthwhile supplement to the broad comparison of frequency given thus far, in order to warn against 'losing' variation between individual texts in the collective view. Below, we compare Judgement in extracts from Text E5-826 and Text H7-4643.

Text E5-826 is a response to a task on 'disposable consumerism'. The candidate wrote the entire text in capital letters. An extract is shown below, with the instances of Judgement underlined in bold, and the type of Judgement indicated in square brackets and italics after each instance. The target of each instance of Judgement is also given after the Judgement type. (Instances of Affect and Appreciation are not shown.)

> PEOPLE ARE **NOT TRAINED TO DO QUALLIFIED WORK THEY DO NOT UNDERSTAND THE IMPORTANCE OF THE PRODUCT QUALITY** [*Capacity* - invoked - "PEOPLE"]. IN THE PAST WORKERS **TAKE PRAIDE ON THEM WORK. THEY PRODUCE QUALITY AND EXTECTICS PRODUCTS FOR PEOPLE TO BAY AND COME BACK TO THE SOME SHOP** [*Tenacity* - invoked - "WORKERS"]. THE SHOP KEEPERS WHERE DIFFERENT IN THE PAST TO, **THEY ANSWER THE QUESTION, THEY KNOWN THE PRODUCT** [*Capacity* - invoked - "SHOP KEEPERS"], **AND THEY GOT A SMILE ON THEM FACES NO MATTER HOW ENTRED THEY SHOP** [*Propriety* - invoked - "SHOP KEEPERS"] IN THE NEW SOCIETY **THEY NOT LOOK AT YOU AND DO NOT STOP TALKING IF A CUSTOMER COME IN** [*Propriety* - invoked - "SHOP KEEPERS ... IN THE NEW SOCIETY]"].

Clearly there are many problems with the spelling of basic lexis, and also with basic grammar and punctuation in this text. Beyond this, there are also shortcomings in the use of Judgement, and these are perhaps best discussed in comparison with the use of Judgement in another text. Text H7-4643 (below) is a response to a task on 'youth in government'.

> Younger people are **enthusiastic, energetic, courageous** [*Tenacity* - "younger people"] to take oppurtunities. They have enough **zeal** [*Tenacity* - "younger people"] to carry out their work. they **think for the perspective of country and its development** [*Capacity* - "younger people"] that a nation requires now for it, in this everyday changing world.
>
> In some sectors of government like Finance, Healthcare, etc it is better to give important position to **experienced** [*Capacity* - "people"] people, because they **have plenty of experience to take vital national decisions, at the same time being aware of consequences** [*Capacity* - "experienced people"].

The two texts from which these extracts are taken have a comparable number of instances of Judgement, but the patterns evident in each differ. One interesting feature of Text E5-826 is that all the Judgement is invoked. This text responds to a task on the broad topic of 'disposable consumerism', and most of the Judgement is in the category of Propriety. But beyond the fact that all but one of the targets of Judgement are people, there is no consistent pattern in the targets of Judgement in this text. The targets of Judgement in the entire text are: *people around the world, the factory, people, workers, shop keepers, shop keepers ... in the past, shopkeepers [in the new society], the person at the counter, people*. The lack of consistency in these targets, and the fact that most are producers and sellers rather than consumers, reflects a lack of coherent (and relevant) development in the arguments overall.

Text H7-4643 has more instances of Judgements of Capacity than of other categories. The targets of Judgement in this text, which responds to a task on the broad topic of 'youth in government', are: *younger people, younger people, younger people, younger people, younger people, [experienced] people, experienced people, experienced people, experienced people, employee, employee*. Clearly there is a consistent development of arguments in the text, and the pattern of Judgement reflects this.

In these texts, the difference in the main semantic resource of Judgement employed (more Propriety for 'disposable consumerism', and more Capacity for 'youth in government') appears related to the topic of each respective task. A closer examination shows also that the use of Judgement in the band 5 text is poorly managed due to the targets of Judgement, whereas the band 7 text uses Judgement in a way appropriate to the topic. As with Affect, the task is likely to have an effect on what areas of Judgement a candidate draws on in their response.

#### 3.2.2.3 Appreciation

**Appreciation** deals with "meanings construing our evaluations of 'things', especially things we make and performances we give, but also including natural phenomena" (Martin and White 2005, p 56). The domain of Appreciation is the semantics of evaluation of objects and things (material or semiotic, abstract or concrete), even when these are associated with human behaviour. To exemplify:

- *He worked professionally*. (Judgement)
- *He was a professional worker*. (Judgement)
- *It was a professional job*. (Appreciation)

Importantly for academic writing, complex processes and evaluations of human behaviours that are construed as 'things' grammatically (e.g. the use of gerund and nominalisation) are analysed in Appraisal as Appreciation, and not Judgement (Martin and White 2005, pp 58ff). Nominalisation and abstraction are both common and valued in academic writing, and complex processes and organisations (such as institutions) are often the subject matter of academic writing, which means that Appreciation is likely to also be a common feature of academic writing. Thus, the *target* of Appraisal matters in the process of analysis (e.g. determining between Judgement and Appreciation as shown above). This also applies to the three sub-divisions of Appreciation:

- **Reaction** (typically evaluations of things with the power to trigger emotion)
- **Composition** (typically evaluations of things in terms of our perception of them)
- **Valuation** (typically evaluations of things without consciousness, but which are implicated in human behaviour).

The sub-system of Appreciation is illustrated in Figure 3.28 below.

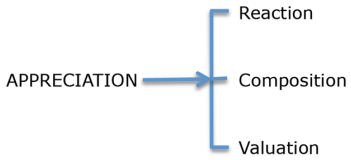



***Figure 3.28: The sub-system of Appreciation***

**Reaction** is concerned with the quality (e.g. *lovely*, *nasty*) or impact (e.g. *arresting*, *pedestrian*) of an object. This is illustrated in the following examples, in which the targets of Reaction are: *life*, *situations*, *to be in the situation not to work throughout the whole year*, and *the environment*.

- *but that doesn't mean that life will be happy* (A6-496)
- *and dealing with difficult situations which need alot of concentration and awarness* (A6-892)
- *it would be nice to be in the situation not to work throughout the whole year* (E7-1161)

- *The need for energy to fuel the machinery has caused serious damage to the environment.* (E7-100)

In the corpus of 54 texts, the scripts of Hindi L1 candidates collectively have a lower proportion of instances of Reaction than those of Arabic L1 and European-based L1 candidates, and band 6 scripts collectively have a lower proportion of instances of Reaction than band 5 or band 7 scripts. Table 3.25 shows the frequency of instances of Reaction as a percentage of all instances of Attitude in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Reaction | Instances of Reaction as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 51 | 13.9% |
| Hindi L1 | 41 | 8.8% |
| European-based L1 | 54 | 14.7% |
| Band 5 | 51 | 13% |
| Band 6 | 37 | 8.8% |
| Band 7 | 58 | 14.8% |

### Table 3.25: Frequency of Reaction

**Composition** deals with expressions evaluating the balance (e.g. *unified*, *contradictory*) or complexity (e.g. *elegant*, *simplistic*) of an object. This is illustrated in the following examples.

- *To come for a suitable position for his home country is not a simple thing.* (H5-4675)
- *In addition teenagers would tend to choose those subject wich are easy for them or those wich are simple* (E6-454)

The European-based L1 group has a higher proportion of instances of Composition than the other L1 groups, but the overall number of occurrences is too small in all groups (L1, and band score) to form the basis of any firm conclusions. Table 3.26 shows the frequency of instances of Composition as a percentage of all instances of Attitude in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Composition | Instances of Composition as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 5 | 1.4% |
| Hindi L1 | 6 | 1.3% |
| European-based L1 | 15 | 4.1% |
| Band 5 | 10 | 2.6% |
| Band 6 | 9 | 2.1% |
| Band 7 | 7 | 1.8% |

### Table 3.26: Frequency of Composition

**Valuation** deals with evaluations of whether an object is worthwhile (e.g. *timely*, *insignificant*). Recalling that Judgement is the domain of meaning dealing with evaluations of humans and their behaviour, Valuation is the closest of the three Appreciation categories to the meanings encompassed by the system of Judgement (see Bednarek 2007). Recalling also that complex processes can be dealt with in Appreciation (particularly when they are nominalised), evaluations of these kind of phenomena tend to fall in the category of Appreciation: Valuation. Examples of Valuation from the corpus of 54 texts are given below.

- *study via distance is good but not a ideal way for student to be a good citizen* (H6-2097)
- *TEENAGERS THEY DON'T HAVE RESOURCES TO STUDY IN A GOOD SCHOOL* (E5-1004)

The frequency of instances of Valuation is higher than all other sub-categories of Attitude. Despite the caveats given at various points above concerning the number of texts in the corpus and the lack of statistical analysis, the marked difference between the proportion of Valuation compared with all other sub-categories of Attitude, and the consistency of this difference across all L1 groups and all band scores (with the arguable exception of band 5) suggests that Valuation is a domain of evaluative meaning that is of importance to candidate success in IELTS Academic Writing Task 2. Table 3.27 below shows that band 5 scripts collectively have a markedly lower proportion of instances of Valuation than band 6 and 7 scripts. This finding is discussed further below.

| Group of blocks (L1 or band score) | Number of instances of Valuation | Instances of Valuation as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 143 | 39% |
| Hindi L1 | 161 | 34.5% |
| European-based L1 | 135 | 36.7% |
| Band 5 | 99 | 25.3% |
| Band 6 | 176 | 42.0% |
| Band 7 | 164 | 41.9% |

### Table 3.27: Frequency of Valuation

We have now looked at the sub-system of Appreciation and its three sub-divisions, and the frequency of occurrences in each sub-division.

As stated earlier, in the 54 texts analysed, expressions of Appreciation occurred more frequently than expressions of both Affect and Judgement in all the following groupings:

- all Arabic L1 scripts, as a block
- all Hindi L1 scripts, as a block
- all European-based L1 scripts, as a block

- all band 6 scripts, as a block
- all band 7 scripts, as a block.

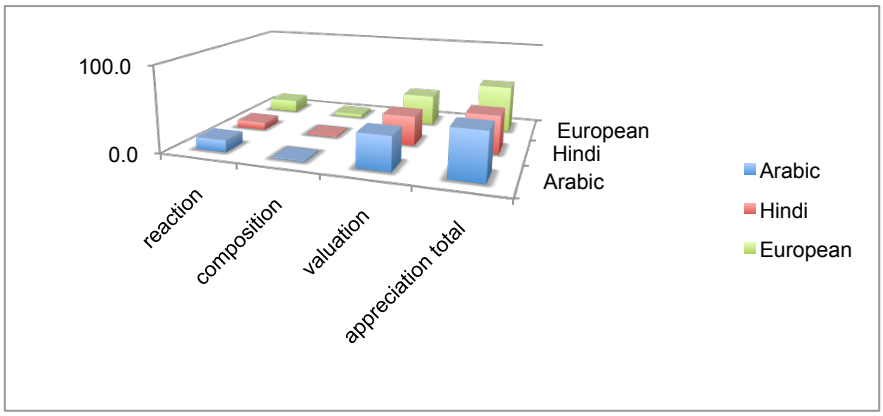



***Figure 3.29: Instances of Appreciation as a percentage of total instances of Attitude: Comparison across L1 groups***

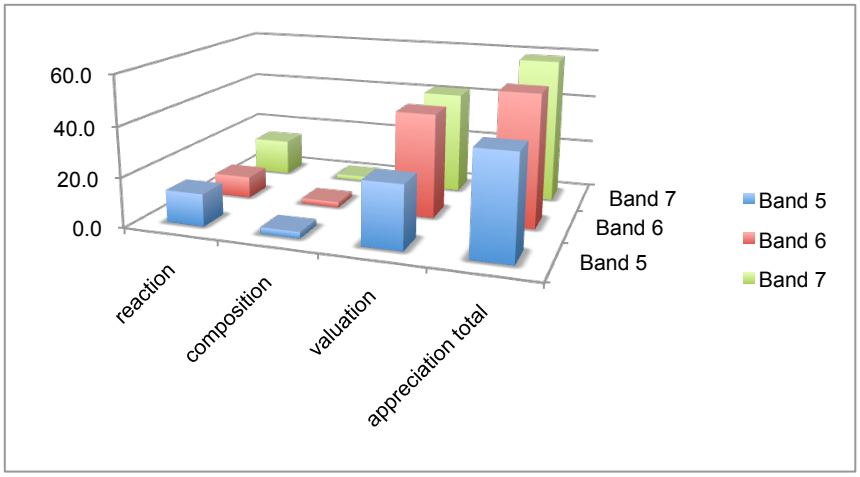



***Figure 3.30: Instances of Appreciation as a percentage of total instances of Attitude: Comparison across band scores***

Expressions of Appreciation occurred less frequently than expressions of Judgement in the following grouping:

- all band 5 scripts, as a block.

In particular, expressions of Appreciation: Valuation occurred in higher proportions than all other sub-systems.

Figure 3.29 provides a visual comparison of the percentages in instances of Appreciation (shown in the Tables 3.25 to 3.26) across the three L1 groups. There are slight differences between the three L1 groups in the overall use of Appreciation, but the very high proportion of Valuation is consistent across the three L1 groups.

Figure 3.30 provides a visual comparison of the percentages in instances of Appreciation (shown in the Tables 3.25 to 3.27 above) across the three band scores: band 5, band 6, and band 7. The lower proportion of instances of Valuation in the band 5 scripts (already mentioned) leads to an overall lower proportion of instances of Appreciation than band 6 or band 7 scripts.

Figure 3.30 shows that collectively, in the 54 scripts analysed, those which score higher have more instances of Appreciation. As with Affect and Judgement, an examination of a number of individual scripts can complement the collective perspective and highlight the potential for differences in individual texts. Below, Texts E5-340 and E7-440 are compared. Both scripts are responses to the same task on 'environmental responsibility'.

Beginning with Text E5-340, an extract is shown below, with the instances of Appreciation underlined in bold, and the type of Appreciation indicated in square brackets and italics after each instance. (Instances of Affect and Judgement are not shown.)

> The companies have a huge impact on the invaroment, so they must act down the pollution, use more **efficient** [*Reaction*] technologies iven they are more **expencive** [*Valuation*]. The company can not keep going polluating just becose of interest or keeping people employed. The climate change **have a big impact on the agriculture, our half** [*Reaction* - invoked]. The friquent ciclones, earthquakes, huricanes, **damege on billions our live and economy** [*Reaction* - invoked], so some interest is **just nothing in compare what we lose** [*Valuation* - invoked].

In Text E5-340, the targets of Appreciation are relatively concrete (e.g. *pollution, technologies, insolations, forests, houses and cars, Earth*), and/or draw on everyday concepts widely in use in public discourse, such as in the mass media (e.g. *climate change, earthquakes, companies, Australia, drought, future*). The language used to evaluate these 'things' is relatively unsophisticated (e.g. *problem, efficient, expencive, poore, not developt enaff, biger, smaller, better*), especially when the mis-spelling of even basic lexical items is taken into account. This can be compared with Text E7-440, a response to the same task shown in full below.

> All over the world, people discuss about the environmental effects of $CO_2$-emission. They complain about cyclons, floods and bushfires. However, not many countries focus on the reduction of pollution. Burning rubbish is **one of the main factor of worldwide $CO_2$-emission** [*Valuation* - invoked]. Why not reduce this amount? For example, in Switzerland every household and every company have to pay to clean up the pollution that they have produced. How much they have to pay depends on the volume they have produced. As a consequence, receycling is very **popular** [*Valuation*]. 80 per cent of bins are returned.
>
> People do not care about something unless they have to pay for Money is probably t**he best motivation to reduce the pollution** [*Valuation* - invoked]. If the people are not sensible of the pollution they produce, they will not change their attitude and keep going to pollute the environment.

As the extracts suggest, Text E7-440 has fewer instances of Appreciation than Text E5-340 (seven as opposed to 23), which is counter to the collective finding for the comparative frequency of Appreciation in band 5 and band 7 texts. The targets of Appreciation in Text E7-440 are not technical, but they involve processes construed grammatically as Participants (*burning rubbish, receycling*), and the named *Swiss system* which re-construes much of the first paragraph into a Participant that twice becomes the target of Appreciation. Apart from accurate use of more common vocabulary in the instances of Appreciation, the instances of invoked Appreciation show a degree of sophistication clearly beyond that achieved in Text E5-340.

Text E5-340 uses a mix of Valuation, Reaction and Composition, whereas Text E7-440 uses mostly Valuation (although as we have already seen, much less overall). This demonstrates that even when responding to the same task, patterns of Attitude in texts can differ, and this raises the question of whether, in scripts responding to the same task, consistent patterns of Appraisal appear at different band scores.

In conclusion, we can see that quantitatively, Appreciation: Valuation is an important discursive resource for candidates sitting IELTS Academic Writing Task 2, and that it is widely used across scripts of all three L1 groups and all three band scores examined here. Qualitatively, we can see that scripts differ in terms of the sophistication in their use of Appreciation.

The finding shown in Figure 3.30 is that higher proportions of Appreciation (as a total of instances of Attitude) correlates with higher band scores. This finding bears further investigation with controls on task. If it is generalisable, it has implications for IELTS preparation, both in the teaching and learning of vocabulary suitable for evaluating objects (as opposed to the language of emotion and judgement of human behaviour), and also in how processes and events can be (re)construed in a way that they can become targets of Appreciation.

#### 3.2.2.4 Attitude: Comparison and conclusion

Figure 3.31 illustrates a point that has been made in earlier sections. In the three L1 groups, we can observe that Appreciation is used more frequently than Judgement, which is in turn used more frequently than Affect. This is clearly the case for the Arabic L1 and European-based L1 groups. For the Hindi L1 group, the frequency of use of Judgement and Appreciation is much closer. To determine whether this finding is a reflection of the L1 of the candidates, research on scripts which responded to the same task would be needed, given the importance of the task as discussed in earlier sections.

Figure 3.32 illustrates another point that has been discussed earlier. In band 7 and band 6 scripts viewed collectively, Appreciation is more frequently used than Judgement, which is more frequently used than Affect. This is likely to reflect the tendency of academic discourse to focus on concrete and abstract phenomena and/or complex processes, and to be objective in evaluation rather than drawing on emotion. The finding that band 5 scripts used Judgement slightly more frequently than Appreciation is an interesting finding; this would be worth pursuing with research on scripts which respond to the same task. Given that this finding is of all scripts across the three L1s, such a study would not necessarily need to control for L1. Indeed, scripts from candidates with a wide variety of L1s would be preferable. Because the band 5/band 6 score is used as a 'cut off' point by many users of the IELTS test, this finding is potentially important for IELTS preparation and the explicit teaching of different strategies of evaluation.

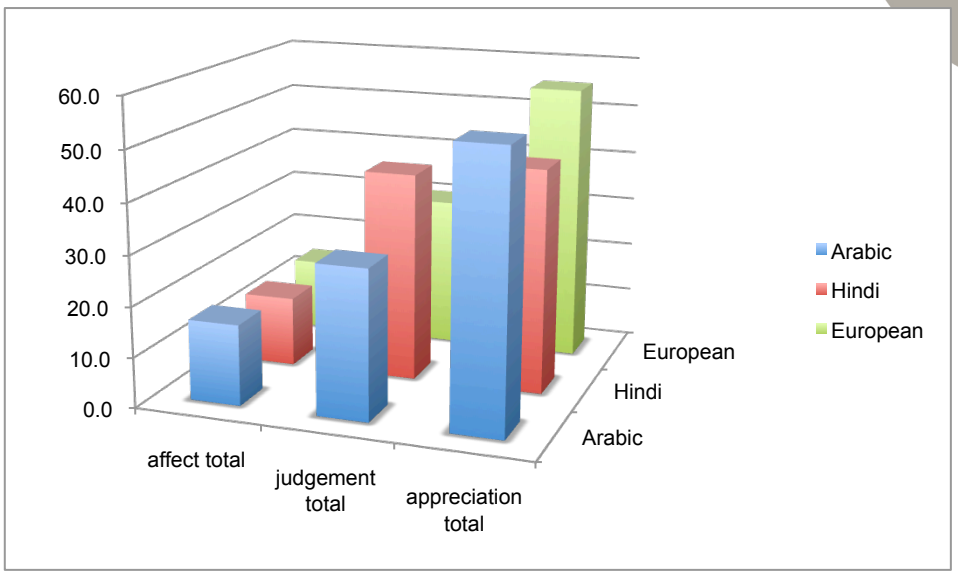



***Figure 3.31: Comparison of Affect, Judgement and Appreciation as a percentage of total instances of Attitude: Comparison across L1 groups***

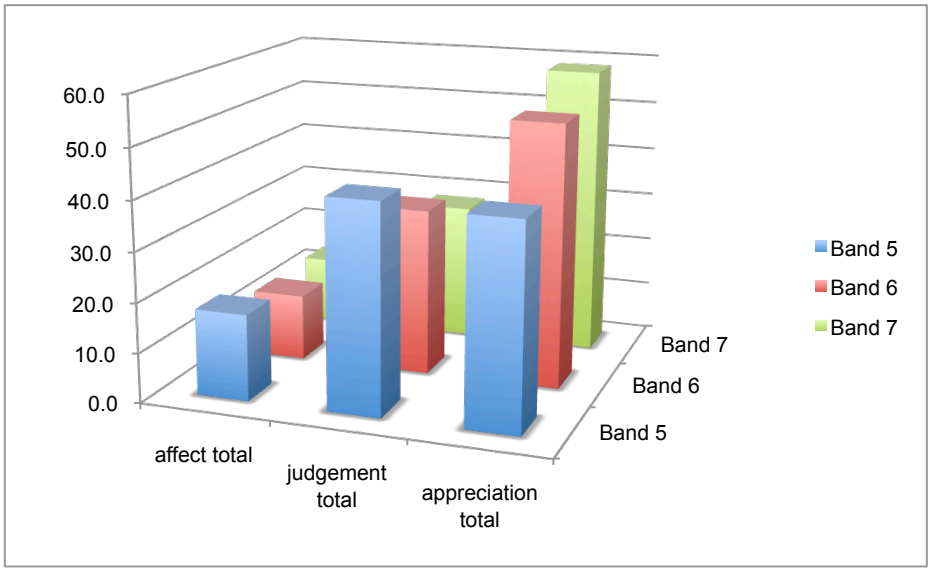



***Figure 3.32: Comparison of Affect, Judgement and Appreciation as a percentage of total instances of Attitude: Comparison across band scores***

The nature of IELTS Academic Writing Task 1 means that Attitude is unlikely to figure prominently in Task 1 responses, although this assumption is worthy of research. If correct, this means that candidates responding to a single Task 2 item may be somewhat limited in the area(s) of Attitudinal lexis they have the opportunity to draw on (Lexical Resource), and the evaluative discursive resources they can draw on to convince the reader of the worth of their arguments (Task Response). It is well established that validity and reliability in the testing of writing are enhanced by the use of a greater number of scripts (e.g. Hughes 2003; Weigle 2002; Weir 1993), and the IELTS Writing Test already addresses this in the use of two tasks. But given that only Task 2 is likely to give candidates a chance to use the Attitudinal resources of English extensively, and given that it appears likely that a strong control of the semantic resources of Appreciation is important for success on Task 2, the possibility of including a third task, or of reconsidering the overall approach and design of the IELTS Academic Writing Test may be worthy of consideration. This is discussed further in Section 4.

### 3.2.2.5 Source of Attitude

An important element of Appraisal analysis is identifying the source of Attitude. If, for example, a Judgement is made of an action or a person, or an Appreciation is made of a thing, it is important to analyse whether the source of the Judgement is the author, or whether the author is reporting a Judgement from elsewhere. The sources of Attitude in texts are one part of the discursive patterning of evaluation.

In this part of the current study, a distinction was made between authorial Attitude (where the author is the source) and 'other' Attitude (where the source of the Attitude was anyone or anything other than the author). This is exemplified in Table 3.28 below.

| Source of Appraisal: Author | Source of Appraisal: Other |
|---|---|
| *... receycling is very popular.* (E7-440)<br><br>(Appreciation: Valuation of "receycling" by the Author) | *When talking about museums many people think of them as places for learning and education, ...* (A6-1287)<br>(Appreciation: Valuation of "museums" by "many people") |

***Table 3.28: Examples of authorial Attitude and non-authorial Attitude***

The 'management' of different voices in a text is an important element of academic writing (e.g. Hood 2010; Hyland 2000). Across the corpus of 54 texts, the vast majority of instances of Attitude had the author as their source. The European-based L1 group in particular had a high frequency of authorial Attitude, but the Arabic L1 and Hindi L1 groups were also over 90%. This is shown in Table 3.29 below, and illustrated in Figures 3.33 and 3.34, which compare the source of Attitude across L1 and band score respectively.

| Group of blocks (L1 our band score) | Instances of authorial Attitude as a % of all instances of Attitude in each group of blocks | Instances of non-authorial Attitude as a % of all instances of Attitude in each group of blocks |
|---|---|---|
| Arabic L1 | 92.9% | 7.1% |
| Hindi L1 | 91.9% | 8.1% |
| European-based L1 | 98.1% | 1.9% |
| Band 5 | 95.7% | 4.3% |
| Band 6 | 93.8% | 6.2% |
| Band 7 | 92.8% | 7.2% |

***Table 3.29: Sources of Attitude***

The following extract from the beginning of Text H7-4733 provides an illustration of how different sources of Attitude can be managed in a text. We can see that the author uses projections (e.g. *Some people say*) to introduce other voices, which also allows the author to use these other voices to present evaluations. In this case, the author uses this device to present different 'sides' of the argument, before overtly introducing their own position: *In my opinions ...* . In the extract below, indications of a 'switch' of voice are shown in ***bold and italics***. Attitude is shown in **bold and underline**, with the source of Attitude given in square brackets. (The red brackets are an indication from the IELTS examiner that the text in brackets was taken verbatim from the prompt. This suggests that this candidate has drawn heavily on the task prompt in 'setting up' the different 'voices' in the text, and their own position among them.)

> 'A child is the father of man' is a very old saying and it is so **true** [Appraiser: author]. It is the young people who **lay the foundation of the future** [Appraiser: author]. There has always been a lot of contradictions about young people being given **important** [Appraiser: author] positions in the government. [***Some people say*** that ... [Appraiser: "some people" x 2].] Young people are usually **inexperienced** [Appraiser: "some people"] in handling **important** [Appraiser: "some people"] matters. They may not have **foresightedness** [Appraiser: "some people"] for making the right decisions for the country. ***People often also say*** that **if young person is given too much power, they can exploit the power** [Appraiser: "people"].
>
> ***There are others who say*** that ... [Appraiser: "others"]. ***They say*** if a young person is given ... [Appraiser: "others"] **they will be able to give the country a modern thinking** [Appraiser: "others"]. Young people are often more **hardworking and zealous** [Appraiser: "others"] when it comes to executing new ideas. A young person is **not afraid** [Appraiser: "others"] of taking risks and tying out new things. A young person in an **important** [Appraiser: "others"] position can also **understand the various problems** [Appraiser: "others"] faced by youth and **tackle them in a better way** [Appraiser: "others"] than their older counterparts.
>
> ***In my opinions*** the young people should be given **important** [Appraiser: author] ... as they are **more energetic and more helpful** [Appraiser: author] for the future. ...

This kind of 'sourcing' of Attitude in a text is an important resource for academic writing. Yet Attitude sourced from voices other than the author rarely occurred in the scripts of candidates in the three L1 groups, and across the three band scores. The extremely high proportion of authorial Attitude and the consistency of this finding (over 90% in all L1 groups and all bands) raises the issue of the management of 'voices' in responses to Task 2 of the IELTS Academic Writing Test, and the extent to which Task 2 tests this important skill. This point is revisited at the end of Section 3.2.3.

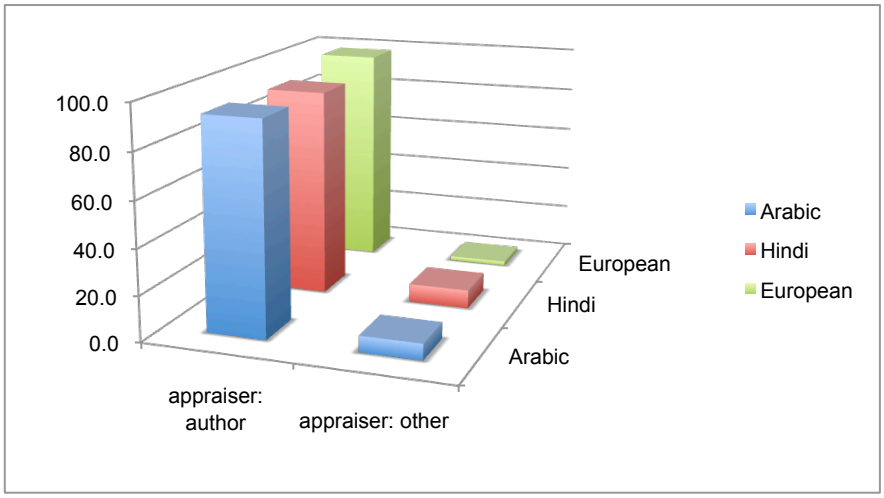



***Figure 3.33: Sources of Attitude as a percentage of total instances of Attitude: Comparison across L1 groups***

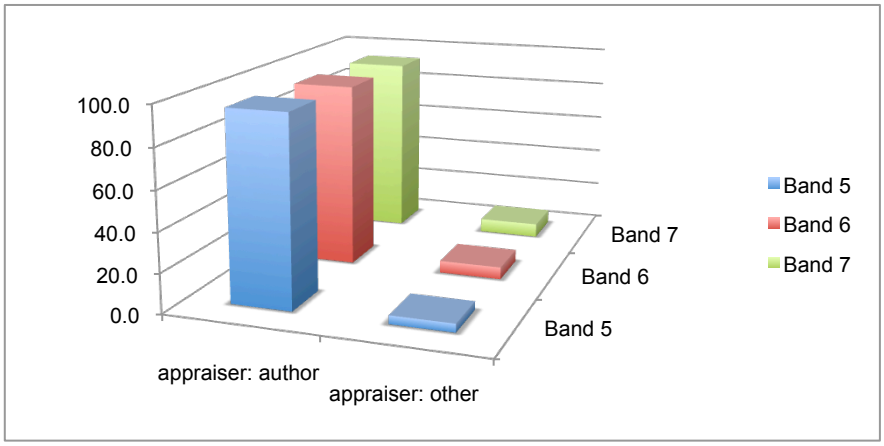



***Figure 3.34: Sources of Attitude as a percentage of total instances of Attitude: Comparison across Band Scores***

Source of Attitude is just one resource by which voices can be managed in writing, and this brings into focus the system of **Engagement** in Appraisal, which is concerned with the management of different 'voices' (including the author's) in discourse. It is to Engagement that we now turn.

### 3.2.3 Analysis of Engagement

The system of **Engagement** is concerned with "the linguistic resources by which speakers/writers adopt a stance towards the value positions being referenced by the text and with respect to those they address" (Martin and White 2005, p 92). Many of the resources (e.g. the use of reporting verbs, modality, negation) are familiar to learners of English as a second language. The system of Engagement organises these and other resources not around their grammatical form, but according to their functions.

#### 3.2.3.1 Monogloss and Heterogloss

The most basic distinction in the system is between **Monogloss** ('bare assertions' that do not overtly recognise the possibility of alternate positions to the one expressed) and **Heterogloss** (any expression which recognises that the position stated is not the only possible one, including devices such as those mentioned in the previous paragraph). The choices in the system under the choice of **Heterogloss** are shown in Figure 3.35 below.

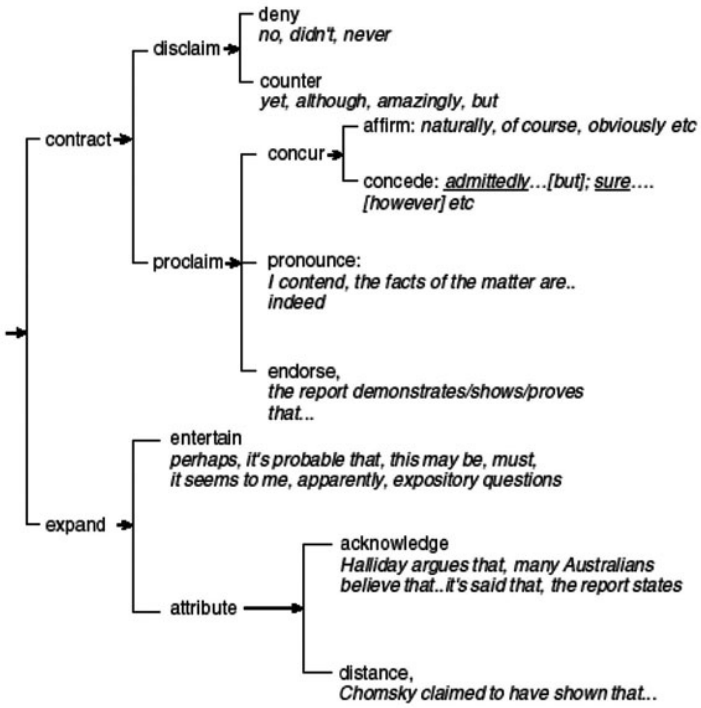



***Figure 3.35: Choices under Heterogloss in the system of Engagement (source: Martin and White 2005, p 134)***

As the figure shows, in the choice of Heterogloss, the basic division is between:

- **Contract:** those linguistic choices that, in some way, function to 'close down' the possible alternate voices (e.g. by proclaiming validity, as in: ... *the fact is* ...)
- **Expand:** those linguistic choices that allow for the possible alternative voice (e.g. by reporting as in: ... *Smith argues*...; or hedging as in: ... *it could be possible to* ...).

Table 3.30 gives examples of each category from the corpus of 54 texts. Table 3.31 gives the frequency of instances of Contract and Expand, and Monogloss across the 54 texts.

| Heterogloss | | Monogloss |
|---|---|---|
| **Contract** | **Expand** | |
| *As a matter of fact not all the people are the same ...* (E6-454)<br><br>*As for as I concered younger people do everything from the older people.* (H5-4675) | *The airways companies should reduce that ...* (A6-9)<br><br>*... some people claim, that it is the governments role to encourage the reduce of pollution.* (E6-979) | *... they seek the best schools...* (A7-7762)<br><br>*Education makes a man perfect.* (H7-7810) |

***Table 3:30: Examples of Heterogloss and Monogloss***

| | Heterogloss | | | Monogloss |
|---|---|---|---|---|
| | **Contract** | **Expand** | **Total** | |
| Arabic L1 | 14.5% | 23.5% | 37.9% | 62.1% |
| Hindi L1 | 20.6% | 25.2% | 45.8% | 54.2% |
| European-based L1 | 17.6% | 20.8% | 38.5% | 61.5% |
| Band 5 | 17.1% | 23.1% | 40.2% | 59.8% |
| Band 6 | 17.7% | 23.2% | 41.0% | 59.0% |
| Band 7 | 18.1% | 23.1% | 41.2% | 58.8% |

***Table 3:31: Frequency of Heterogloss and Monogloss***

The ratio of instances of Heterogloss to Monogloss is collectively consistent across the three band scores. The Hindi L1 group has a slightly higher proportion of instances of Heterogloss compared to the other two L1 groups, but overall the most noticeable finding of this broad comparison is the relative homogeneity.

But despite the relative homogeneity when viewed collectively, individual texts can vary widely in terms of their use of the resources of Engagement, including how these resources are distributed within a text. Below, we compare extracts from two texts from candidates sharing the same L1 (Arabic) and which scored the same band score (band 7) in terms of the degree to which they use Heterogloss.

In Text A7-9464, shown immediately below, all instances of Heterogloss are indicated by bold font and underlining. The categories of Heterogloss are also indicated, although at this point the existence (or absence) of Heterogloss is all we are concerned with.

> People in the past used to have ... . This features **may** [*Expand: Entertain*] incloude housing, customs, food and language as well as religion. This features used to be notesable when people travel from country to another. Nowadays, places ... and more similarities are found. **In my openion** [*Expand: Entertain*], there are many causes of this ... .
>
> Firstly, globalisation plays big role in ... . Globalisation aims to make ... . This is the great reason that made places all over the world ... .
>
> Secondly, ... is also a reason to have ... . Australia is a good example to show the effect of ... . People who ... , practise similar life-style in ... .
>
> Moreover, turisim make ... . For example, Dubai provides these things, that why its one of the first countries that attract turist.
>
> There are many advantages for having ... . First, people **will** [*Expand: Entertain*] feel ... and the **will not** [*Contract: Disclaim: Deny*] feel that they are ... . People **will be able to** [*Expand: Entertain*] practis ... .
>
> **On the other hand** [*Contract: Disclaim: Counter*], there are also some disadvantages for this issue. As each country will ... . Furthermore, new generations **will not** [*Contract: Disclaim: Deny*] know ... . It **may** [*Expand: Entertain*] also creat crimes and problems.
>
> Having a one world that ... **may** [*Expand: Entertain*] be a good thing **but** [*Contract: Disclaim: Counter*] many other thing as the disadvantegs **should be** [*Expand: Entertain*] counted to avoid the bad secomostances.

This text uses a number of the heteroglossic resources of Contract (e.g. *will not, on the other hand*) and Expand (e.g. *may, in my openion*). (These choices and their sub-types are discussed in detail in the sub-sections following.) A feature of this text, though, is the relatively high number of monoglossic, 'bare assertions'. We can see that paragraphs 2-4 are completely monoglossic, and that the use of Engagement in this text is related to the generic structure of the text (which is an analytical exposition topologically close to a discussion – see Section 3.1.1 above). The Argument stages are monoglossic, but the Thesis and Conclusion (as might be expected) and the Argument For and Argument Against stages (where, in its structure, the text structure becomes heteroglossic) use the resources of Heterogloss. Thus, in addition to the topic of the task potentially influencing patterns of Attitude, we can see that the genre of a script potentially influences patterns of Engagement.

The next script, A7-9, is also an analytical exposition (though topologically 'further away' from a discussion than Text A7-9464), yet this text uses the resources of Heterogloss throughout the text. The exception is the Reiteration stage of the text, which begins with an instance of **Contract** (*Unfortunately, ...*), and then is monoglossic.

> **Although** [*Contract: Disclaim: Counter*] ... are ... than they once were, **it is hard to believe** [*Contract: Disclaim: Counter*] that this **could** [*Expand: Entertain*] a positive trend. With the Earth's population growing at a rapid rate, **it is only natural** [*Contract: Proclaim: Concur*] that ... **will** [*Expand: Entertain*] also increase to meet this growing demand. **Yet** [*Contract: Disclaim: Counter*] as demand grows, the rate at which ... **will** [*Expand: Entertain*] also increase. **Inevitabily** [*Contract: Proclaim: Concur*] this ... **will** [*Expand: Entertain*] cause the quality of ... to deteriorate drastically. An example of this is the ... scandal of 2008, where ... . When the ... is lowered, one **needs to** [*Expand: Entertain*] ask themselves whether ... . It **is not** [*Contract: Disclaim: Deny*] feasible that at such low costs, ... **would** [*Expand: Entertain*] have bothered with ... . **On the contrary** [*Contract: Disclaim: Counter*], **even though** [*Contract: Disclaim: Counter*] a ... **might seem** [*Expand: Entertain*] relatively cheap
>
> and **you** **may**[*Expand: Entertain*] **believe** [*Expand: Acknowledge: Attribute*]
>
> that you have made ... , **in most likelihood** [*Expand: Entertain*] you are receiving ... . **Unfortunately** [*Contract: Disclaim: Counter*], many ... have escaped ... , so it is ultimately down to ... .

In the discussion above, we pointed out that the genre of discussion is heteroglossic in its structure. This does not mean, though, that discussions will be heteroglossic and expositions monoglossic. Skilful writers draw on the resources available to them at the level of text structure (genre), and at the level of discourse (e.g. Appraisal), and

just as the resources at these (and other) levels can be used to 'work together', they can also be 'played off' against each other.

The analysis of the two texts above reminds us that while the ratio of instances of Heterogloss to Monogloss is, collectively, relatively consistent across the three band scores, individual texts will vary in which resources they employ (even in the same L1 group and the same band score), and the ways in which they employ them. Further, this variation can be related to a number of factors, including (but not limited to) topic and genre. In the sub-sections that follow, we examine more delicate choices in the sub-systems of Contract and Expand, and the discursive choices made (and, importantly, not made) by candidates in responding to the IELTS Academic Writing Task 2.

### 3.2.3.2 Heterogloss: Contract

The sub-system of **Contract** has two basic sub-divisions and a number of sub-divisions beneath that. The first to be dealt with here are the two categories of **Disclaim: Deny**, and **Disclaim: Counter**. Disclaim: Deny involves the use of negation. By using the negative, a position is introduced in the discourse in order to reject it. Examples from the corpus follow.

- *... people may feel that they don't have freedom.* (A6-496)
- *... cheaper doesn't mean better ...* (E7-7)

Arabic L1 candidates use Deny less frequently than the other two L1 groups in the 54 texts analysed, and scripts scored at band 5 use this strategy slightly more than scripts at bands 6 and 7. The proportional difference between its use by the Arabic L1 group and the other groups suggests that this may reflect discursive strategies (e.g. a tendency not to use negation in argumentation) from the L1. Table 3.32 shows the frequency of instances of Deny as a percentage of all instances of Engagement in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Deny | Instances of Deny as a % of all instances of Engagement in each group of blocks |
|---|---|---|
| Arabic L1 | 38 | 4.8% |
| Hindi L1 | 88 | 10.8% |
| European-based L1 | 76 | 9.0% |
| Band 5 | 86 | 10.1% |
| Band 6 | 57 | 6.7% |
| Band 7 | 59 | 7.9% |

*Table 3.32: Frequency of Deny*

**Disclaim: Counter** is a discursive strategy typically achieved with conjunctive devices. It is where one position 'replaces' another. Examples follow.

- *Another reason that many people throw things away rather than repair them, is the change of the value of things to people.* (E6-1189)
- *Instead, a wide variety of programs ranging from educational, political to entertaining are available just through a push of a button.* (A7-116)

The difference between L1 groups, and between band scores, in terms of proportion of instances of Counter as shown in Table 3.33 is unlikely to be significant. All groups (L1 and band score) have a low proportion of Counter, though band 5 scripts have a slightly lower proportion than the other two band scores. Table 3.33 shows the frequency of instances of Counter as a percentage of all instances of Engagement in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Counter | Instances of Counter as a % of all instances of Engagement in each group of blocks |
|---|---|---|
| Arabic L1 | 53 | 6.7% |
| Hindi L1 | 66 | 8.1% |
| European-based L1 | 57 | 6.7% |
| Band 5 | 45 | 5.3% |
| Band 6 | 74 | 8.7% |
| Band 7 | 57 | 7.6% |

*Table 3.33: Frequency of Counter*

Deny and Counter are 'negative' strategies of Contract. In contrast, there are three strategies under the category of **Proclaim**, and these are:

- **Proclaim: Concur** (an overt signal that the author has the same position as a putative dialogic partner - Martin and White 2005, p 122), such as:
  - *Obviously, sentences have to follow delicts and unlawful behaviour.* (E6-698)
  - *... of course the governments should pay a high salary for those people who work very hard ...* (A6-892)
- **Proclaim: Pronounce** (explicit authorial statements of intervention into the argument - Martin and White 2005, p 127), such as:
  - *As we all know health is very important in today's life.* (H5-512)
  - *As a matter of fact not all the people are the same ...* (E6-454)
- **Proclaim: Endorse** (the use of reporting verbs and similar forms of projection that give validity to the projected content - Martin and White 2005, p 126), such as:

- *This indicates that readings has a direct influence on our brain physiological activities.* (A7-116)
- *Some evidence is to be found in the way companies produce and export in many different countries.* (E6-1189)
- *Furthermore, the past experiences revealed that employers are eager to have a long term taskforce ...* (E7-1161)

There are very few instances of any/all of the strategies of Proclaim in the 54 texts, so these three strategies (i.e. Concur, Pronounce, and Endorse) are presented collectively in Table 3.34.

| Group of blocks (L1 or band score) | Number of instances of Proclaim | Instances of Proclaim as a % of all instances of Engagement in each group of blocks |
|---|---|---|
| Arabic L1 | 23 | 2.9% |
| Hindi L1 | 14 | 1.7% |
| European-based L1 | 16 | 1.9% |
| Band 5 | 14 | 1.6% |
| Band 6 | 20 | 2.3% |
| Band 7 | 19 | 2.5% |

***Table 3.34: Frequency of Proclaim***

The fact that there are so few instances of Proclaim in the 54 texts again raises questions about the content validity of the IELTS Academic Writing Test. Similar quantification of the use of Proclaim in academic discourse more broadly is needed to determine whether this set of discursive resources is under-represented (and therefore 'under-tested') in candidate responses to IELTS Academic Writing Task 2, or whether these discursive resources are also typically little used in the Target Language Use (TLU) domain. But the finding here is that, like non-authorial Attitude (Section 3.2.2.2), the resources of Proclaim are little used in the 54 analysed texts. This suggests that research investigating whether Proclaim is under-represented in candidate responses to the IELTS Academic Writing Test (as compared to student writing in the TLU domain) is warranted.

Figure 3.36 compares the use of the different discursive resources of Contract across the three L1 groups visually. It shows that the Arabic L1 group uses Contract less (proportionately) than the European-based L1 group, who in turn uses these resources less than the Hindi L1 group. The resources of Proclaim are little used by all three groups, and these differences are largely a reflection of the proportions of the use of Deny (and to a lesser extent, Counter) as discussed above.

Similarly, Figure 3.37 compares the use of the different discursive resources of Contract across the three band scores visually. Despite some minor differences in the proportion of instances of Counter and Deny (which are used in small numbers overall), the general finding shown in Figure 3.37 is a consistent quantitative use of these resources across the three band scores.

In summary, Arabic L1 candidates use the discursive resources of Disclaim: Deny less than the other two L1 groups in the 54 texts, and band 5 scripts have fewer instances of Disclaim: Counter than the other band scores, but overall the resources of Contract, and particularly the resources of Contract: Proclaim are used relatively little in these texts.

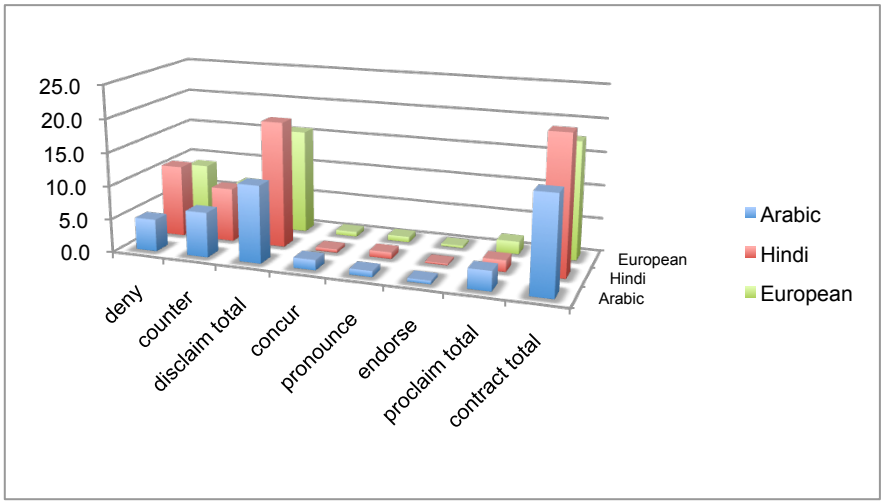



***Figure 3.36: Resources of Contract as a percentage of total instances of Engagement: Comparison across L1 groups***

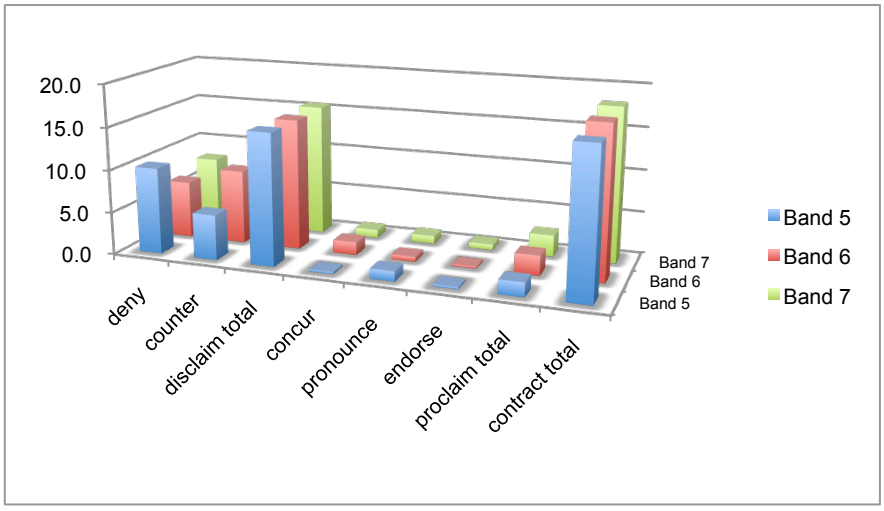



***Figure 3.37: Resources of Contract as a percentage of total instances of Engagement: Comparison across band scores***

Having looked at the sub-system of Contract in the system of Engagement, we now turn to the sub-system of Expand.

### 3.2.3.3 Heterogloss: Expand

In contrast to the resources of **Contract**, the resources of **Expand** "[open] up the dialogic space for alternative positions" (Martin and White 2005, p 103). There are three resources under Expand:

- **Entertain**
- **Attribute: Acknowledge**
- **Attribute: Distance**.

**Entertain** is the category of wordings whereby the authorial voice allows for the possibility of other voices through the use of modality and closely related linguistic devices, including prepositional phrases indicating the point of view of the author (e.g. *in my opinion*) and projecting mental Process clauses such as *I think* and *I suspect* (Martin and White 2005, pp 104-5). Examples from the 54 scripts are given below.

- *Poeople should have space to controll their lifestyle.* (A5-502)
- *We can learn new upgrades our study or our subjects with all these.* (H6-2097)
- *Why not reduce this amount?* (E7-440)
- *In my opinion, it is possible to improve the Swiss system.* (E7-440)

Entertain is quite frequently used in the 54 scripts analysed, and accounts for approximately 20% of the instances of Engagement (including monogloss) in each block. The frequency of Entertain does not appear to be different between L1 groups, nor to be an indicator of band score. But the high frequency of Entertain across all blocks does indicate that this is a domain of meaning that is certainly 'in play' in candidate responses to IELTS Academic Writing Task 2. Table 3.35 shows the frequency of instances of Entertain as a percentage of all instances of Engagement in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Entertain | Instances of Entertain as a % of all instances of Engagement in each group of blocks |
|---|---|---|
| Arabic L1 | 164 | 20.8% |
| Hindi L1 | 181 | 22.2% |
| European-based L1 | 167 | 19.8% |
| Band 5 | 174 | 20.5% |
| Band 6 | 180 | 21.1% |
| Band 7 | 158 | 21.2% |

***Table 3.35: Frequency of Entertain***

Moving on from the category of **Entertain**, the two categories under **Attribute** are those where the voice of a proposition is explicitly marked as not being the voice of the author. **Attribute: Acknowledge** is where the reporting verb (or other wording that indicates a semantic projection of another voice) is 'neutral', and does not indicate the author's position in relation to the projection (e.g. *say*, *report*, *according to*) (Martin and White 2005, pp 112-3). Examples taken from the 54 texts follow.

- *Nowadays many citizen believe that every criminal has to be put in jail, although there are other voices who provide different suggestion ...* (E6-698)
- *... many people think of them as places for learning and education ...* (A6-1287)
- *To sum up, people think the have to be up-to-date ...* (E5-1199)

The European-based L1 group has a lower proportion of instances of Acknowledge than the other two L1 groups, but the numbers overall in this category are too small to make any conclusions. Given the importance of Acknowledge and the conventional ways it is managed in academic writing (e.g. through standardised referencing, in addition to other more general resources of reported speech and so on), the fact that Acknowledge is so little used in the 54 scripts analysed here is a surprising finding. This is discussed below in relation to **Distance**. Table 3.36 shows the frequency of instances of Acknowledge as a percentage of all instances of Engagement in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Acknowledge | Instances of Acknowledge as a % of all instances of Engagement in each group of blocks |
|---|---|---|
| Arabic L1 | 21 | 2.7% |
| Hindi L1 | 24 | 2.9% |
| European-based L1 | 7 | 0.8% |
| Band 5 | 22 | 2.6% |
| Band 6 | 16 | 1.9% |
| Band 7 | 14 | 1.9% |

*Table 3.36: Frequency of Acknowledge*

**Attribute: Distance** is similar to Acknowledge above, except that the choice of wording indicates that the author does not share the perspective of the other voice. Examples from the 54 texts analysed illustrate.

- *On the other hand many people claim, that prisons in our decades are similar to hotels.* (E6-698)
- *However, some people claim, that it is the governments role to encourage the reduce of pollution.* (E6-979)

The archetypal expression of Distance is the reporting verb *claim*, and both examples above us this wording. In fact, the two examples above are the only two instances of Distance in the 54 texts analysed, and this is a very surprising finding. Table 3.37 shows the frequency of instances of Distance as a percentage of all instances of Engagement in each group of blocks (either L1 or band score).

| Group of blocks (L1 or band score) | Number of instances of Distance | Instances of Distance as a % of all instances of Engagement in each group of blocks |
|---|---|---|
| Arabic L1 | 0 | 0% |
| Hindi L1 | 0 | 0% |
| European-based L1 | 2 | 0.2% |
| Band 5 | 0 | 0% |
| Band 6 | 2 | 0.2% |
| Band 7 | 0 | 0% |

*Table 3.37: Frequency of Distance*

The resources of Attribute (i.e. Acknowledge and Distance) are crucial devices for written academic discourse, where authors need to position themselves in relation to the existing literature in their field. The finding that Acknowledge is very little used in the 54 texts analysed, and that Distance is used almost not at all suggests that these important resources of managing voices through projection and related devices are actually little tested in IELTS Academic Writing Task 2. This is supported by Moore and Morton (1999), who found that university assignments require students to draw predominantly on primary and secondary sources, whereas Task 2 of the IELTS Writing Test requires candidates to draw on prior knowledge. Mayor et al. (2007) made a similar observation, with the pronoun *I* figuring prominently in Theme position in clauses in their data.

> In complying with the rubric to '*present an argument ... to an educated reader*', candidates are thrown back on their own resources, which is not a situation similar to that encountered in academic writing at tertiary level. However, it may well be that candidates that can cope successfully in this situation will also be successful in more traditional forms of academic writing in English. (p 301)

In contrast to Mayor et al., we would argue that their findings, in concert with those of Moore and Morton (1999), and with our own findings on the frequency of Attribute (immediately above), Proclaim (Section 3.2.3.2), and the sources of Attitude (Section 3.2.2.3 above) suggest that further investigation is warranted into the content validity of IELTS Academic Writing Test with respect to these discursive domains of academic writing in English.

We speculate that Task 1 can be expected to generate instances of Proclaim: Endorse (e.g. *The graph shows ...*) or Attribute: Acknowledge (e.g. *According to the graph ...*) in candidate responses, but that it is highly unlikely to generate instances of Attribute: Distance (e.g. *The graph claims ...*). Further, conventions of Attribute particular to academic writing (including in-text referencing and the use of footnotes in different disciplines) are not tested in any way.

Thus, we would argue that the demands placed on candidates to succeed in Task 2 (and probably Task 1) 'under-test' their ability to 'manage voices' in their writing – a crucial skill in academic writing. Therefore, the discourse of candidate responses to Task 2, at least, differs in important ways from the demands of the Target Language Use domain. Addressing these issues will place pressure on IELTS test designers and item writers. These issues are discussed in the conclusion to this study.

Figure 3.38 compares the use of the different discursive resources of Expand across the three L1 groups visually. It shows that the European-based L1 group uses Expand less overall, and Hindi L1 groups uses Expand more overall, but that the differences are minor.

Similarly, Figure 3.39 compares the use of the different discursive resources of Expand across the three band scores visually. It shows a relatively consistent frequency across all three band scores.

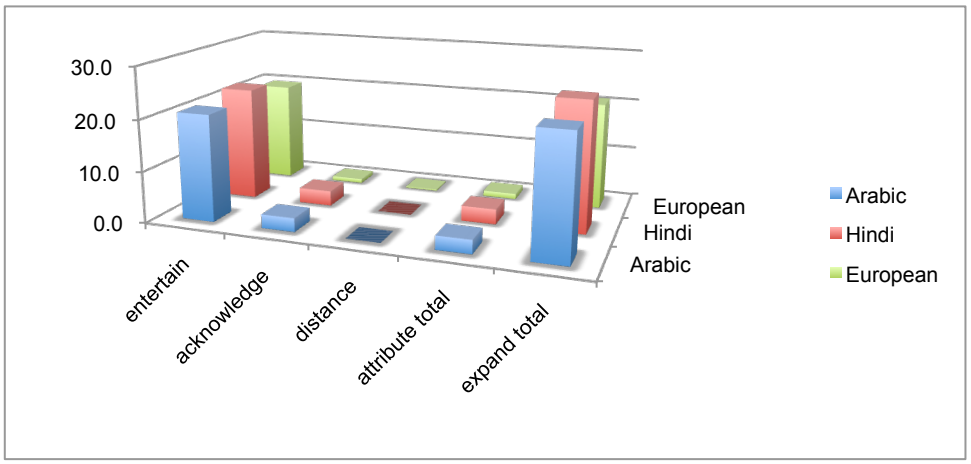



***Figure 3.38: Resources of Expand as a percentage of total instances of Engagement: Comparison across L1 groups***

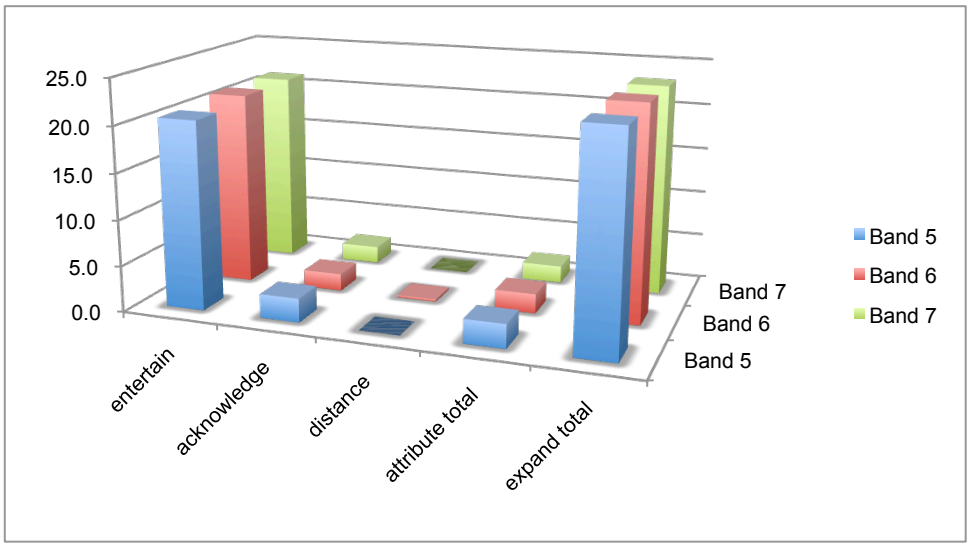



***Figure 3.39: Resources of Expand as a percentage of total instances of Engagement: Comparison across band scores***

#### 3.2.3.4 Engagement: Conclusion

In conclusion, we can see that while there is individual variation in texts, and relative consistency across L1 and band score in the frequency of use of Engagement resources in the 54 scripts analysed, there are important resources of Engagement that are little used in these scripts, namely: **Proclaim** and **Attribute**. These findings, already discussed above, are revisited in Section 4.

### 3.2.4 Appraisal analysis: Conclusion

In Sections 3.2.2 and 3.2.3, we compared the frequency of instances of Attitude and Engagement respectively, according to L1 (Arabic, Hindi, and European-based) and band score (band 5, 6, and 7) to address the research questions. The number of texts is relatively small, and as explained earlier, the study of the distribution and frequency of Appraisal resources has been conducted not in order to determine statistical significance, but to better understand the interpersonal resources used in candidate responses to Task 2. In addition, we examined Appraisal patterns in individual texts, which showed that individual variation in the type of resources used, and the way they are used in individual texts, is influenced by factors including task and genre, and findings related to L1 and band score must be understood in this light.

The study of Attitude found that Hindi L1 candidates tend to use the resources of Judgement more frequently, and the resources of Appreciation less frequently than Arabic L1 and European-based L1 candidates. In terms of band score, band 5 scripts use Judgement in greater proportion than Appreciation, whereas band 6 and band 7 scripts use the resources of Appreciation in greater proportion than Judgement. This finding suggests that further research that controls for task is warranted in this area. Further, across all L1 groups and band scores, the source of Attitude is overwhelmingly the author of the script. Further research to investigate the extent to which this is, or is not, consistent with discourse typical of the Target Language Use domain is warranted.

The study of Engagement found that the resources of Contract: Proclaim, and Expand: Attribute are little used in the 54 analysed texts, and that this is consistent with the finding (stated above) of the overwhelming use of authorial Attitude, and with the findings of Moore and Morton (1999) and Mayor et al. (2007). This suggests that research exploring these features in responses to Task 1 of the IELTS Academic Writing Test, and in texts in the TLU domain, is warranted, as there is potentially an important issue with the content validity of the IELTS Academic Writing Test. Indeed, as one anonymous reviewer pointed out, research that investigates the patterns of Attitude and Engagement in IELTS scripts in general, and how these compare against a broader corpus of academic writing, is warranted.

## 3.3 Discourse analysis: Conclusions

We are now in a position to return to the two research questions which are the focus of this section, and to provide answers in relation to genre and Appraisal. It bears repeating that these findings are based on an analysis of 54 texts (18 from each of the three L1 groups; 18 from each of the 3 band scores).

***Research Question 1*: What systematic differences are there in the linguistic features of scripts produced for IELTS Academic Writing Task 2 at bands 5, 6, and 7?**

In terms of genre, the findings are as follows:

- **Band 5 scripts** are more likely to be *atypical* in their generic structure, and to have a generic structure that *is not* matched to the demands of the task than band 6 and band 7 scripts.
- **Band 6 scripts** are more likely to be atypical in their generic structure, and to have a generic structure that is not matched to the demands of the task than band 7 scripts, but are more likely to be *typical* in their generic structure, and to have a generic structure that *is* matched to the demands of the task than band 5 scripts.
- **Band 7 scripts** are more likely to be *typical* in their generic structure, and to have a generic structure that *is* matched to the demands of the task than band 6 scripts and band 5 scripts.

While there is variation within the group of scripts in each band score according to genre (see Figure 3.19), the findings regarding genre are consistent with what would be expected of a valid, reliable test of writing, with candidate responses more likely to be more closely aligned with the demands of the task and with established conventions of academic writing the higher their band score.

In terms of Appraisal, the findings are as follows. In the system of Attitude:

- **Band 5 scripts** use Judgement in a slightly higher proportion than Appreciation, and use Appreciation in a higher proportion than Affect.
- **Band 6 scripts and band 7 scripts** use Appreciation in higher proportions than Judgement, and Judgement in higher proportions than Affect.
- **Band 5, 6, and 7 scripts** have a very high proportion (over 90%) of authorial Attitude.

In the system of Engagement:

- **Band 5, 6, and 7 scripts** are relatively consistent in their use of the resources of Contract.
  - **Band 5, 6, and 7 scripts** use the resources of Proclaim very little.
- **Band 5, 6, and 7 scripts** are relatively consistent in their use of the resources of Expand.
  - **Band 5, 6, and 7 scripts** use the resources of Attribute very little.

***Research Question 2*: What systematic differences are there (if any) in the linguistic features of the scripts produced for IELTS Academic Writing Task 2 for European-based, Hindi, and Arabic L1 backgrounds?**

In terms of genre, the findings are as follows:

- **the scripts of all three L1 groups** are more likely to be matched to task in terms of their generic structure than otherwise
  - this is particularly the case for scripts of **the Arabic L1 group**
- **the scripts of all three L1 groups** are relatively unlikely to be atypical in generic structure
- **the scripts of the Arabic L1 and Hindi L1 groups** are more likely to have a variation on a conventional generic structure than the European-based L1 group
  - **the scripts of the European-based L1 group** are more likely to have a conventional generic structure than the scripts of the Arabic L1 and Hindi L1 groups.

In terms of Appraisal, the findings are as follows. In the system of Attitude:

- **all L1 groups** use Appreciation in a higher proportion than Judgement, which is in turn used in a higher proportion than Affect (see Figure 3.31)
  - **in the Hindi L1 group**, the proportion of Appreciation to Judgement is much closer than in the other two L1 groups
- **all L1 groups** have a very high proportion (over 90%) of authorial Attitude
  - **the European-based L1 group** has a proportion of over 98% of authorial Attitude.

In terms of Engagement:

- **Hindi L1 scripts** use a slightly higher proportion of Contract resources than European-based L1 scripts, which in turn use a slightly higher proportion than Arabic L1 scripts. The differences are small and could easily be due to factors other than L1.
  - **All L1 groups** use the resources of Proclaim very little.
- **Hindi L1 scripts** use a slightly higher proportion of Expand than Arabic L1 scripts, which in turn use a slightly higher proportion than European-based L1 scripts. The differences are small and could easily be due to factors other than L1.
  - **All L1 groups** use the resources of Attribute very little.

In the course of the research, other findings have also been made (and a number of these have been discussed in earlier sections in this report). In Section 4, we discuss all significant findings from the project, and their implications.

# 4 CONCLUSIONS

## 4.1 Overview

This study has used two methodological approaches to examine the discourse of candidate responses to Task 2 of the IELTS Academic Writing Test. Scripts of candidates from three first-language (L1) background groups (Arabic, Hindi, and European-based), and from candidates who scored three different bands on the Writing Test (band 5, 6 and 7) were collected and analysed.

Mapping the three L1 groups against the three band scores gave nine 'blocks' of scripts for comparison: Arabic L1 Band 5, Arabic L1 Band 6, Arabic L1 Band 7, Hindi L1 Band 5, Hindi L1 Band 6, and so on (see Table 1.1).

Computational Text Analysis (CTA) was used to examine 254 scripts (approximately 30 from each block). These scripts were analysed in terms of:

- text length
- readability index
- Word Frequency Level (WFL)
- lexical diversity
- syntactic complexity
- incidence of all connectives
- coreferentiality (Stem and Argument overlap)

Systemic Functional Linguistics (SFL) was used to examine a subset of 54 scripts (six from each block). These scripts were analysed in terms of:

- genre:
  - typicality
  - match to task
- Appraisal
  - Attitude
  - Engagement.

Broadly speaking, the CTA provided evidence for some differentiation of linguistic features of scripts rated at the three bands of 5, 6, and 7 and across the three L1 backgrounds. The SFL analysis raised issues primarily related to content validity. That said, as might be expected, there was some 'overlap' in the findings from the two approaches. A number of the findings of the study were unexpected, and the final recommendations draw also on these unexpected findings.

After discussing the limitations of the study, the sub-sections that follow consider the conclusions under the broad headings of *Differentiation according to L1*, *Differentiation according to band score*, *Rating and reliability, Genre and task difficulty*, *Presence and absence of discoursal features in scripts*, and *Handwritten scripts*.

## 4.2 Limitations

The study has used multiple methods. Both quantitative and discourse-analytic analyses of the data were performed. The strength of such a multi-method approach is that it enables researchers to combine quantitative and qualitative analyses for the sake of triangulation and complementarity (Johnson and Christensen, 2008). There remain, however, some limitations.

The first limitation is the number of scripts analysed. The initial plan was to analyse 30 scripts from each 'block' (a total of 270) using CTA. It was not possible to source 30 scripts from each block, and as Table 2.2 detailed, four of the nine blocks had fewer than 30 scripts.

For the SFL analysis, due to the labour-intensive nature of detailed, 'manual' discourse analysis, only six texts from each block were analysed.

These shortages in numbers pose potential problems if we aim to generalise the findings presented. In one sense, generalisation is important for the current research project. From another perspective, however, *explanation* of the data rather than generalisation is an important aim of this research. Issues that apply to any subset of candidates for the IELTS test are worthy of attention given the high-stakes decisions for which the results are used.

The second limitation is specific to the SFL analysis, and is related to the first limitation. The 54 candidate scripts analysed using SFL responded to 26 different tasks. This means that the findings of the Appraisal analysis in particular could reflect differences in scripts attributable to task differences rather than L1 or band score. Due to the larger number of scripts analysed, and the nature of the analysis, this use of scripts responding to different tasks is not a limitation for the CTA analysis, as it provides for a 'spread' of discourse features in the sample analysed.

The third limitation relates to the selection of texts for the SFL analysis. Because the texts analysed in this part of the study all approximated 250 words in length (see introduction to Section 3), the effect of text length (Section 2.4) could have had an impact on the findings. A sample of texts in each band score with a greater variety in text length might have led to different results.

The fourth, and most important limitation of this research is the operationalisation of 'first language'. In the case of the 'European-based' L1 group, candidates actually came from four L1 backgrounds (Dutch, German, Portuguese, Romanian). For the Arabic L1 group, scripts were collected from candidates with the following nationalities: Egyptian, Jordanian, Kuwaiti, Lebanese, Libyan, Omani, Syrian, and nationals from the UAE. In terms of homogeneity, we can state definitively that the candidates in this L1 group identified Arabic as their first language on the form they completed to sit the IELTS test. Clearly though, there will be a great deal of difference in the reality of L1 use, and in the L1 itself in the different national and cultural contexts from which these candidates come. Turning to the Hindi L1 group, given the linguistic and cultural diversity of the sub-continent, the issue of diversity must apply also to any group of people assumed as being homogeneous on the basis of identifying as a speaker of Hindi as an L1.

## 4.3 Summary of findings, and implications

The findings (discussed also throughout Sections 2 and 3) and their implications include the following.

### 4.3.1 Differentiation according to L1

Many of the quantitative and discourse-analytic measures found little or no significant difference between scripts on the basis of candidate L1. However, some differences were found, and these suggest areas for further attention.

The quantitative analysis found that scripts from candidates with a European-based L1 measured higher on a number of quantitative measures (e.g. lexical diversity, word frequency, and reading ease) compared to scripts produced by test-takers from the other L1 backgrounds. The genre analysis found that European-based L1 candidates tended to use a more typical generic structure the higher they scored, whereas Arabic L1 candidates became more likely to use a variation on a typical generic structure the higher they scored. The Appraisal analysis found that Arabic L1 and European-based L1 candidates used Appreciation more than Judgement more than Affect. The same trend was observed with the Hindi L1 candidates, but there was a much smaller difference between the amount of Appreciation and Judgement used, and Hindi L1 candidates used Judgement more, and Appreciation less than the candidates of the other two L1 groups.

The differences in the findings of the quantitative analysis, in particular, suggest that L1 could be a potential factor affecting the band score candidates achieve. This raises concerns about some discoursal features (e.g. lexical diversity, word frequency, and cohesion) in the scripts produced by test-takers which are discussed further in Section 4.3.3.

### 4.3.2 Differentiation according to band score

The findings in relation to band score were mixed. In the quantitative analysis, the measures of Readability (Flesch Reading Ease) and Word Frequency were able to significantly differentiate scripts at bands 5, 6 and 7. Likewise, in the discourse analysis, the extent to which the genre of scripts was matched to task, and to which scripts were typical in their generic structure, was consistent with band score. The Appraisal analysis found differences between band 5 scripts (which used more Affect, and which used Judgement more than Appreciation) on one hand, and band 6 and 7 scripts (which used less Affect, and which used Appreciation more than Judgement) on the other. All these measures provide evidence for validity and reliability in Task 2 of the IELTS Academic Writing Test.

However, several quantitative measures did not differentiate scripts according to band score. Also, many facets of the Appraisal analysis also did not differentiate between band scores (e.g. in the source of Attitude, and the system of Engagement). The implications of this finding, together with those of the findings in Section 4.3.1, are discussed in Section 4.3.3.

### 4.3.3 Rating and reliability

Based on the descriptive and inferential results of the quantitative analysis and on the qualitative analysis, it seems that the complexity of the texts (lower readability index, greater lexical diversity, and lower-frequency words), generic structure (conventionality and match to task), and relative frequency of Attitude (Judgement and Appreciation over Affect) were more distinctive features of higher band scores than text cohesion (coreferentiality and index of all connectors). These findings provide support for the reliability and validity of the IELTS Academic Writing Task 2.

However, as discussed above, the quantitative analysis of the scripts showed that some scripts rated at the same band levels across the three L1 categories differed significantly in terms of some of their textual features, such as lexical diversity, cohesion, and word frequency. And the frequency of Engagement, an important resource in academic writing, was consistent across the band scores in the texts analysed for this study.

Thus, while some findings provide evidence for the reliability and validity of the scoring system of the IELTS Academic Writing Task 2, others highlight the fact that there is always an imperative to achieve higher reliability and validity in the scoring of scripts in a high-stakes test like the IELTS. Therefore, it seems crucial to ensure that Task 2 is designed in a manner that generates a representative sample of linguistic features consistently and sufficiently (see Section 4.3.5 below), and that IELTS trainers and examiners are sensitised to these features of test-takers' scripts.

The rating scales were not examined in this project, but on the basis of the findings summarised in this section, it appears that further research is warranted into the rating scales and their relation to the linguistic features of scripts (cf. Brown's 2006 study of the use of the rating scales in the Speaking Test, and Mickan's 2003 study of the use of rating scales in the General Training Writing Test). It would be informative to conduct further research that specifically investigates which discoursal features:

- *do* figure in the current band scales, but appear not to predict band score
- do not figure in the current band scales, but *do* appear to predict band score.

### 4.3.4 Genre and task difficulty

The 54 scripts analysed using genre theory were identified as belonging to genres which can be mapped topologically along two clines: single-perspective / multiple-perspective on one hand, and analytical / hortatory on the other. Most texts clearly belonged to the genres of exposition and discussion (e.g. Martin and Rose 2008; see also Mayor et al. 2007).

In mapping the genres and tasks topologically, it appeared that candidates could use (variations on) a hortatory discussion to meet the demands of almost any task, but that an analytical exposition would only meet the demands of a task which expected an analytical exposition. The washback effect of this is likely to be that candidates are prepared to write a hortatory discussion (giving a 'two-sided' argument, and including statements and/or a section about what 'should be' the case in addition to 'what is'), regardless of the instructions of a particular task, which is unlikely to be useful preparation for the demands of writing assignments in universities (Moore and Morton, 1999).

Research is warranted into the relation between task and genre in the IELTS Writing Test. First, greater understanding of the demands on item writers, who are required to be at once creative and 'scientific', while conforming to necessarily strict guidelines of structure and subject matter, would be beneficial. Would a further restriction on task structure (e.g. requiring all tasks in Task 2 to include a multiple-perspective argument, and a hortatory element) pose problems for item writers and task development or not (cf. Green and Hawkey 2012)?

Second, research into the washback effect of the current tasks (both Task 1 and Task 2) in terms of genre-related instruction would be valuable (cf. Mickan and Motteram 2008, pp 16-17). Classroom-based investigation of genre-focused preparation strategies and their relative success would inform understanding of the impact of the current approach to testing writing ability as operationalised in Task 2.

Third, regardless of the demands on item writers, and the washback effect, if variations in the genre requirements of Task 2 of the IELTS Academic Writing Test (changed since Mayor et al.'s 2007 study – see Section 3.1.15 above) equate with variations in the difficulty of the task, then there is an issue of reliability, and this requires investigation (cf. Mickan and Slater 2003). Identification of 'required' or 'expected' genres for a representative sample of tasks would be relatively straightforward, and this classification could form the basis for large-scale quantitative analysis of candidate performance (according to band score) on tasks requiring different genres.

### 4.3.5 Presence and absence of discoursal features in scripts

The Appraisal analysis identified a number of discoursal features that are important to academic writing, but that featured little in the scripts analysed. The finding that very little Attitude in the scripts comes from sources other than the author of the script is consistent with the finding that various areas of the Engagement system (those which are used to overtly project the voice of the author or of others) are used very little. It appears that the important skill of 'managing voices' in academic writing is little tested by IELTS Task 2 (cf. Moore and Morton 2003; Mayor et al. 2007).

There appears to be a major discrepancy between the 'management of voices' in the IELTS Academic Writing Test on one hand, and in the TLU domain on the other. The implications for the test are significant. The kinds of voices which are acceptable and valued in academic writing, and the ways in which authors are expected to introduce and evaluate such voices are highly conventionalised, and issues of validity and washback (and therefore impact) are of great concern in this area.

Further research is warranted which investigates the extent to which IELTS scripts vary from texts in the TLU domain (for instance, student assignments in English-medium universities). Content validity is crucial for the predictive power of the IELTS Academic Writing Test, and if the writing generated by the current Task 2 does not provide a representative sample of the linguistic features that figure in the Target Language Use domain, the nature of the task may need to be reconsidered (e.g. the inclusion of content that test-takers must integrate into their responses).

This may have ramifications for the structure of the IELTS Academic Writing Test (e.g. number of tasks, type of tasks), and perhaps even for the structure of the entire IELTS test if tasks which integrate writing, reading and listening were included. These potential ramifications are discussed in Section 4.5.

### 4.3.6 Handwritten scripts

The use of handwriting in the IELTS test was not an object of study in this research. However, it quickly became an issue in the 'transcription' stage of the process, something we had considered would be relatively straightforward.

Handwriting of scripts was found to be a problematic aspect of the IELTS Writing Test. Scripts vary widely in terms of their legibility, and handwriting allows candidates to 'fudge' some aspects of writing (e.g. punctuation, capitalisation, spelling, paragraphing). While many of the transcription decisions were relatively minor matters, others had ramifications for grammatical and discursive understanding of the scripts (cf. the discussion of error analysis and reliability in Mayor et al. 2007). It became clear that many candidates could use the 'flexibility' of handwriting to their advantage, in a way that would not be acceptable in submitting academic assignments (which are now usually required to be submitted typed in many or most English-medium universities).

Whether a result of individual style or intentional ambivalence, the ambiguity of aspects of handwriting in many scripts poses threats for the reliability of the test. (Does the quality of a candidate's handwriting affect a rater's judgement? Is one rater more accustomed to a style of handwriting, or simply more patient, than another?)

Brown (2003) found that handwriting (as opposed to typing) responses advantaged candidates on Task 2 of the IELTS Writing Test, and further that poor handwriting gave more of an advantage than relatively legible handwriting. In contrast, Weir, O'Sullivan, Yan and Bax (2007) found that handwritten versus computer input had no significant difference in test-taker performance on Task 2 of the IELTS Writing Test, and found it "highly plausible that the two versions [i.e. computer-input and handwritten input] were testing the same language ability" (p 24). But they also concluded that the method of input could lead to problems in reliability, and suggested further research in this area (pp 25-6).

Personal computers are now widespread in schools and households, including in many developing nations. The availability of technology is changing quickly for the candidature of the IELTS; the availability and pervasiveness of computers in the TLU domain (which, for the IELTS Academic Writing Test, is in practice English-medium universities) is changing; the resources available for testing organisations like IELTS are changing.

Research is warranted into the relation between changes in the availability and use of technology in writing practices, and the needs of users of the IELTS test (something that has not, to our knowledge, been researched).

There may be, for example, important differences between the needs of users of the IELTS Academic Writing module (e.g. English-medium universities) and the IELTS General Training (GT) Writing module (e.g. immigration decision makers in national governments). To illustrate, anecdotally, most universities require typed assignments from students. Users of the GT module might or might not have similar expectations of 'computer literacy' in the writing practices in their institutional contexts. Regardless, if the writing practices tested in the IELTS are different from those in the TLU domain, the validity of a handwritten test would be in question. This issue can only be resolved through research.

Research is also warranted into the test-takers' contexts. How much access do they have to computers, and in what contexts do they use computers to write? Can they type? How much variation is there according to national and/or economic background? Weir et al. (2007) found that the subjects of their study were familiar with computers, but that some variables in their social backgrounds (i.e. accessibility of public computers and frequency of word processing activity) were correlated with differences in performance. It is over five years since that research was published – much may have changed in the interim.

There are now people of an age suitable to sit the IELTS Test who rarely use handwriting. For instance, until recently, all Australian school students received a laptop in Year 9 so much of their academic reading and writing was done on computer and not with pen and paper. How many IELTS candidates have a similar background, and to what extent are they disadvantaged by sitting a handwritten test (just as some candidates might be disadvantaged by sitting a typed-input test)?

In terms of validity and reliability, typed input of scripts could standardise the medium, allowing examiners to focus on the discourse of scripts rather than handwriting (see Brown 2003). It would also more closely match the medium of the TLU domain for the Academic Writing module.

The use of typed input would also allow for computerised text analysis (CTA) of scripts to be conducted in tandem with human rating, enhancing reliability (see Section 4.3.3). A text analysis program like Coh-Metrix or an e-rater could be used to complement human ratings. Such an approach would generate a strong body of evidence for the reliability of scoring, and may also help to identify which specific areas, if any, pose problems for reliability in the scoring of the test.

In terms of practicality, the financial cost of the switch from handwriting to typed input would be significant in the short term, but in the long term the savings in the production, distribution, storage, and destruction of paper, and resulting savings in administration of results may even make the test more economical – a potential saving which could be passed on to test-takers and result in a positive outcome for access to the test, and therefore for the equity of the test.

IELTS test-takers come from a wide variety of social, cultural, and economic backgrounds, so equity in access to the test is both very important, and very difficult to provide. However, in terms of reliability and validity, it seems that the issues related to testing academic writing with the computer-based IELTS (Blackhurst 2005; Green and Maycock 2004; Maycock and Green 2005) are, at the very least, worth further investigation. In terms of the rapidly evolving social and technological context of the 21st century, a move to computer-based testing of writing in the IELTS test appears inevitable.

But how and when the computer-based IELTS is rolled out needs to be informed also by research into the social contexts of test users and test-takers.

## 4.4 Recommendations

On the basis of the findings, we recommend the following.

1. IELTS conduct research into the relations between the rating scales and the linguistic features of texts in the Writing module (see Section 4.3.3).
2. IELTS conduct research on the relations between genre, task difficulty, task development and candidate preparation. Such research could also be conducted in relation to Task 1, and could involve examination of the genres in the TLU domain (see Section 4.3.4).
3. IELTS conduct research on the extent to which the 'management of voices' in academic writing is suitably and adequately tested in the IELTS Academic Writing Test, in order to determine whether the introduction of one or more 'integrated' tasks that would require candidates to integrate provided sources into their response is warranted (see Section 4.3.5).
4. IELTS seriously investigate using typed input for the IELTS Writing Test. Such investigation would include research into the issues surrounding handwritten and typed input (see Section 4.3.6).

In the medium-long term, these recommendations could lead to changes in the rating scales for the IELTS Writing Test (Recommendation 1), the design of Task 1 and 2 of the Writing Test (Recommendations 2 and 3), the design of the Writing Test and the entire IELTS Test (Recommendation 3) and in the administration and implementation of the entire IELTS test (Recommendation 4). Nevertheless, on the basis of our findings, we believe investigation along the lines of these recommendations is warranted.

## 4.5 Conclusion

The recommendations above outline a relatively modest research agenda. But it is one that could have potentially far-reaching implications for the IELTS test.

A shift from handwritten to typed input would be a major change for the way the test is conducted, and would also have fundamental institutional implications for the organisations involved in the running of the IELTS test. At the same time, it would offer the opportunity to take advantage of information technologies that could improve the practicality, reliability, and validity of the Test.

Any move to reconsider the task structure of Task 2 of the Writing Test and/or the band scales would need to be done in conjunction with a consideration of Task 1, and the introduction of an integrative section or sections would involve a reconsideration of the entire structure of the IELTS, not just the Writing Test.

None of these decisions would be simple or easy. But IELTS has shown a willingness to change with the times, most recently with major revisions of rating scales in the Writing Test, and of the entire Speaking Test. The social environment in which the IELTS test operates is changing, and our understanding of the nature of language use in academic and professional contexts has moved a long way from when the IELTS test was first conceived. Just as the shift to communicative language testing involved a move away from the 'discrete-item' understanding of language and the 'discrete-item' approach to testing it, so testing in the 21st century seems destined to move on from the 'pen-and-paper' understanding of writing and the 'pen-and-paper' approach to testing it, and from the 'four skills' understanding of language and the 'four skills' approach to testing it.

IELTS was at the forefront of the last major shift in international standardised testing. An appropriately targeted research agenda, and a willingness to act on the findings could keep it at the forefront through the next shift.

# 5 ACKNOWLEDGEMENTS

The authors would like to thank the following people:

- David Caldwell and Lai Ping Florence Ma, research assistants on the project
- Sumin Zhao and Lai Ping Florence Ma for transcription of the scripts
- Jenny Osborne for her efficiency, professionalism and tireless assistance throughout every stage of the project
- the three anonymous reviewers for their constructively critical feedback on earlier submissions, which has led to an improved report.

In addition, we would like to thank IELTS Australia for funding this project.

## 6 REFERENCES AND BIBLIOGRAPHY


Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford University Press, Oxford.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice.* Oxford University Press, Oxford.

Banerjee, J., Franceschina, F., Smith, A. M. (2007). 'Documenting features of written language production typical at different IELTS band score levels' in *IELTS Research Reports*, Vol. 7, pp 241-309. IELTS Australia Pty Ltd, Canberra and British Council, London.

Bateman, J. A. (2008). *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. Palgrave Macmillan, Hampshire.

Bednarek, M. (2007). 'Polyphony in Appraisal: Typological and topological perspectives' in *Linguistics and the Human Sciences, 3*(2), pp 107-136.

Blackhurst, A. (2005). 'Listening, Reading and Writing on computer-based and paper-based versions of IELTS' in *University of Cambridge ESOL Examinations Research Notes, 21*, pp 14-17.

Brown, A. (2003). 'Legibility and the rating of second language writing: An investigation of the rating of handwritten and word-processed IELTS Task 2 essays' in *IELTS Research Reports*, Vol. 4, pp 132-151. Ed R Tulloh, IELTS Australia Pty Limited, Canberra.

Brown, A. (2006). 'An examination of the rating process in the revised IELTS Speaking Test' in *IELTS Research Reports*, Vol. 6, pp 1-30. IELTS Australia Pty Limited, Canberra.

Brown, J.D. (2004). 'Performance assessment: Existing literature and directions for research' in *Second Language Studies, 22*(2), pp 91-139.

Christie, F. (1997). 'Curriculum macrogenres as forms of initiation into a culture' in F. Christie and J. R. Martin (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp 134-160). Continuum, London.

Celestine, C., & Su Ming, C. (1999). 'The effect of background disciplines on IELTS scores' in *IELTS Research Reports*, Vol. 2, pp 36-61. IELTS Australia Pty Limited, Canberra.

Cobb, T. (2003). VocabProfile, The Compleat Lexical Tutor. From http://www.lextutor.ca

Cobb, T. (2010). 'Learning about language and learners from computer programs' in *Reading in a Foreign Language, 22*(1), pp 181-200. Retrieved 16 April 2010 from: http://nflre.hawaii.edu/rfl

Coffin, C. (2004). 'Arguing about how the world is or how the world should be: The role of argument in IELTS tests' in *Journal of English for Academic Purposes, 3*(3), pp 229-246.

Coffin, C., & Hewings, A. (2005). 'IELTS as preparation for tertiary writing: distinctive interpersonal and textual strategies' in L. Ravelli & R. Ellis (Eds.), *Analysing Academic Writing: Contextualized Frameworks* (pp 153-171). Continuum, London.

Coffin, C. (2006). *Historical discourse: The language of time, cause and evaluation*. Continuum, London.

Cotton, F., Wilson, K. (2008). 'An investigation of examiner rating of coherence and cohesion in the IELTS Academic Writing Task 2' in *IELTS Research Reports Volume 12*. Retrieved 15 Dec. 2011 from: http://www.ielts.org/pdf/Vol12_Report6.pdf, IELTS Australia Pty Ltd, Canberra and British Council, London.

Crossley, S. A., & McNamara, D. S. (2009). 'Computational assessment of lexical differences in L1 and L2 writing' in *Journal of Second Language Writing, 18*, pp 119–135.

Crossley, S. A., & McNamara, D. S. (2010). 'Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication' in *Journal of Research in Reading, 35*(2), pp 115-135.

Crossley, S.A., & McNamara, D.S. (2011). 'Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing' in *International Journal of Continuing Engineering Education and Life-Long Learning*, *21*(2/3), pp 170–191.

Crossley, S.A., Salsbury, T., & McNamara, D.S. (2011). Predicting the proficiency level of language learners using lexical indices. *Language Testing, 4*, pp 561-580.

Crossley, S.A., Salsbury, T., McNamara, D.S., & Jarvis, S. (2011). 'Predicting lexical proficiency in language learner texts using computational indices' in *Language Testing, 28*(4) pp 561–580.

Crossley, S.A., Weston, J.L., McLain Sullivan, S.T., & McNamara, D.S. (2011). 'The development of writing proficiency as a function of grade level: A linguistic analysis' in *Written Communication, 28*(3), pp 282-311.

Ferris, D. (1994). 'Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency' in *TESOL Quarterly, 28*, pp 414–420.

Frase, L., Faletti, J., Ginther, A. & Grant, L. (1997). Computer analysis of the TOEFL test of written English (TOEFL Research Report no. 64). Educational Testing Service, Princeton, NJ.

Gerot, L., & Wignell, P. (1994). *Making sense of functional grammar: An introductory workbook*. Queensland: Antipodean Educational Enterprises.

Graesser, A.C., McNamara, D.S., & Kulikowich, M. (2011). 'Coh-Metrix: Providing multilevel analyses of text characteristics' in *Educational Researcher, 40*(5), pp 223-234.

Graesser, A., McNamara, D., Louwerse, M., & Cai, Z. (2004). 'Coh-Metrix: Analysis of text on cohesion and language' in *Behavioral Research Methods, Instruments, and Computers, 36,* pp 193-202.

Green, A., & Hawkey, R. (2012). 'An empirical investigation of the process of writing Academic Reading test items for the International English Language Testing System' in *University of Cambridge ESOL Examinations Research Notes, 11*, pp 1-100.

Green, T., & Maycock, L. (2004). 'Computer based IELTS and paper based versions of IELTS' in *University of Cambridge ESOL Examinations Research Notes, 18*, pp 3-6.

Green, A., Unaldi, A., & Weir, C. (2010). 'Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading' in *Language Testing, 27*(2), pp 191-211.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman, London.

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

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed.). Arnold, London.

Heydari, P., & Riazi, M. (2012). 'Readability of texts: Human evaluation vs. computer index' in *Mediterranean Journal of Social Sciences, 3*(1), pp 177-191.

Hood, S. E. (2010). *Appraising research: Evaluation in academic writing*. London: Palgrave Macmillan.

Hood, S. E. (2004). 'Managing attitude in undergraduate academic writing: A focus on the introductions to research reports' in L. Ravelli & R. Ellis (Eds.), *Analysing Academic writing: Contextualized frameworks* (pp 24-44). Continuum, London.

Hood, S. E., & Martin, J. R. (2007). 'Invoking attitude: the play of graduation in appraising discourse' in R. Hasan, C. M. I. M. Matthiessen & J. J. Webster (Eds.), *Continuing discourse on language: A functional perspective* (Vol. 2, pp 739-764). Equinox, London.

Huang, J., & Mohan, B. (2009). 'A functional approach to integrated assessment of teacher support and student discourse development in an elementary Chinese program' in *Linguistics and Education, 20*(1), pp 22-38.

Hughes, A. (2003). *Testing for language teachers.* 2nd ed. Cambridge University Press, Cambridge.

Hyland, K. (2000). *Disciplinary discourses : social interactions in academic writing*. Longman, Harlow.

Hyland, K. (2006). *English for Academic Purposes: An advance resource book*. Routledge, London.

IELTS. (2009). 'IELTS Test-taker Performance 2009' in *Research Notes, 40*, pp 26-29.

Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage, Thousand Oaks, CA.

Kane, M., Crooks, T., & Cohen, A. (1999). 'Validating measures of performance' in *Educational Measurement: Issues and Practice, 18,* pp 5-17.

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Hodder Arnold, London.

Laufer, B., & Nation, P. (1995). 'Vocabulary size and use: Lexical richness in L2 written production' in *Applied Linguistics*, *16*(3), pp 307–322.

Leung, C., & Mohan, B. (2004). 'Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse' in *Language Testing, 21*(3), pp 335-359.

Malvern, D., & Richards, B. (1997). 'A new measure of lexical diversity' in A. Ryan & A. Wray (Eds.), *Evolving models of language* (pp 58–71). Multilingual Matters, Clevedon, UK.

Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave MacMillan, Houndmills, England.

Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture*. Equinox, London.

Martin, J. R., & Rose, D. (2012). *Learning to write, Reading to learn*. Equinox, Sheffield.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan, Hampshire.

Maycock, L., & Green, T. (2005). 'The effects on performance of computer familiarity and attitudes towards CB IELTS' in *University of Cambridge ESOL Examinations Research Notes, 20*, pp 3-8.

Mayor, B., Hewings, A., North, S., Swann, J., & Coffin, C. (2007). 'A linguistic analysis of Chinese and Greek L1 scripts for IELTS Academic Writing Task 2' in L. Taylor & P. Falvey (Eds.), *IELTS collected papers: Research in speaking and writing assessment* (pp 250–315). Cambridge University Press, Cambridge.

McCarthy, P. M., Lehenbauer, B. M., Hall, C., Duran, N. D., Fujiwara, Y., & McNamara, D. S. (2007). 'A Coh-Metrix analysis of discourse variation in the texts of Japanese, American, and British Scientists' in *Foreign Languages for Specific Purposes, 6*, pp 46–77.

McKee, G., Malvern, D., & Richards, B. (2000). 'Measuring vocabulary diversity using dedicated software' in *Literary and Linguistic Computing,15*(3), pp 323-337.

McNamara, D.S., Crossley, S.A., & McCarthy, P.M. (2010). 'Linguistic features of writing quality' in *Written Communication, 27*(1), pp 57-86.

McNamara, D.S., Louwerse, M.M., Cai, Z., & Graesser, A. (2005). Coh-Metrix version 1.4. Retrieved from: http://cohmetrix.memphis.edu/cohmetrixpr/index.html draft.docx

McNamara, D.S., Louwerse, M.M., McCarthy, P.M., & Graesser, A.C. (2010). 'Coh-Metrix: Capturing linguistic features of cohesion' in *Discourse Processes, 47*, pp 292-330.

Mickan, P. (2003). 'An investigation into language descriptors for rating written performance' in *IELTS Research Reports*, Vol. 5, pp 126-155. IELTS Australia Pty Ltd, Canberra and British Council, London.

Mickan, P., & Motteram, J. (2008). 'An ethnographic study of classroom instruction in an IELTS preparation program' in *IELTS Research Reports*, Vol. 8, pp 1-26. Ed J Osborne, IELTS Australia, Canberra,

Mickan, P., & Slater, S. (2003). 'Test analysis and the assessment of academic writing' in *IELTS Research Reports*, Vol. 4, pp 59-88. Ed R Tulloh, IELTS Australia Pty Limited, Canberra.

Mohan, B., & Slater, T. (2004). 'The evaluation of causal discourse and language as a resource for meaning' in J. A. Foley (Ed), *Language, education and discourse: Functional approaches* (pp 255-269). Continuum, London.

Moore, T., & Morton, J. (1999). 'Authenticity in the IELTS academic module writing test: A comparative study of Task 2 items and university assignments' in R. Tulloh (Ed.), *IELTS Research Reports, Vol 2*, pp 64-106. IELTS Australia Pty Ltd, Canberra.

O'Loughlin, K., & Wigglesworth, G. (2003). 'Task design in IELTS Academic Writing Task 1: The effect of quantity and manner of presentation of information on candidate writing' in *IELTS Research Reports*, Vol. 4, pp 89-131. Ed R Tulloh, IELTS Australia Pty Limited, Canberra.

Pallant, J. (2007). *SPSS survival manual: a step by step guide to data analysis using SPSS for Windows* (Third ed.). Open University Press, Maidenhead.

Perrett, G. (1997). 'Discourse and rank: The unit of transaction in the oral interview' in *Australian Review of Applied Linguistics, 20*, pp 1-20.

Rasti I. (2009). 'Iranian Candidates' Attitudes towards IELTS' in *Asian EFL Journal 11, 3.* Retrieved from http://www.asian-efl-journal.com/September_2009_ir.php

Ravelli, L. (2004). 'Signalling the organization of written texts: Hyper-Themes in management and history essays' in L. Ravelli & R. Ellis (Eds.), *Analysing academic language: Contextualized frameworks* (pp 104-130). Continuum, London.

Ravelli, L. J., & Ellis, R. A. (Eds.). (2004). *Analysing academic writing: Contextualized frameworks*. Continuum, London.

Scott, M. (2006). Oxford WordSmith Tools 4.0. Retrieved from: http://www.lexically.net/downloads/version4/html/index.html

Shaw, S, & Falvey, P. (2008). 'The IELTS writing assessment revision project: Towards a revised rating scale' in *Research Reports*, Volume 1, January 2008.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Lawrence Erlbaum Associates, Publishers, Mahwah, NJ.

Taylor, L. (2004). 'Second language writing assessment: Cambridge ESOL's ongoing research agenda' in *Research Notes, 16*, pp 2-3.

Veel, R. (1997). 'Learning how to mean - scientifically speaking: Apprenticeship into scientific discourse in the secondary school' in F. Christie & J. R. Martin (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp 161-195). Continuum, London and New York.

Weir, C. (1993). *Understanding and developing language tests*. Prentice Hall, Hertfordshire.

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

Woodward-Kron, R. (2005). 'The role of genre and embedded genres in tertiary students' writing' in *Prospect, 20*(3), pp 24-41.

Wray, A., & Pegg, C. (2005). 'The effect of memorized learning on the writing scores of Chinese IELTS test-takers' in *IELTS Research Reports*, Vol. 9, pp 191-216. IELTS Australia, Canberra and British Council, London.

Yu, G. (2010). 'Lexical density in writing and speaking task performance' in *Applied Linguistics, 31*(2), pp 236-259.

Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). 'Relationship between lexical competence and language proficiency – variable sensitivity' in *Studies in Second Language Acquisition*, *27*(4), pp 567–595.