

An examination of discourse competence at different proficiency levels in IELTS Speaking Part 2

Authors: Noriko Iwashita, *The University of Queensland*
Claudia Vasquez, *The University of Queensland*

Grant awarded: Round 19, 2013

Keywords: “IELTS, Speaking test, discourse competence, cohesion and coherence in oral performance, IELTS Speaking Band Descriptors, test-taker performance, oral proficiency, communicative competence, L2 proficiency”

Abstract

This study investigates characteristics of test-taker performance on IELTS Speaking Part 2 at Levels 5, 6 and 7 focusing on test-takers’ strategies for producing comprehensible, high-quality speech with various devices. The features of performance identified in the study were co-referenced with the IELTS Speaking Band Descriptors.

The study investigated the features of discourse competence observed in IELTS Speaking Part 2 performance and how the distinctive features of performance correlate to the IELTS Speaking Band Descriptors. Scholars attempted to elaborate the notion of discourse competence as a part of their pursuit for further understanding of communicative competence (eg, Bachman & Palmer 1996; Chalhoub-Deville 2003; Purpura 2008). While there seems to be consensus on the importance of greater understanding of discourse competence as a means of further understanding communicative language ability and L2 proficiency in general, a detailed study of discourse competence appears to have been somewhat neglected (Kormos 2011; Purpura 2008), particularly into speaking performance. Discourse competence is one of the four categories identified in the IELTS Speaking Band Descriptor.

In order to fill this gap, the current study undertook detailed examination of test-taker oral discourse at three proficiency levels. The transcribed 58 speech samples (18–20 examples at each level) of IELTS Speaking Part 2 were analysed both quantitatively and qualitatively. The features of discourse competence analysed in the current study included both cohesive devices (use of reference, ellipsis and substitution, lexical cohesion, conjunctions) and coherence devices (ie, text generic structure and theme-rheme development).

The in-depth analysis revealed that some features of discourse (eg, use of a wider range of conjunctions, more accurate use of referential expressions) were more distinctively observed in the higher-level test-taker performance than the lower level test-takers, but other features (eg, ellipsis and substitution, use of reference) were not clearly distinguished across the levels.

These findings contribute to further understanding of the nature of oral proficiency; they also supplement IELTS Speaking Band Descriptors with features of test-taker discourse empirically identified in the test-taker performances. Furthermore, the results will inform language teachers of characteristics of oral proficiency to be targeted in L2 instruction.

Publishing details

Published by the IELTS Partners: British Council, Cambridge English Language Assessment and IDP: IELTS Australia © 2015. This online series succeeds *IELTS Research Reports Volumes 1–13*, published 1998–2012 in print and on CD. This publication is copyright. No commercial re-use. The research and opinions expressed are of individual researchers and do not represent the views of IELTS. The publishers do not accept responsibility for any of the claims made in the research. Web: www.ielts.org

AUTHOR BIODATA

Noriko Iwashita

Noriko Iwashita is a senior lecturer in Applied Linguistics at The University of Queensland. Her research interests include task-based assessment, and cross-linguistic investigation of four major language traits and the interfaces of language assessment and SLA. Her work has appeared in *Language Testing*, *Applied Linguistics*, *Language Learning* and *Studies in Second Language Acquisition*.

Claudia Vasquez

Claudia Vasquez is a PhD candidate at The University of Queensland. The topic of her thesis is the effect of planning time on discursive aspects of learner performance. Her research interests include characteristics of discourse competence, analysis of learner performance, task-based teaching and assessment, and second language acquisition in general. Prior to coming to Australia, Claudia was a lecturer in Linguistics and EFL at a university in Chile.

ACKNOWLEDGEMENTS

We would like to express thanks to the IELTS partners for their support and allowing us to use the data for the research. We are also very thankful for Professor India Plough at Michigan State University, Professor Liz Hamp-Lyons at the University of Bedfordshire, and two anonymous IELTS reviewers for their helpful comments on the draft research report.

IELTS Research Program

The IELTS partners – British Council, Cambridge English Language Assessment and IDP: IELTS Australia – have a longstanding commitment to remain at the forefront of developments in English language testing.

The steady evolution of IELTS is in parallel with advances in applied linguistics, language pedagogy, language assessment and technology. This ensures the ongoing validity, reliability, positive impact and practicality of the test. Adherence to these four qualities is supported by two streams of research: internal and external.

Internal research activities are managed by Cambridge English Language Assessment's Research and Validation unit. The Research and Validation unit brings together specialists in testing and assessment, statistical analysis and item-banking, applied linguistics, corpus linguistics, and language learning/pedagogy, and provides rigorous quality assurance for the IELTS test at every stage of development.

External research is conducted by independent researchers via the joint research program, funded by IDP: IELTS Australia and British Council, and supported by Cambridge English Language Assessment.

Call for research proposals

The annual call for research proposals is widely publicised in March, with applications due by 30 June each year. A Joint Research Committee, comprising representatives of the IELTS partners, agrees on research priorities and oversees the allocations of research grants for external research.

Reports are peer reviewed

IELTS Research Reports submitted by external researchers are peer reviewed prior to publication.

All IELTS Research Reports available online

This extensive body of research is available for download from www.ielts.org/researchers.

INTRODUCTION FROM IELTS

This study by Noriko Iwashita and Claudia Vasquez of the University of Queensland 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 or 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 joint-funded research program started in 1995; over 100 empirical studies having 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://www.cambridgeenglish.org/silt>), and in *IELTS Research Reports*. Since 2012, in order to facilitate timely access, individual research reports have been made available on the IELTS website immediately after completing the peer review and revision process.

In this report, Iwashita and Vasquez considered the notion of discourse competence and investigated candidates' use of a range of relevant features in Part 2 of the IELTS Speaking test. To do this, they employed a combination of quantitative and qualitative analyses to discover differences in the spoken performances of candidates at IELTS bands 5, 6 and 7.

It should be noted that the candidates in the study were divided according to their overall Speaking scores, whereas the analysis focused only on their performance in Part 2 of the test. In any event, statistically significant differences were found in the use of comparative conjunctions, of lexical cohesion such as hyponymy and repetition, and in the accurate use of referential expressions. While these are positive findings, it would in some ways be quite disappointing if these were the only discursive features that distinguished candidates at different IELTS bands. However, those were not in fact the only ways in which weaker and stronger candidates were found to differ.

The qualitative component of the authors' work found that "the compliance index of the text generic structure show clear differences according to the band levels, even though the statistical analysis reveals no significant difference across the levels". In addition, analyses of theme-rheme development patterns revealed other differences, with higher band candidates producing speech characterised by higher levels of cohesion resulting in richer content.

This study shows once again the usefulness of using multiple and mixed-methods in research. As automated text analysis tools become more widely available, producing hundreds of statistics and indices regarding pieces of texts, it can become too easy to just depend on the numbers produced and draw conclusions that are potentially misleading. More careful analysis, as employed in this study, can show that there is more going on in texts than would initially appear, and that examiners are apparently able to perceive these.

Indeed, as previously noted, the researchers focused in this study on Part 2 of the Speaking test, the 'long turn', where candidates speak uninterrupted about a particular topic. But the test also has a Part 3 which, for many examiners, is the part of the test that really allows them to distinguish higher ability candidates. In this part of the test, examiners and candidates interact about the given topic more broadly, with greater unpredictability. It would not be unreasonable to suppose that discourse competence is even more crucial in that part of the test, and would exhibit itself in myriad, more complex and interesting ways. That, of course, is the subject of future research.

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

CONTENTS

1	INTRODUCTION	6
2	BACKGROUND TO THE STUDY	6
2.1	Communicative competence	6
2.2	Discourse competence	6
2.2.1	Discourse and text	6
2.2.2	Discourse competence	7
2.3	Investigation of discourse competence in learner performance	8
3	RESEARCH QUESTION	8
4	METHODOLOGY	8
4.1	Data	8
4.2	Method	8
4.3	Analysis	9
4.3.1	Cohesion	9
4.3.1.1	Conjunction	9
4.3.1.2	Reference	10
4.3.1.3	Lexical cohesion	12
4.3.2	Coherence	12
4.3.2.1	Text generic structure	12
4.3.2.2	Theme–rheme development	14
4.3.3	Lexical richness	17
5	RESULTS	18
5.1	Cohesion analysis	18
5.1.1	Conjunction	18
5.1.2	Reference	21
5.1.3	Lexical cohesion	23
5.2	Coherence analysis	26
5.2.1	Text generic structure	26
5.2.2	Theme–rheme development	28
5.3	Lexical richness	35
5.4	Summary of the results	36
5.5	Co-reference with the IELTS Speaking Band Descriptor	38
5.5.1	Level 7	38
5.5.2	Level 6	38
5.5.3	Level 5	38
6	DISCUSSION AND CONCLUSION	38
	REFERENCES	41
	APPENDIX 1: TEST-TAKERS' L1 AND LEVEL	43
	APPENDIX 2: DETAILED RESULTS OF LEXICAL RICHNESS	44

List of tables

Table 1: Example of analysis of referential expression (Level 6, ID606).....	10
Table 2: Example of analysis of referential expression (Level 6, ID263).....	11
Table 3: Descriptive statistics of use of conjunctions (frequency, per 100 words)	19
Table 4: Results of Kruskal–Wallis analysis (conjunction use).....	19
Table 5: Distribution of conjunction use (%)	19
Table 6: Descriptive statistics of use of referential expressions (frequency, per 100 words) and percentage of accurate use	21
Table 7: Example of Level 7 performance (excerpt) (Level 7 ID709)	22
Table 8: Example of Level 5 performance (excerpt) (Level 5 ID226)	23
Table 9: Example of Level 5 performance (excerpt) (Level 5 ID501)	23
Table 10: Descriptive statistics of use of lexical cohesions (frequency, per 100 words)	24
Table 11: Results of Kruskal-Wallis analysis (lexical cohesion)	24
Table 12: Distribution of lexical cohesion (%).....	24
Table 13: Descriptive statistics of generic structure compliance index scores	26
Table 14: Descriptive statistics of quantity of text in each proficiency group.....	28
Table 15: Theme–rheme patterns in each proficiency group	29
Table 16: Descriptive statistics of lexical richness measures	35
Table 17: Results of one-way ANOVA analysis (lexical richness).....	35
Table 18: Summary of the results of quantitative analysis	37

List of figures

Figure 1: Example of theme–reiteration/constant theme pattern 1	15
Figure 2: Example of theme–reiteration/constant theme pattern 2.....	15
Figure 3: Example of combination of theme–reiteration and zigzag/linear theme pattern 1.....	16
Figure 4: Example of combination of theme–reiteration and zigzag/linear theme pattern 2.....	16
Figure 5: Example of combination of theme–reiteration and zigzag/linear theme pattern 3.....	17
Figure 6: Distribution of conjunction use.....	19
Figure 7: Distribution of lexical cohesion	24
Figure 8: Theme–rheme development (Reiteration/constant theme pattern) (Level 5-1).....	29
Figure 9: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 6-1)	30
Figure 10: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 7-1)	31
Figure 11: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 5-2)	32
Figure 12: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 6-2)	33
Figure 13: Theme–rheme development (combination of theme–reiteration/constant theme and zigzag/linear theme patterns) (Level 7-1)	34
Figure 14: Distribution of the four word categories	36

1 INTRODUCTION

This study aimed to identify features of test-taker oral performance at three different levels of the IELTS Speaking Test Band Descriptors by focusing on discourse competence. Thus, this research examined the construct of discourse competence observed in test-taker performance and investigated how distinctive features of performance correlate to the IELTS Speaking Test Band Descriptors.

This examination of discourse competence was undertaken as scholars attempted to elaborate the concept of communicative competence in the learning, teaching, and assessment of second languages (L2) (eg Bachman 1990; Bachman & Palmer 1982, 1996; Canale 1983; Canale & Swain 1980; McNamara 1996; Celc -Murcia 2008; Celc -Murcia, Dornyei & Thurrel 1995; Purpura 2008). The search for deeper understanding has prompted the development of various frameworks to provide theoretical foundations of the nature of communicative competence. These various theoretical proposals conceptualise communicative competence as a composite of different sub-competencies which explain the degrees of learner mastery of a L2.

Among the sub-competencies suggested to constitute communicative competence, discourse competence has been considered to be at the core of the knowledge required to communicate in a L2 (see Bachman 1990; Bachman & Palmer 1996; Canale 1983; Celc -Murcia et al. 1995; Celc -Murcia 2008). While there seems to be consensus on the importance of greater understanding of discourse competence as a means for further clarity of communicative ability in general, a detailed study of discourse competence has been somewhat neglected (Kormos 2011; Purpura 2008; van Lier 1989). This is particularly the case for speaking. To fill this gap, the current study examined discourse competence through the detailed examination of test-taker performance at given levels of proficiency.

2 BACKGROUND TO THE STUDY

2.1 Communicative competence

The conceptualisation of communicative competence has been at the centre of an ongoing debate, with multiple research efforts attempting to elucidate comprehensively what it means to know a language. Over the years, a model of communicative competence has been gradually elaborated and articulated to suit specific purposes (eg assessment, pedagogy) by various scholars (eg Bachman 1990; Bachman & Palmer 1982, 1996; Canale 1983; Canale & Swain 1980; Celc -Murcia 2008; Celc -Murcia, Dornyei & Thurrel 1995). Although no agreement has been reached on the definition of communicative competence, researchers seem to agree that communicative competence is a multi-componential phenomenon constituted by a gamut of sub-competencies, the articulation and interactions of which explain the degree of a learner's mastery of a language.

Thus, the notion of communicative competence is believed to minimally encompass dimensions relating to the following:

- knowledge of how to arrange formal units of language into unified units of discourse
- knowledge and understanding of the socio-cultural and communicative context in which communication takes place
- knowledge of how to interact successfully with an interlocutor in a communicative exchange in a L2.

Bachman (1990) and Bachman and Palmer (1996) conceptualised communicative competence as "a capacity that enables language users to create and interpret discourse" (p 33). Their model encompasses two main components: language competence, or knowledge of language; and strategic competence, a set of metacognitive strategies that control the manner in which language users interact with the characteristics of the language use situation.

This model asserts that language knowledge includes two major categories, namely organisational and pragmatic knowledge. In turn, these two main components break down into a number of sub-components addressing a wide range of language dimensions. On the one hand, organisational knowledge refers to how utterances or sentences and texts are organised, and it further comprises grammatical and textual knowledge. Pragmatic knowledge, on the other hand, refers to how utterances or sentences and texts are related to the communicative goals of language users and to the features of the language use setting; therefore, this knowledge is made up, in turn, of functional and sociolinguistic knowledge. Although this model has been regarded as an elaborate and comprehensive representation of language proficiency (eg Alderson & Banerjee 2002; Chalhoub-Deville 1997), it is not clear how each component of knowledge contributes to communicative language ability, and therefore it is hard to implement in assessment practice (McNamara 1996). With regard to spoken communication, the concept of communicative competence has been expanded to encompass 'interactional competence' (eg Hall & Doehler 2011; Young 2011).

2.2 Discourse competence

2.2.1 Discourse and text

Discourse is usually contrasted with the notion of text. However, these terms have been used interchangeably. Notwithstanding this, some scholars have distinguished between the connotations the two terms have and the phenomena to which they refer. Widdowson (1984), for example, elaborates on the discourse-text dichotomy by arguing that "discourse is a communicative process by means of interaction. Its situational outcome is a change in a state of affairs. Its linguistic product is text" (p 100). He further elaborated on the view that a prerequisite of communication is the negotiation of meaning through interaction, a process he defined as discourse: "the process whereby language users negotiate a 'reciprocity

of perspectives' for the conveyance of information and intention" (p 100). Christiansen (2011) argued that, from a non-specialised perspective (ie, in non-linguistic and non-semiotic circles) 'text' is "sometimes used for examples of written language and discourse for the spoken" thus basing the discussion on a "distinction between medium and channel" (p 34). However, he (2011) also noted that this distinction is simplistic and proposed a differentiation based on "text for the form and discourse for the content" (p 34).

From the perspective of functional-systemic linguistics, it is the notion of text that encompasses "all forms of oral and written communication" (Eggins 1994, p 85). Thus, a text is "any passage (of language), spoken or written, of whatever length, that [...] forms a unified whole" (Halliday & Hasan 1976, p 1). To describe the way in which a text enacts itself as a unified whole, Halliday and Hasan (1976) advanced the notion of 'texture', a property that holds "the clauses of a text together to give them unity [and] distinguishes text from non-text" (p 2). Texture is achieved through the resources of cohesion and coherence, among others. A subsequent revision of the concept of text by Halliday and Matthiessen (2013) defined it as "any instance of language, in any medium, that makes sense to someone who knows the language [...] we can characterize text as language functioning in context" (p 3). The concept of discourse, within this approach, describes "the different types of texture that contribute to making text: the resources the language has for creating text" (Eggins 1994, p 85).

Within the context of text linguistics, text has been defined as a "communicative occurrence which meets seven standards of textuality" (De Beaugrande & Dressler 1981, p 3). Textuality is thus a defining characteristic of text, which distinguishes text from non-text. According to these scholars, textuality is achieved through compliance with standards that include cohesion and coherence. If any of the standards of textuality is not satisfied, the text becomes 'non-communicative' and thus is considered 'non-text'.

2.2.2 Discourse competence

Discourse competence concerns the creation and understanding of text and was defined by Canale as follows: "the mastery of how to combine and interpret meanings and forms to achieve unified text in different modes by using (a) cohesion devices to relate forms [...] and (b) coherence rules to organise meanings [...]" (p 335, 1993). Building on a pedagogically-oriented model, Celc -Murcia (2008) argued that discourse competence lies at the core of communicative

competence because this competence is where the linguistic, actional, and sociocultural competences converge to articulate and shape the production of discourse. In this model, discourse competence is conceptualised as "the selection, sequencing and arrangement of words, structures, and utterances to achieve a unified spoken message" (Celc -Murcia 2008, p 46), with four main sub-areas contributing to discourse processing. These sub-areas are cohesion, deixis, coherence, and generic structure.

The centrality of discourse competence in communicative competence can be further justified on the basis of language use. That is, we use language to interpret or negotiate intended meanings as well as to convey meaning. To achieve this we create discourse. A strong case can be made arguing that (successful) language use requires the articulation of the different types of knowledge embedded in language ability in the production of discourse. As discussed above, the concept of connectedness or 'textuality' is a core notion in the study of text, as it represents a decisive criterion for a group of sentences (or utterances) to be considered discourse as opposed to a disjointed passage. Thus textuality is the standard that a sequence of sentences (or utterances) needs to meet in order to qualify as text (De Beaugrande & Dressler 1981; Eggins 1994; Halliday & Hasan 1976).

Of the seven standards that texts need to display in order to qualify as text, the features of cohesion and coherence are the core standards that provide texts with 'connectivity' (ie unity) (De Beaugrande & Dressler 1981). Cohesion refers to semantic relations between sentences within a text, which offer a text a degree of unity (Cameron, Lee, Webster, Munro, Hunt & Linton 1995). Coherence also concerns textual unity and includes elements that make a text meaningful (De Beaugrande & Dressler 1981). Because these features have been identified as the standards that endow text with connectedness, these textual properties have been seen as the main contributors to the construction of unified discourse. It comes as a natural consequence that the textual properties of cohesion and coherence have received most attention in discourse studies (Halliday & Matthiessen 2013). Accordingly, one way to analyse the degree of discourse competence observed in oral or written performance is to examine the features of cohesion and coherence displayed in performance (Kang 2005). Therefore, in the present study, these textual attributes have been identified as pivotal in the operationalisation of discourse competence.

2.3 Investigation of discourse competence in learner performance

The importance of discourse competence has been well acknowledged in studies investigating the quality of learner performance in both writing and speaking. In these studies, coherence and cohesion have been two of the most frequently observed aspects of discourse competence. Earlier studies examined the use of discourse markers used by international teaching assistants (ITA) in universities in the US to identify the source of difficulty in comprehending the speech of non-native speakers (e.g. Tyler 1992; Williams 1992). Some studies compared the quality of non-native speaker oral production with that of native speakers and other studies investigated factors influencing quality of speech, such as proficiency, tasks, and the amount of preparation time. The findings showed that infrequent or inappropriate use of discourse markers caused difficulties in comprehension (eg Fung & Carter 2007; Tyler 1992) and that the frequency and type of discourse markers differed according to the learner's proficiency, task types (Geva 1992), and planning time (Williams 1992).

Despite the acknowledgement of the importance of discourse competence, for some reason learners were not aware of discourse devices and teachers paid little attention to these devices in contrast to the attention paid to other aspects of L2 proficiency, especially grammar and vocabulary (Hellermann & Vergun 2007).

An increasing amount of language testing research has analysed various features of the language produced by test-takers in oral assessment in both monologue and interaction. Van Lier (1989) stressed the importance of speech analysis, especially the importance of looking at oral tests using data from test performances (ie, what test-takers actually said), in order to address issues of validity. Douglas and Selinker (1992, 1993) argued that raters, despite working from the same scoring rubrics, may well arrive at similar ratings for very different reasons. In other words, speakers may produce qualitatively quite different performances and yet receive similar ratings. Earlier studies compared the scores assigned by raters with what test-takers produced through in-depth analysis (eg Douglas 1994; Fulcher 1996). More recent studies (eg Brown, Iwashita & McNamara 2005) investigated features of performance identified in rating scales focusing on individual performance, while other studies concerned interactional features observed in oral interview or peer interaction assessment (eg Brooks 2009).

In a context of speaking scale development for the TOEFL iBT Brown et al. (2005) examined the relationship between detailed features of the spoken language produced by test-takers and holistic scores awarded by raters to these performances. Their analysis included data collected from spoken test performances representing five different tasks and five different proficiency levels (200 performances in all), using a range of measures of grammatical accuracy and complexity, vocabulary, pronunciation, and fluency.

It revealed that features from each category helped distinguish overall levels of performance, with particular features of vocabulary and fluency having the strongest impact. Though Brown et al. (2005) examined a few aspects of discourse competence, including use of conjunctions and schematic structure for a sub-set of data, other aspects of discourse competence were not fully investigated.

3 RESEARCH QUESTION

Despite the widely held view concerning the centrality of discourse competence in the communicative competence model, and empirical findings arising from discourse analysis of learner performance on various tasks in pedagogic contexts, little is known about how features of discourse competence are reflected in speaking test performance. Hence the current study addressed the following research question:

What are the distinctive features of performance that characterise test-taker discourse in IELTS Speaking Task 3 at each of the Levels 5, 6 and 7?

4 METHODOLOGY

4.1 Data

The present study analysed transcribed speech samples provided by IELTS. The data comprised a total of 58 test-taker performances corresponding to the three proficiency levels (ie, Levels 5, 6 and 7).

The 58 participants had a diverse range of L1 backgrounds (22 languages). The largest L1 group was Chinese (N=8) followed by Tagalog (N=6), Urdu (N=6), Vietnamese (N=5), and Arabic (N=5). No single L1 group dominated one level. The 22 different L1 groups are roughly spread across the levels.

Approximately 32% of the performances in the data correspond to female test-takers (Female=22, Male=36). Detailed information about the test-takers' L1 and levels is summarised in Appendix 1.

4.2 Method

The methodology for this study built upon the previous studies investigating the characterisation of test-taker performance through the analysis of writing performance, such as IELTS Academic Writing Task performance at various IELTS proficiency levels (Banerjee, Franceschina & Smith 2004; Mayor, Hewing, Swann & Coffin 2007). Unlike previous research, this study set out to identify distinctive features of oral discourse construction as observed in test-takers' performances in IELTS Speaking Part 2. This section of the speaking test provided test-takers with an opportunity to talk about a particular topic; test-takers had two minutes preparation time and were allowed to make notes that could be used during the interview. The examiner may ask them one or two questions on the same topic to finish this part of the test.

As explained earlier, in the data analysis, discourse competence was operationalised in terms of the textual features of cohesion and coherence. Not only have these textual aspects been identified as core contributors to the construction of discourse (Canale 1983; Halliday & Matthiessen 2013), but they have also been integrated as “important aspects of the IELTS rating scale” (Banerjee et al. 2004, p 11). Based on the identification of the textual properties of cohesion and coherence as core contributors to discourse competence in the test-taker performance discussed above, the current study explored discourse competence by examining the levels of cohesion and coherence displayed in the performances.

In order to do so, we undertook both qualitative and quantitative analyses to identify the textual resources used by test-takers to achieve cohesion and coherence in their discourse. For analysis, a section of the transcribed data of the IELTS Speaking Part 2 performance, where test-takers provided a response (in the form of monologue), was first compiled in a database. It was then subjected to both qualitative and quantitative analysis of cohesion and coherence as explained below.

4.3 Analysis

4.3.1 Cohesion

In order to examine cohesion in the test-taker discourse, it was considered best to employ the method used by Banerjee et al. (2004), which explored writing performance in the IELTS test through a number of features including the use of cohesive devices, levels of lexical richness, syntactic complexity and grammatical accuracy. Although Banerjee et al. (2004) claimed that the use of these measures in the analysis attempted to produce “a [reliable] learner language profile” (p 8), the categories used in their investigation did not necessarily characterise test-taker performance in discursive terms. Furthermore, although the study acknowledged the importance of the features of cohesion and coherence in IELTS rating scales, cohesion devices examined in the study were only demonstratives as anaphoric reference and use of ellipsis and substitutions. For that reason, we chose the four aspects of cohesion proposed by Halliday and Matthiessen (2013) outlined below. The four ways to identify cohesion achieved in English are listed below.

1. **Conjunction:** This resource “creates cohesion by linking whole clauses or combinations of clauses” (Halliday & Matthiessen 2013, p 604). It represents logico-semantic relationships between components of a text at the clause level.
2. **Reference:** This generates cohesion by creating links between an element of the text and something else (entities, facts, or phenomena) in reference to which it is interpreted.
3. **Lexical cohesion:** This resource operates at the lexical level and “it is achieved through the choice/selection of lexical items [...] these cohesive relations [may] hold between single lexical units [or] wordings having more than one lexical item in them” (Halliday & Matthiessen 2013, p 642).

4. **Ellipsis and substitution:** These cohesive resources function at the level of the clause or a smaller item. Ellipsis “allows for the language user to leave parts of a structure when they can be presumed from what has gone before” (Halliday & Matthiessen 2013, p 606) while substitution enables the replacement of one item by another.

In order to investigate the degree of cohesion observed in test-taker discourse, the data analysis focused on the examination of three of the four cohesive resources detailed above. The analysis of ellipsis and substitution in IELTS test-taker discourse was addressed in the previous study (Banerjee et al. 2004). We also attempted to identify the use of ellipsis and substitution in test-taker performance in our initial analysis. However, the analysis was eventually abandoned as these features were not frequently observed in the performances.

4.3.1.1 Conjunction

Each transcribed file was scanned to identify the use of conjunctions to signal textual relations. Once identified, conjunctions were further classified into one of the four categories below depending on the logico-semantic relationship being enacted in the text. This analysis was based on Martin’s (1992) classification of conjunctive relations where he identifies the four main types of conjunctions as:

- **additive**
(eg and, or, moreover, in addition, alternatively)
- **comparative**
(eg whereas, but, on the other hand, likewise, equally)
- **temporal**
(eg while, when, after, then, meanwhile, finally)
- **consequential**
(eg so that, because, thus, since, if, therefore).

After detailed scanning of the transcribed performances of each file in each band level, the total number of each type of conjunction, as well as the total number of conjunctions used was calculated. An example of the analysis and coding is presented below.

113	the <i>job</i> i'd like to do is sports management (.6) then what i	consequential
114	need to bring the job is to manage the different kinds of	
115	<i>sports</i> (.5) and share advice with different sports people and	additive
116	note their (.7) feelings or attitude towards the sport they	additive
117	do (.8) then (.7) why- wa- they skills i need to the job	consequential
118	(.2) can be:: (.8) i have to be active in the sports i	
119	manage (.5) and i have to be able to:: (.8) know how the	additive
120	sports:: is played and	

(Level 5 ID202)

In the excerpt above, conjunction use was found in lines 113, 115, 117 and 119. Once identified, each conjunctive element was highlighted in bold and analysed with respect to the context of use in the utterance. Finally, the conjunctive element was further categorised. The sub-category identified for the conjunctive element was recorded in the adjacent column. The mean and median of the total number of conjunctions and conjunction types observed in the test-taker performance in each band level was recorded for statistical analysis. It should be noted that frequent use of additive and consequential conjunctions might be attributed to the nature of oral language, and also additive conjunctions (such as 'and') and consequential conjunctions (such as 'because') might have been used as a filler.

4.3.1.2 Reference

The analysis of the use of referential expressions in the current study was limited to the identification of the use of anaphoric reference. This decision was based on the fact that, when analysing the use of reference in test-taker performance, the previous research (Banerjee et al. 2004) identified the majority of the occurrences in its data as corresponding to instances of anaphoric reference. The following example in Table 1 provides a sample of data coding.

Line	Text	Reference	Referent	Accuracy RI – right WR – wrong
67	okay (.) alright (.) thank you (1.0) the piece of equipment that I find			
68	very useful (.7) is in the home (1.2) and is the rice cooker (.8) the		1 rice cooker	
69	rice cooker is very useful to cook (.) because it is very easy to use (.)	it	1	RI
70	erm (1.5) the instruction when I bought the equipment is very easy to			
71	follow (1.5) and I got it in 1997 (1.3) I think that er:: (3.0) this	it	1	RI
72	equipment is very useful to for everyone (.7) and especially for me (.)			
73	besides I used it for cooking rice (.4) and I also use to warm the rice	it	1	RI
74	(.9) the night before I cook the rice (1.6) but I can use it when I want	it	1	RI
75	to make some cookies (1.1) I can use it for that (.7) and also I can you	it	1	RI
76	it to boil some water (.5) this piece of equipment (.) I find it very	It	1	RI
	Assessment	No. of referential expressions – 6, Accuracy 6/6 100%		

Table 1: Example of analysis of referential expression (Level 6, ID606)

The instances of anaphoric reference were found in lines 69 and 76 through the use of the pronoun 'it' and recorded in the reference column. The analysis of the use of this referential expression allowed us to trace back its antecedent to line 68 where the noun 'rice cooker' was first used. Once the referent of 'it' in line 69 was identified as 'rice cooker', the referent was recorded in the referent column as '1 rice cooker'. The number 1 refers to 'rice cooker', being the first referent introduced. Once the reference item and its proposed antecedent (referent) were identified, the agreement between reference and antecedent was assessed as right (RI) or wrong (WR) as shown in column five.

In the example above, the use of the referential expression ‘it’ (in line 69) can be traced back to ‘rice cooker’ with which it stands in agreement and therefore this was noted as ‘RI’ in column five. In line 75, the test-taker introduced the referential expression ‘it’ which, upon closer examination of the surrounding linguistic context, appears to be related to referent 1 (ie the ‘rice cooker’) in line 68. Accordingly, this was noted as ‘it’ in the reference column and 1 for ‘rice cooker’ was recorded in the referent column. The total number of referential expressions and also a percentage showing the proportion of accurate use of the referential expression were calculated and shown in the bottom of the table (note: the example is not the whole speech of the test-taker). In this example, six anaphoric references in total were used and the referent of all six references was correctly identified and, therefore, the accuracy was recorded as 100%.

The example in Table 2 shows five referential expressions (‘they’) are observed, but unlike the example above, four of the five references could not be traced back to a referent and, therefore, the accuracy of the referential expression was assessed as 20%.

Line	Role	Text	Reference	Referent	Accuracy RI- right WR- wrong
200		schoo::ls: (0.4) i believe has a really (.) really great impact		1 school	
201		on students:: (0.6) they for onc::e (0.7) are the big (0.3)	they		WR – not clear
202		factor (0.7) for? (0.2) for the:: growth and development (0.2)			
203		of every students:?=			
204	E	=m hm:?			
205		0.3			
206		of every children that comes along- that comes:: they are (used	they	No referent	WR
207		to) ((inaudible))			
208		0.2			
209	E	so do they effectively prepare people for jobs?	they	1 school	RI
210		0.2			
211		er yes i believe [so:::] (0.4) er::m:: (0.3) it is a stepping=			
212		[m hm::]			
213		=sto::ne? (0.7) erm (0.7) learning in- a- in an (.)			
214		institutional school (0.2) would serve an es- (.) (plispect)			
215		(0.3) erm stepping stone for an individual. (0.2) to be			
216		effective (0.3) to be fruitful (0.3) erm::: (1.1) in their	their	No referent	WR
217		er:m (0.2) in the jobs that they er:: want there to be.	they	No referent	WR
218		(.)			
E – examiner	Assessment	No of referential expressions – 5 Accuracy 1/5 20%			

Table 2: Example of analysis of referential expression (Level 6, ID263)

4.3.1.3 Lexical cohesion

Strategies to achieve cohesion involve the use of linguistic resources (grammatical items), however, cohesion can also “operate within the lexical zone of lexico-grammar” (Halliday & Matthiessen 2013, p 642). This means that cohesion can be created in discourse through the use of lexical items to establish “relationships among lexical items in a text, particularly, among content words” (Paltridge 2000, p 134). The index of lexical cohesion in the data was analysed in terms of the four lexical relations: repetition, synonymy (a word with the same or similar meaning, eg pupil vs. student), hyponymy (a specific term to refer to a member of a class, eg animal for dog, kangaroo, koala), and meronymy (a word to refer to a member of something, eg apple is a meronym of apple tree). The following example illustrates the analysis procedure and coding.

123	(.3) i'm not sure (.) i'm not er:: (.5) i don't know ab-			
124	(.3) about foreign countries (1) like canada (a) america (b) or (.4)	foreign countries (1)	hyp 1a	hyp 1b
125	er::: eu- (.6) europa (.4) [but] in iran (c) i am:: (.6) a c=			
126	[mhm]			
127	accustomed to iran's(c) house (.3) [er] (.2) in (.5) in the=		hyp 1c x 2	
128	[mhm]			
129	=past the houses were very big [for] example er a- i mean=			

(Level 7 ID206)

(Note: hyp – hyponymy)

It can be observed that in line 124 the test-taker referred to the lexical category of (foreign) ‘countries’ which became the starting point in the lexical chain and was noted with a number 1 as ‘foreign countries (1)’ in the adjacent column. The test-taker subsequently introduced three additional lexical items (‘canada’ and ‘america’ in line 124, and ‘iran’ in lines 125 and 127) relating to the lexical category of ‘countries’ through the semantic relationship of hyponymy. These subsequent items are highlighted in bold and underlined, and then recorded as ‘hyp 1a’, ‘hyp 1b’, and ‘hyp 1c’ because they mark the first (a), second (b), and third (c) hyponyms of ‘countries’ used respectively in the discourse. This information was recorded in the column adjacent to the text. The mean and median of the total number of lexical cohesion devices and their types, used in the performance in each band level were recorded for statistical analysis.

4.3.2 Coherence

While cohesion refers to the internal properties of a text, coherence refers to its “contextual properties; that is the way in which it relates to and makes sense in the situation it occurs” (Paltridge 2000, p 139). For a text to make sense, successful interaction between the knowledge a text presents and knowledge of the world listeners or readers possess is required (De Beaugrande & Dressler 1981). Coherence also stems from the discourse relations established among propositions in the text, the connections among its concepts and ideas, and the logical organisation and development of the text’s thematic content. In order to make discourse cohere, language users adhere to conventional discourse organisation principles, which regulate discourse relations in a text. In the current study, the examination of the degree of coherence observed in test-taker performance was operationalised in terms of text generic structure and theme and rheme development.

4.3.2.1 Text generic structure

Text generic structure refers to the text’s formal schemata that allow users to identify an oral discourse segment as a conversation, narrative, interview, service encounter, report, lecture, sermon, etc. Generic structure mappings have been developed adapting the model developed by (Egins 2012).

In order to examine whether test-takers were able to conform to a particular text structure, we analysed the texts produced by the 58 test-taker performances in two steps. In the first step we examined the prompt presented to the test-taker in the speaking task. The example below shows an examiner’s discourse elicitation.

101	E: [he]re’s some paper? (.) and a pencil for making notes
102	(0.5) .hhh and here’s your topic (0.6) i’d like you to
103	describe a JO::B (.) you think would be interesting.

(Level 5 ID202)

In this extract, the examiner presented the test-taker with the prompt in Speaking Part 2. In line 103, the examiner requested the test-taker to provide a description of a job they would find interesting. On the basis of the analysis of the examiner’s prompt, the text type elicited was classified as descriptive text.

Secondly, we conducted analysis to establish the degree of test-takers' compliance with the conventional generic structure associated with particular text types (Paltridge 2000). The following is an example taken from the data file in the study of an IELTS Speaking Part 2 question asked of a test-taker by the IELTS examiner.

[...] good, paper and pen there for making notes (.2) and here's your topic (.3) so i'd like you to describe your favourite newspaper or magazine.

(Level 5 ID216)

Our analysis in the first step followed Paltridge (2000), who presented a characterisation of text types, as well as structures (in terms of a text's goal and the moves expected in a particular genre) that are conventionally associated with a given text type. For example, descriptive text was characterised as the type of discourse that aims "to describe a particular person, place or thing" (p 111). Its schematic structure was suggested to include two items or moves, namely, identification (id) and description (de). It has the conventional structure as illustrated.

Description

Purpose: To describe a particular person, place or thing

Schematic structure:

1. Identification
2. Description

(Paltridge 2000, p 111)

Analysis of the prompt showed that the test-taker was required to provide a descriptive text. In the second step of the analysis we established the degree of compliance observed in the test-taker performance when compared to

the suggested generic structure of that text in particular, in this case descriptive text. The degree of discourse compliance to conventional generic structure patterns was estimated using a scale developed for the current study based on the description of prototypical text schematic structure presented in Paltridge (2000). The scale included three levels of compliance (1, 0.5, and 0) indicating the degree of test-taker discourse adherence to the suggested structure for a given type of text: 1 for conforming to generic schematic structure to a full extent, 0.5 for some extent, and 0 for no obvious structure conforming to generic schematic structure. It should be noted that the instructions given in the task card required test-takers to provide an item and then not only to describe it, but also to explain the reason for choosing the item (eg why the item was useful). Therefore, test-takers' descriptions included some explanation. The examples below illustrate the generic structure analysis.

In the following example, the test-taker was asked to describe their favourite newspaper. The test-taker identified it as 'the times of india' in line 208, noted as 'id a' in the adjoining column. (Note that the letters 'a', 'b', and 'c' allow for the tracing back of the identified element in the discourse.) As observed in the example, the test-taker provided a description of the newspaper by explaining the type of information included in the paper (eg line 218). Next, the performance introduced the different sections in the paper together with providing an explanation of the type of information contained in each of these sections. The test-taker fulfilled the task requirement by producing a descriptive text in compliance with the schematic features associated with this type of discourse, and therefore was rated as 1.0.

208	yeah like er my (.3) favourite newspaper is the times of india (a)	id a
212	and like er: (.4) only cos i have seen (.) it's the (.2)	
213	newspaper which i have been using since the time i was born	
217	and it's (.) really (.2) unique (.5) in aspect like <u>it covers</u>	
218	<u>all parts er (.2) life (a)</u> (.) like (.6) starting <u>the front page</u>	de a
219	<u>is the headlines (b)</u> (.6) where is all the political news and	de b
220	whatever happening (.) around the world	
224	and then comes into the bangalore city (.6) and you get all	
225	<u>the crime news and whatever happened in the previous day (c)</u> (.3)	de c
226	in bangalore (.7) and then you have the <u>business (.7) where</u>	
227	<u>you get all the business related materials (d)</u> (.6) and <u>the sports</u>	de d
228	<u>section (e)</u> (.2) and you have an exclusive section for the	de e

(Level 7 ID230)

(Notes: the items in bold only are identification, while italic, bold, and underlined parts of the utterance refer to description; each item is labelled with 'a', 'b', etc; id – identification, de – description)

In the next example, the test-taker was asked to describe a piece of equipment. The test-taker identified a watch (line 90) as the item for description. As seen in the example, the test-taker attempted to provide a description of the watch by listing possible uses of a watch, (lines 90, 91, 92, and 93, labelled 'de a', 'de b', and 'de c' respectively). Although the discourse incorporates a descriptive element (ie, description of the usefulness of a watch), the performance fails to provide further elements to describe the watch (eg materials it is made of, brand, colour). For that reason, this performance was rated 0.5 as it does not comply fully with a schematic structure of descriptive text.

- 90 The **watch (a)** (.6) *we: use it for uh: (.6) every, kind of (.8) in any in ev-* id a
 91 *every time (.2) in the time ((inaudible 2.7)) do anything (de a)* (.5) eh we de a
 92 (.5) *we finishing eh fin- finishing and work (de b)* uh: (.3) when uh we have de b
 93 ha- uh (.5) *ha- have an appointment (de c)* (1.1) when uh (.9) we uh with de c
 94 with with other (1.1) eh men (1.5) we uh: (1.0) ((inaudible)) (.9)
- (Level 5 ID510))

In the following example, first the test-taker chose their favourite building for the description in lines 146–147 ('Bayoke building') noted as 'id a', and then moved on to describe this building's properties and characteristics (text in italics) in lines 148, 149, 152, 153 and 155 (noted as 'de a', 'de b', 'de c', 'de d', 'de e', and 'de f' respectively). Because this performance conformed to descriptive discourse schematic structure to a full extent, it was rated as 1.0. No performance was rated 0 on this scale. Data from all 58 test-taker performances was assessed using the generic text structure compliance index first, then the mean and standard deviation was calculated and compared across the levels.

- 145 [erm t]he building that i'm
 146 going to (.) to say is:: (.6) is the **bayoke (.8) bayoke**
 147 **building (a)** (.7) it (.3) it is::: *the (1.3) tallest building in* id a
 148 *(.2) thailand (a)* (.8) erm (2.2) erm it is::: *situated (.7)* de a
 149 *on: (.) pr- (.7) pat- (.) pat- (.6) pratunam (b)* (.6) road (.6) de b
 150 °(i should think)° (.3) .hhhh yes (.) erm::: (1.4) it is::: u-
 151 (1.2) it is the 1.1 building fo:r (.3) erm (.5) *used for like*
 152 *(.3) hotel (c):: (.6) um like (.5) apartment (d)* (1.3) a::nd (.4) de c / de d
 153 *some restaurants (e):: (.6) in there (.8) yeah erm::: (.8) i think* de e
 154 it (.) it is::: (1) you can say that it is the *land- (.2)*
 155 *landmark of:: (.2) bangkok (f)* (1.2) erm (3) although it- (.2) it de f
- (Level 5 ID209)

4.3.2.2 Theme–rheme development

Theme–rheme development refers to the organisational pattern of propositions in discourse (ie, thematisation and staging of propositions) to constitute a unified message (Halliday & Matthiessen 2013). According to Halliday and Matthiessen (2013), the theme of the clause is the element that acts as the departure point in the message, where “the first structure functioning as participant, a circumstance or [...] process” (p 91) is the theme. Accordingly, the part of the clause where the theme is developed is referred to as a rheme. Thus, the structure of the message comprises two elements: a theme and a rheme. The way in which theme and rheme distribute in a text is referred to as thematic progression, defined as “the way in which the theme of a clause may pick up, or repeat a meaning from a preceding theme or rheme” (Paltridge 2000, p 140). In the present study, identification of theme–rheme development was limited to main clauses only (not subordinate nor embedded clauses). Spoken texts were separated into clauses where the clause boundary was set through the identification of tone-units, which have been suggested by Roche (2000) to mark grammatical units such as clauses. Thus, clause boundaries are established by speech pauses longer than 0.6 seconds. Examples of this analysis are shown below.

- 113 **the job (t1) i'd like to do** *is sports management (r1)* (.6) then **what i**
 114 **need to bring the job (t1)** *is to manage the different kinds of*
 115 *sports (.5) and share advice with different sports people and*
 116 *note their (.7) feelings or attitude towards the sport they (r2)*
- (Level 5 ID202)

(Notes: t – theme, r – rheme)

In the example above, the word 'job' was identified as a point of departure in the message and, consequently, it was labelled as 'theme'; the counterpart 'rheme' developed as a complement in the rest of the clause. We observe that theme 1 (coded as 't1' – 'the job') was recovered again and repeated in the next clause thus indicating there was something else to be said about theme 1 ('t1') in rheme 2 (coded as 'r2'). This type of thematic progression is referred to as 'theme–reiteration' or 'constant theme' (Paltridge 2000). In this pattern, the first theme was used and repeated in the consecutive theme as shown below (Figure 1).

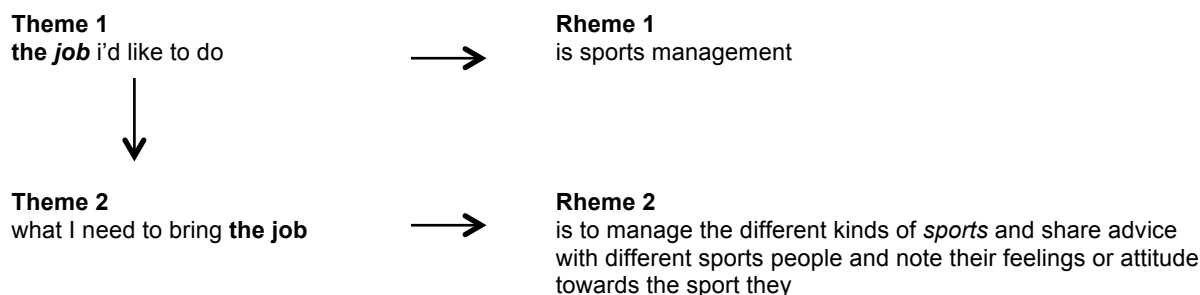


Figure 1: Example of theme–reiteration/constant theme pattern 1

Similarly, in the example below, the element ‘I’ in rheme 1 (in italics, line 22) was taken up as theme 2. At the same time, the (implicit) subject in rheme 2 (italics, lines 23–24) was picked up as theme 3. These patterns are illustrated below.

19 C: so **I (t1)** am: (0.6) talking about eh (0.4) computers ((inaudible)) (0.4) normal
 20 eh central processing unit (0.7) [which]
 21 E: [mm]
 22 C: is a device (0.9) that (.) I've been very much in touch (r1) // (0.8) and (.) **I (t2)** like
 23 C.P.U. because (.) it (0.6) m- add while at sitting (0.4) while sitting at
 24 home (r2) // (0.9) **I (t3)** can explore the world (0.4) and have any information (0.6)
 25 what sort of information I need (1.6) and eh (0.6) because my field (0.5)
 26 is being (0.6) computer (1.3) (r3) // **a:nd this (t4)** is why (.) I (0.8) did my
 27 correspondence through (0.7) my C.P.U. I (0.6) eh (0.4) through my
 28 computer, (0.9)(r4) // and **I f-first time I-I (t5) used it (0.3) I (t5)** was bit [confused] (r5)

(Level 6 ID613)

(Notes: // signals clause boundary; C – candidate; E – examiner; t – theme; r – rheme)

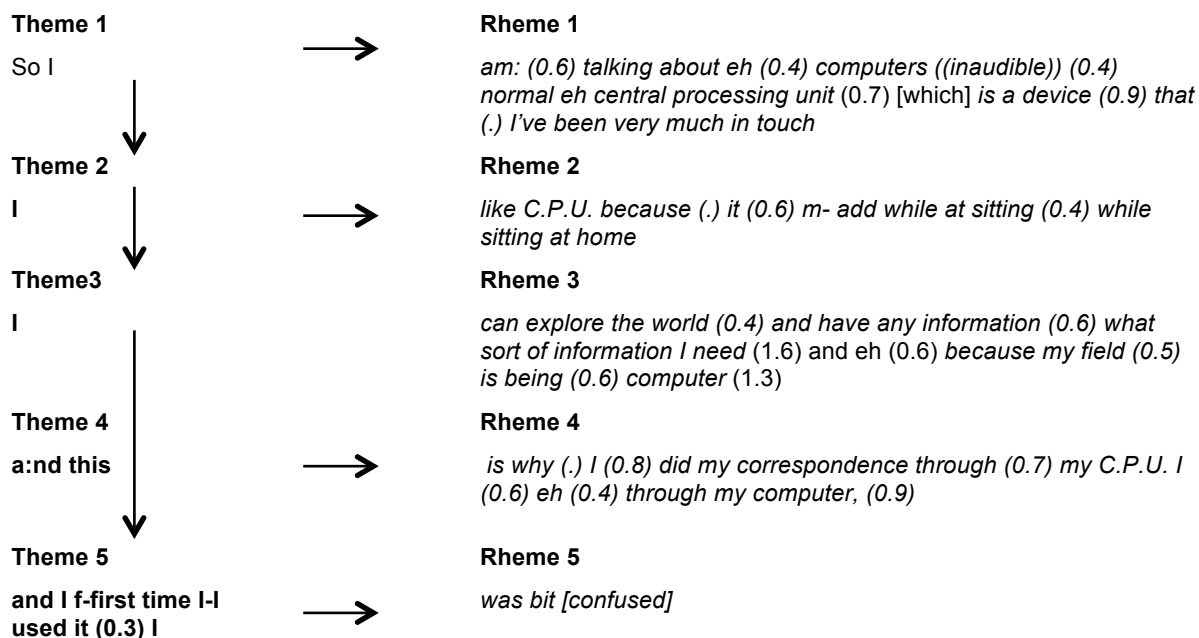


Figure 2: Example of theme–reiteration/constant theme pattern 2

The structure of the example taken from Level 5 below is a combination of constant theme (or theme–reiteration) and zigzag/linear theme patterns. A constant theme pattern occurs when the first theme is used and repeated in the consecutive themes. In the example above ‘I’ in theme 2 was picked up in theme 4 as seen in the diagram in Figure 3. A zigzag/linear theme means that the information in the rheme of a clause was taken up as the theme of the subsequent clauses. In the example below, the information in rheme 2 was taken up as the theme in the following clause (theme 3). The patterns are graphically represented in Figure 3.

78 erm (0.6) **a piece of equipment that I find very useful to me (t1)** (0.2) is (0.4)
 79 my paint box um (r1) // (1.8) **I (t2)** use this paint box of course to put all my paint in
 80 it (r2) // (0.2) **it (t3)** is a erm birthday gift from my father (r3) // (0.6) **I (t4)** first use it when I was ten years old //
 (Level 5 ID511)

(Notes: // signals clause boundary; C – candidate; E – examiner; t – theme; r – rheme)

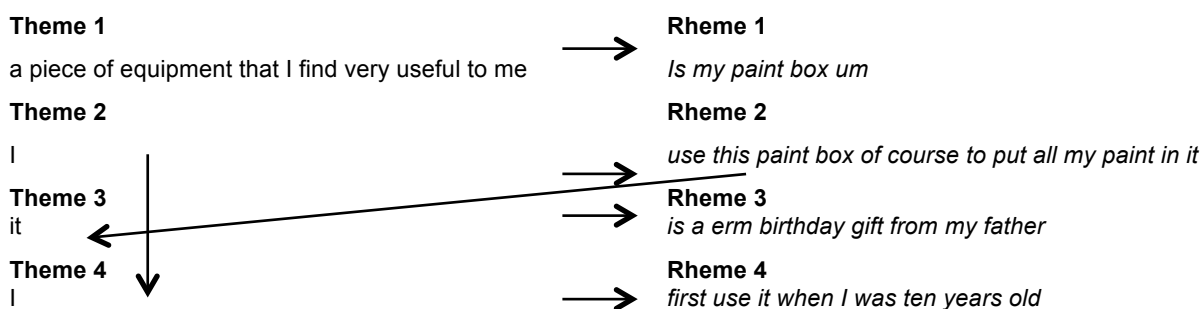


Figure 3: Example of combination of theme–reiteration and zigzag/linear theme pattern 1

In this extract, we observe that theme 2 (‘I’ in line 91) is repeated as theme 3, thus establishing a theme–reiteration/constant theme pattern. Additionally, a piece of information in rheme 3 (‘it’ in italics, line 94) is taken up as theme 4 (‘it’ in bold, line 94) in a zigzag linear theme pattern. The theme–rheme configuration is illustrated in Figure 4.

90 okay (0.7) err: (1.4) in this day and age (.) **a computer I would say (t1)** is
 91 a very useful piece of equipment (.) especially for me (r1)//(.) **I (t2)** log onto the
 92 computer get my data information (.) research materials (.) any patient
 93 which I feel I need to access more information from overseas (.) or from
 94 my professors (.) or from my friends (r2) // (1.2) so **I (t3)** find it very useful (.) (r3) // **it (t4)**
 95 gives me such a broad view of what I can do for my patient (r4) // – and then

(Level 7 ID716)

(Notes: // signals clause boundary; C – candidate; E – examiner; t – theme; r – rheme)

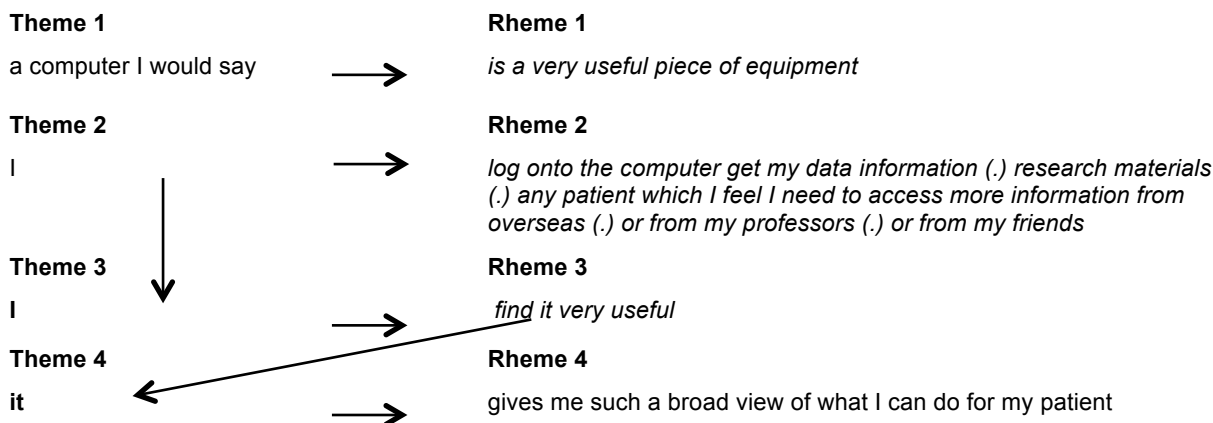


Figure 4: Example of combination of theme-reiteration and zigzag/linear theme pattern 2

The theme–rheme pattern in the example below is also referred as a zigzag/linear theme pattern because a rheme includes information which is taken up as a theme in the subsequent themes. In this example, rheme 1 ‘bayoke’ is picked up in subsequent themes (ie themes 2, 3, and 4) more than once. Recurrence of the theme ‘it’ also shows a theme-reiteration/constant theme pattern as seen in Figure 5.

145 [erm t]he building that i'm
 146 going to (.) to say is:: (0.6) (t1) is the bayoke (0.8) bayoke
 147 1.building (0.7) (r1) it (0.3) it (t2) is::: the (1.3) tallest building in
 148 (0.2) thailand (0.8) (r2) erm (2.2) erm it (t3) is::: situated (0.7)
 149 on: (.) pr- (0.7) pat- (.) pat- (0.6) pratunam (0.6) road (0.6)
 150 °(i should think)° (0.3) .hhhh yes (.) (r2) erm::: (1.4) it is::: u-
 151 (1.2) it (t4) is the 1.1 building fo::r (0.3) erm (0.5) used for like
 152 (0.3) hotel:: (0.6) um like (0.5) apartment (1.3) a::nd (0.4)
 153 some restaurants: (0.6) in there (0.8) yeah erm::: (0.8) i think
 154 it (.) it is::: (1) you can say that it is the land- (0.2)

(Level 5 ID209)

(Notes: // signals clause boundary; C – candidate; E – examiner; t – theme; r – rheme)

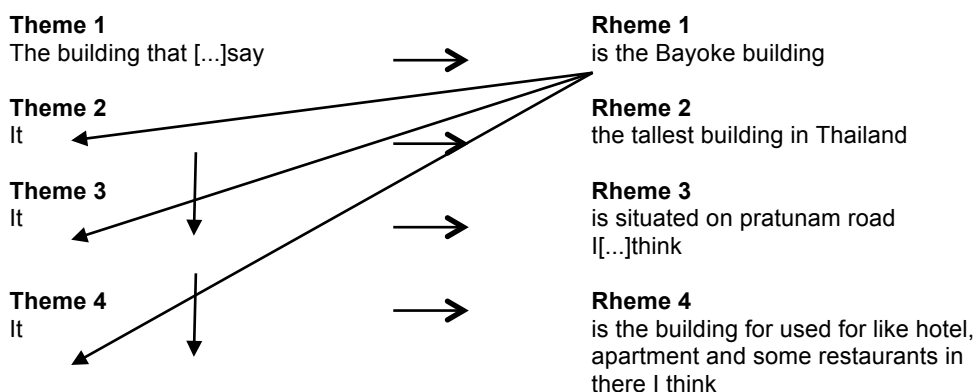


Figure 5: Example of combination of theme–reiteration and zigzag/linear theme pattern 3

Apart from theme-reiteration/constant theme and zigzag/linear theme patterns, there is another pattern called a ‘multiple/split theme’ and refers to a situation where a rheme includes more than one piece of information which may be picked up as a theme in subsequent clauses (Paltridge, 2000). However, this thematic pattern was not observed in the data of the current study. In this study, the theme–rheme development of 58 test-taker performances was first analysed by counting the number of main clauses and identifying theme and rheme in the texts and the theme–rheme pattern, and then the results were compared across the levels.

4.3.3 Lexical richness

In addition to examining the degree of discourse competence in test-taker discourse described above, the current study incorporated the examination of lexical richness in test-taker performance. According to Skehan (2009), lexis can be used as a measure of general performance as it “represents a form of complexity that [...] has to be assessed in second language speech performance if any sort of complete picture is to be achieved” (p 514).

Therefore, we believed it was important to include measures of lexical richness in the analysis of test-taker performance in an attempt to produce a broader characterisation of test-taker performance at different IELTS levels. The index of lexical richness in this study was operationalised in terms of the following categories of analysis:

1. lexical output: analysis to look at the number of tokens produced in test-takers’ performance at different IELTS levels
2. lexical variation: analysis to explore the number of different word types in test-taker performance in relation to the total number of words produced
3. lexical density: analysis aiming at calculating the proportion of lexical words to grammatical words in test-taker performance
4. lexical sophistication: analysis to provide information about the number of unusual words (in relation to specified word lists) in test-taker performance at different IELTS levels of proficiency.

In order to examine these four lexical features, we used the web-based program, *VocabProfile* (Cobb 2013), which measures the proportions of low and high frequency vocabulary used by speakers, both native and non-native. The program is based on the Vocabulary Profile (Laufer & Nation 1995) and performs lexical text analysis using the Academic Word List (Coxhead 2000). In addition to calculating measures for word-token, word-type, and the type-token ratio, *VocabProfile* calculates the percentage of words in each of four categories: the most frequent 1000 words of English (K1), the second most frequent 1000 words of English (K2), words found in the Academic Word List (AWL), and any remaining words (Off List). It also calculates lexical density and word-token per word-type.

In order to undertake the analysis, the transcribed speech was pruned to exclude features of repair and imported into *VocabProfile*. Frequency counts were then developed for each of the nine measures. The word-token measure and word-token per type were used to investigate lexical output. The word-type measure was used to assess the range of vocabulary used; we hypothesised that more proficient speakers would use a wider range of word-types. Lexical sophistication was examined in terms of the percentage of words in the four categories above (K1, K2, AWL and Off List).

In addition to these measures we also investigated the type-token ratio. The type-token ratio is a measure of the semantic density of speech, which may vary according to proficiency level; increases in tokens across levels may not be matched proportionately by increases in types, so speech may be relatively more or less dense at different levels.

5 RESULTS

In this section, we report the results of the data analysis in light of the research question, including the findings of both quantitative and qualitative data analyses. For cohesion, we report the results yielded from both quantitative and qualitative analyses. In reporting the results of the analysis of the three categories of cohesion (ie, conjunction, reference and lexical cohesion), we use frequency data (ie, per 100 words) as the length of speech varied.

To compare the performance across the three levels, use of the bootstrapping method was initially considered as the distribution of the data was not always normal and there was a large individual variation with large standard deviations. Bootstrapping refers to robust statistics which “randomly resamples from an observed data set to produce a simulated but more stable and statistically accurate outcome” (Plonsky, Egbert & Laflair in press, p 1). However, in the end, we used non-parametric statistics (Kruskal-Wallis test) as the frequency of each feature under study was very small and we report the findings of qualitative analysis to validate the results of statistical analysis.

In reporting descriptive statistics, means, standard deviations, and medians are presented. For coherence, because of the nature of the data, our analysis was mainly qualitative apart from the generic text structure compliance index in section 5.2.1, and the comparison of the descriptive statistics of main clauses in section 5.2.2. For lexical richness, we report the results of quantitative analysis using parametric statistics (one-way ANOVA).

5.1 Cohesion analysis

As explained in the methodology section, the examination of the textual property of cohesion in the study considered three major types of cohesive resources used to achieve text cohesion: conjunction, reference and lexical cohesion.

5.1.1 Conjunction

Table 3 summarises the descriptive statistics of conjunction use in terms of the number of conjunctions used in each data file, as well as the type of conjunction favoured by the test-takers according to the taxonomy introduced by Martin (1992).

Across the levels, the most frequently used conjunction was additive, followed by consequential. Test-takers used a relatively small number of comparative and temporal conjunctions across the levels. The frequency of temporal conjunctions was lowest in all three levels. Test-takers’ frequent use of additive and consequential conjunctions was partly due to the nature of oral language. In particular, additive conjunctions such as ‘and’, and consequential conjunctions such as ‘because’ might have been used as fillers. The results of independent non-parametric test analysis (Kruskal-Wallis test) are summarised in Table 4 below. The significant difference in the use of conjunctions among the three levels is observed only for comparative conjunctions, and post-hoc testing showed the difference was found between Levels 5 and 6 ($U = 10.5, p = .018$), and between Levels 5 and 7 ($U = 58, p = .000$).

Type	Level 5			Level 6			Level 7		
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Additive	1.87	1.10	1.66	2.17	1.35	1.88	1.71	.70	1.62
Comparative	.37	.39	.22	.57	.61	.45	.66	.49	.60
Temporal	.27	.22	.30	.17	.21	.12	.23	.25	.13
Consequential	1.66	1.01	1.29	1.17	.60	1.01	1.23	.66	1.10
Total	4.15	1.90	3.75	4.07	2.15	3.71	3.83	1.22	3.77

Table 3: Descriptive statistics of use of conjunctions (frequency, per 100 words)

Type	H	Df	p
Additive	.72	2	ns
Comparative	8.08	2	.02
Temporal	1.95	2	ns
Consequential	2.80	2	ns
Total	.01	2	ns

Table 4: Results of Kruskal–Wallis analysis (conjunction use)

Table 5 and Figure 6 display the distribution of each of the four conjunction types used by test-takers of the three levels. For Level 5, the proportion of additive and consequential conjunctions together is larger than in Levels 6 and 7. As the level goes up the proportion of comparative conjunctions increases. The proportion of temporal conjunctions is very different between Level 5 and 7.

Type	Level 5	Level 6	Level 7
Additive	45.12	53.26	44.71
Comparative	8.81	13.92	17.15
Temporal	6.39	4.25	6.03
Consequential	40.11	28.64	32.14

Table 5: Distribution of conjunction use (%)

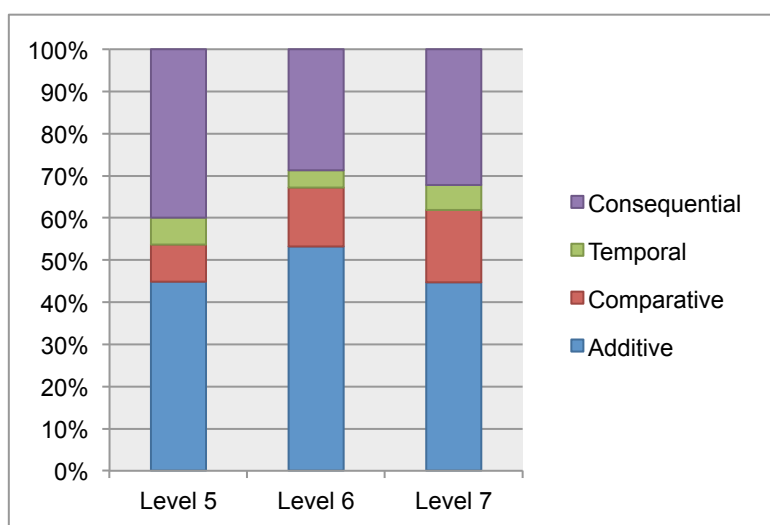


Figure 6: Distribution of conjunction use

The qualitative analysis focused on two aspects: identifying the type of conjunction found in the performances and examining the level of agreement between the use of a particular conjunction and the relation that particular element actually established between the clauses in the text. For example, in the category of additive relations, we examined whether the use of the additive conjunction ‘and’ effectively enacted a relationship of addition in a given test-taker’s performance. The extract below shows the use of additive conjunctions (in bold) in a Level 5 performance.

- | | | |
|----|---|---------------|
| 83 | help immediately (.8) I first used it er (1.0) three years ago I remember | |
| 84 | well (.) when I was in high school <i>my mother er (.8) bought me one for 85</i> | temporal |
| 85 | <i>pr- for my birth day (.6) and I have it since (.2) er with me and er (.8)</i> | additive |
| 86 | now er (.) now (2.0) because I am in another town from my parents they | consequential |
| 87 | (.) find me any time (.2) and we (.) talk (.6) er I think it is useful (.) | additive |
- (Level 5 ID505)

In this example, an additive conjunction ‘and’ was used twice in lines 85 and 87 (in bold). Because the conjunction ‘and’ enacts a relation of addition, it was expected that the content of the second clause would add (new) information to the first one in both cases. In the first case, this conjunction was used to establish a relation of addition between the clauses ‘my mother bought me one for my birthday’ (lines 84–85, in italics) and ‘I have it since with me’ (line 85 in italics). In the case of the second ‘and’ item (line 87, in bold), this conjunction was used to link the clause ‘I am in another town from my parents they find me any time’ (in lines 86–87) along with the clause ‘we talk’ (in line 87).

The analysis revealed that the second clause did indeed add new information in both instances. In the second case, we discovered that the test-taker still kept the gift presented to him by his mother (something we were informed about in clause 1 in lines 84–85), and that the test-taker could talk with his parents even if he was in a different town (something we learned in clause 1 in lines 86–87 respectively).

It is observed that the use of additive conjunctions indeed enacts relations of addition among clauses in performance across band levels. Additionally, the analysis showed that there exist no observable differences in the types of additive conjunctions preferred in discourse across the levels. In other words, test-takers at the three levels of performance did not appear to favour a particular additive conjunction. The use of additive conjunctions ‘and’ and ‘or’ were used indistinctively across the band levels to enact additive relations in discourse; no other examples of additive conjunctions were observed in the performances.

Test-takers of all levels used a variety of conjunctions, but more frequent and more varied comparative conjunctions are observed as the level goes up. Higher proficiency test-takers were also able to use comparative conjunctions for comparison in both positive and negative senses (Paltridge 2000). In the example taken from Level 5 performance data below, the test-taker compared different types of television shows and expressed their preference for news over other shows by establishing a comparison using the conjunction ‘but’ in line 134 (in bold).

- | | | |
|-----|--|-------------|
| 132 | (0.4) the kind of mo- er of (1) the <i>l.show</i> that i like is (0.3) | |
| 133 | ((inaudible)) or talk show quiz show (0.4) er (0.3) and news | additive |
| 134 | (0.2) because it’s not just only for relax (0.5) but is give me | comparative |
| 135 | an (0.9) er (0.2) about (1.8) er the thing that happen | |
- (Level 5 ID210)

The enacting of comparative relations using comparative conjunctions was least favoured in Level 5 in comparison with Level 6 and 7 performances. In the following example, which corresponds to a Level 6 performance, it can be observed that the test-taker used varied comparative conjunctions such as ‘like’ and ‘as’ (in bold, line 260) as well as ‘but’ (in bold, line 262).

In the following extract, the test-taker established comparison (in a positive sense), referring to their father’s expectations for the test-taker to undertake the same job as their father. The comparative relation was thus enacted through the use of the conjunctions ‘like’ and ‘as’ (line 260, in bold). However, in line 262, the test-taker establishes their stance regarding this job opportunity by using the conjunction ‘but’ (line 262, in bold) to express their dislike of that job in contrast to their father’s view. Thus, the test-taker enacted a comparative relation in a negative sense.

- | | | |
|-----|---|---------------|
| 251 | m:: (0.5) the (0.2) the the: international situation (0.3) | |
| 252 | an:::d the:: (0.3) the weapons (0.2) and the cars (0.3) and so | additive |
| 253 | so many (0.4) so i like the (0.5) the- (0.6) the time magazine | consequential |
| 254 | (0.9) and erm (0.7) m: (0.8) er the fame- the famous part (0.2) | additive |
| 255 | er (0.2) i like is car (0.2) er (0.7) cos m:: (0.6) (clears | |
| 256 | throat) (0.7) my father was er er was (producter) (0.4) er was | |
| 257 | m:: (0.6) was a car fact- (.) er was a car factory (0.3) worker 258 | |
| 258 | (0.6) so: (0.5) m: (1.2) before (0.3) before (0.4) before i | |

259	came here my father hoped me (0.3) to: to do er.: (0.4) to	
260	continue to work (car) (0.6) like the same as my father (0.5) 261	comparative
261	to do er car (0.2) to to product the car with him (0.5) but eh	
262	(0.5) er.: (2) but i don't (0.3) i don't like it (0.3) because	consequential
263	i don't like to (0.4) erm (.) to product i like to see it (0.6)	
264	to to drive it	

(Level 6 ID275)

The example below illustrates the use of comparative conjunctions (in bold) in a Level 7 performance. In this extract, four instances of comparison are observed ('as', 'but', 'but' and 'like', in bold, lines 229, 233, 245, and 249, respectively) in a discussion about medical equipment. The conjunctions used in the text enact comparative relations in a positive sense ('as' and 'like', lines 229 and 249) as well as in a negative sense ('but' in lines 233 and 245) (Paltridge 2012).

229	equipment which eh (0.6) which I found very useful, (.) as I told you (.)	comparative
230	as a doctor (0.5) we prefer our equipments (0.7) and eh there are one eh	additive
231	like in modified situation like CT scan and many things like that (0.4)	additive
232	but eh the basic eh (0.4) equipment (.) which is long lasting time tested	comparative
233	(0.6) is our stethoscope (1.1) yeah and eh (0.3) I think doctors are eh	additive
234	identified by the stethoscopes (.) no ma- (0.4) no matter if it's really	consequential
235	((inaudible)) (0.8) and when I wear the (.) stethoscope people feel that he	additive
236	is a doctor (0.5) and (.) this is e- this is equipment which gives us eh	additive
237	(0.3) identification which g- gives us (0.3) eh a help in our eh (0.4)	
238	diagnosing a common patients (1.4) I eh (0.5) when I became doctor I	temporal
239	was in final year (0.5) I first time used the stethoscope and was quite eh	additive
240	(0.3) eh convinced to s- to feel that it it it is going to have (0.3) eh in	
243	particularly when I (0.3) listen the heart sounds (0.6) and eh th- i-it	temporal, additive
244	gives you a specific feelings eh (.) eh you know (0.3) that's the	
245	difference (0.4) but then I change my field and went into just medicine	comparative, additive
246	speciality (0.5) where I found that this (0.3) this (.) particular instrument 247	
247	(0.3) is the key in my field (0.6) because (.) it it really helps in	consequential
248	diagnosing disease- (.) diseases (0.6) which even (.) other invasive	
249	investigations like j- eh like CT scan, (0.3) like eh (.) bone scan like eh	comparative

(Level 7 ID710)

A wider variety of comparative conjunctions in Level 7 may indicate a more developed ability to establish a relation of comparison in text as the proficiency level increases. At the same time, as observed in the example above (ID710) where the test-taker used a variety of conjunctions, the enacting of comparative relations in more proficient discourse appears to develop in a more harmonic balance with other types of logical relations (temporal, additive, and consequential).

5.1.2 Reference

Table 6 presents the results of the analysis in terms of the total number of referential expressions used and the percentage of referential expressions that were used accurately.

	Level 5			Level 6			Level 7		
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
No. of referential expressions	2.60	1.39	2.35	1.88	.91	1.94	2.14	1.06	2.04
Accuracy (%)	77.25	15.94	73.50	78.45	2.52	83.50	92.13	7.92	92.00

Table 6: Descriptive statistics of use of referential expressions (frequency, per 100 words) and percentage of accurate use

The number of referential expressions (eg, it, they, there) observed in test-taker performance is largest in the Level 5 performance, and lowest in the Level 6 performance. The difference in frequency of referential expressions is, however, not significant ($H = 2.72$, $df = 2$, ns). Nevertheless, the percentage of referential expressions used accurately was as expected. That is, the accuracy of referential expressions was highest among Level 7 test-takers and lowest in the Level 5 performance. These differences are statistically significant ($H = 1.24$, $df = 2$, $p < .006$) and post-hoc testing showed the difference was found between Levels 5 and 7 ($U = 80$, $p = .003$), and between Levels 6 and 7 ($U = 109$, $p = .013$).

The qualitative analysis revealed that at higher levels, test-takers were able to enact an accurate relation between the use of a reference item and its antecedent. The example below in Table 7 taken from a Level 7 performance illustrates this point. The example shows that the test-taker made use of a number of reference items in their performance (eg, 'it' in line 120, 'they' in line 123, and 'there' in line 126). We examined the degree of agreement between the use of these expressions and their referents (antecedent). In the case of 'it' in lines 119 and 120 for example, we were able to trace back the use of this particular reference item to its referent 'Maldives' in line 119.

This means that when the test-taker uttered expressions such as 'it's such a beautiful place' and 'it make me breathless', the listener could identify 'Maldives' as the antecedent and thus understand what the test-taker was referring to. This is also the case, for instance, for the successful relationships the text established by test-takers in using the referential expressions 'they' in lines 122 and 123 and its antecedent 'white sand beaches' in line 122, and 'they' in line 125 with its antecedent 'people' in line 125.

Line	Text (excerpt)	Referent	Reference	Accuracy RI – right WR – wrong
119	but eh (.) Maldives is such a nice place (.3) and <i>it's</i> such a beautiful	1 Maldives	it	RI
120	place (.3) that eh (.6) mm I just (.6) it ga- it (.3) <i>it</i> gave me (.3) it	1	it	RI
121	make me (.) breathless (.5) and the scenery (.3) and eh (.4) eh the (.)			
122	beaches (.) white sand beaches (.) <i>they</i> were (.4) I mean (.4)	2 white sand beaches	they	RI
123	outrageously (.6) and eh (.) outrageous (.) <i>they</i> were (.3) <i>they</i> were so	2	they	RI
124	beautiful (.4) and em (.4) what I (.) found over there (.) that eh (.3)	2	there	RI
125	the people (.) over there (.6) <i>they</i> are Islamic in a sense that when I	1 3 the people	1 there 3 they	1 RI, 3 RI
126	went over there it was the holy month of Ramadan (.) and eh (.3) I	1	1 there	
127	found out that eh (.) the shops were closed (.6) eh around six o'clock			

Table 7: Example of Level 7 performance (excerpt) (Level 7 ID709)

In contrast, it was not possible to observe this phenomenon to the same extent in performances at the lower levels. For example, some Level 5 performances failed to establish a relationship between a reference item and its antecedent as illustrated in Table 8. It can be observed that the test-taker introduces a number of reference items (for example 'this' and 'they' in line 140, 'this' in line 142), but these reference items cannot be satisfactorily traced back to an antecedent.

In the case of the use of 'this' in line 140, it is difficult to assess whether the test-taker was referring to the paper or magazine mentioned in line 139, which made the recovery of the reference item 'this' somewhat difficult and ambiguous. A similar phenomenon is observed in Table 9 where the antecedent for the reference element 'they' (line 65) cannot be retrieved from the text.

Line	Text (excerpt)	Referent	Reference	Accuracy RI – right WR – wrong
139	information in er (.4) in this (.4) paper (.4) this magazine			
140	(.4) erm some tier (.2) in this/they are many (.3) er (.3)	No ref	this/they	WR
141	(page) for entertainment .hhherm an er: especially i like uh			
142	(.3) er the (1.4) information about competition in this (.3)	No ref	this	WR
143	because when i finished er university i went to .hhher (.2)			
144	join in er (.2) some competition with (.4) everybody (2.3)			

Table 8: Example of Level 5 performance (excerpt) (Level 5 ID226)

Line	Text (excerpt)	Referent	Reference	Accuracy RI – right WR – wrong
64	(.4) choose France (.) err: Paris is the capital of France (.) err:: and it is	1 Paris	it	RI
65	also (1.5) a fashion city (.4) erm (.2) they also have destinies land there	No Ref	they	WR
66	(.) so I went there (.8) err::m (2.1) I went there by Eurostar and	1 Paris	there	RI
67	I think this is a great city to visit (.) and (.) because erm (2.5) the people	1 Paris	this	RI
68	are very nice (.) and also the place is also very nice (.7) and (.) I think			
69	this is there is (.4) a good place to buy (.8) clothes (.7) so I spent a lot			

Table 9: Example of Level 5 performance (excerpt) (Level 5 ID501)

These examples show that by establishing an accurate relationship between the reference item and its referent, more proficient test-takers are able to provide the listener with the means to recover the identity of an element introduced earlier in the text. In other words, they demonstrate a higher degree of agreement between the reference item and its referent. This, in turn, allows the use of the cohesive resource of reference as a whole to be more successful in the case of more proficient test-takers, which results in producing a comprehensible text. As shown in the results of the quantitative analysis, higher accuracy in the use of reference enabled these performances to demonstrate a higher degree of coherence.

5.1.3 Lexical cohesion

Table 10 summarises the use of lexical cohesion devices. On the whole the frequency of these devices is very low except for repetition and hyponymy. The higher the level, the more frequently hyponymy was used.

The statistically significant difference across the levels is seen only in hyponymy and repetition, as shown in Table 11. Post hoc analysis showed the difference in hyponymy occurred between Levels 5 and 7 ($U = 32, p = .000$). Unlike the use of hyponymy, more repetition was observed as the levels decreased. The difference in the use of repetition was highly significant, and the difference was observed between Levels 6 and 7 ($U = 66, p = .000$) and Levels 5 and 7 ($U = 35, p = .000$).

Table 12 and Figure 7 display the distribution of lexical cohesion devices in the performance of each level. The percentage of each type in each level was calculated out of the total frequency. As reported above, the distribution of hyponymy and repetition in Level 5 data was reversed in the Level 7 data. In the Level 6 performance, the proportion of repetition and hyponymy in the total use of lexical cohesion devices appears to be similar to Level 7 performance. The qualitative analysis included annotation of the use of lexical items as cohesive resources. The examples below show the identification and coding of lexical items in the data.

Type	Level 5			Level 6			Level 7		
	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Synonymy	.08	.27	.000	.06	.10	.00	.10	.11	.08
Antonymy	.07	.11	.000	.08	.12	.00	.10	.12	.04
Hyponymy	.55	.30	.46	1.16	1.03	.84	1.31	.47	1.35
Repetition	1.21	.73	.97	.94	.72	.72	.35	.49	.21
Meronymy	.06	.18	.000	.12	.27	.00	.05	.18	.000
Total	1.94	.90	1.70	2.43	1.15	2.19	1.90	.45	1.85

Table 10: Descriptive statistics of use of lexical cohesions (frequency, per 100 words)

Type	H	df	p
Synonymy	5.91	2	ns
Antonymy	1.08	2	ns
Hyponymy	15.81	2	.000
Repetition	22.07	2	.000
Meronymy	3.13	2	ns
Total	1.75	2	ns

Table 11: Results of Kruskal-Wallis analysis (lexical cohesion)

Type	Level 5	Level 6	Level 7
Synonymy	4.04	2.34	5.39
Antonymy	3.44	3.45	5.26
Hyponymy	28.00	49.30	68.57
Repetition	61.33	39.75	18.17
Meronymy	3.15	5.20	2.49

Table 12: Distribution of lexical cohesion (%)

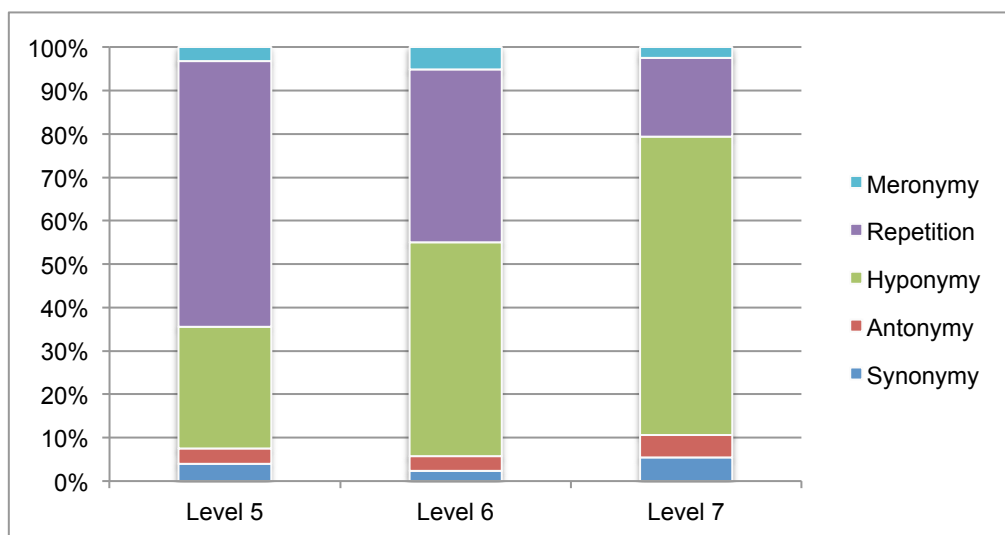


Figure 7: Distribution of lexical cohesion

As can be observed in the extract below, the term ‘computer’ is first introduced in line 97 and then it is subsequently repeated in lines 97, 100, 101, 102 and 108. In this case, the test-taker favoured repetition as a means to establish lexical cohesion. However, as proficiency increases (Levels 6 and 7), the use of repetition as a cohesive device decreases. This is illustrated in the examples taken from Level 6 and 7 performances.

95	(3.5) today I am going to talk about (.2) a 1.piece of equipment (.4)	equip 1		
96	which I find is useful (.4) in my opinion the equipment is (.2)			
97	2.computer (.6) as you know in modern society computer (a) is many (.6) 98	comp 2	hyp 1a	rep 2a
98	bring many benefit to our live (.5) um firstly (.5) um: in study (.5)			
99	secondly in work (.3) and finally (.2) in daily life (.5) firstly I will 100			
100	talk about the advantage of using computer (b) (.) in study (1.0) um I think 101			rep 2b
101	computer (c) is most useful because (.4) er in er study (.3) you can use er 102			rep 2c
102	computer (d) (.) to finish your homework (.) your assessment (.3) even you 103			rep 2d
103	use the computer (e) um (.6) especially using internet (.2) to find (.2) er 104			rep 2e
104	huge information (.3) in the internet (.2) and (.3) it can save your			
105	time and your money (.2) because you can do (.) you can finish your 106			
106	(.5) you can finish your homework (.2) very quickly (.2) and			
107	sometime you can find (.4) some er large information (.) using			
108	computer (f) (.2) you don't need to go to library or (.2) to book (.2) or to 109			rep 2f

(Level 5 ID512)

(Notes: hyp – hyponymy; rep – repetition)

In the example below (Level 6), it can be observed that instances of repetition as a device to establish lexical cohesion is less frequently observed than in the Level 5 performances shown above. In this example, in particular, the lexical resource of repetition was supplemented with the use of hyponymy, which demonstrates the test-taker's ability to use a larger variety of resources to establish lexical cohesion in the text.

57	uh::m washing machine (1) (1.6) uh::m ten years before the washing	w machine1
58	machine (a) is very simple made ((inaudible)) (.6) and now I saw the	rep 1a
59	washing machine (b) got lots of (1.1) got lots of ((inaudible)) different (.4)	rep 1b
60	ways for working (.2) they can wash the clothes and dry the clothes	
61	(.5) this is that impossible when I was a child (.6) an:d uh:: (.4) TV (e)	
62	(.2) digital TV (1.8) uh:m maybe I don't know when (.7) the TV	co-hyp 1c
63	before (.5) is black and white (.4) now it's colour and changing to the	

(Level 6 ID603)

(Notes: rep – repetition; co-hyp – co-hyponymy)

In the following extract (Level 7), only one instance of repetition is observed as shown by the code ‘rep 1a’ while the lexical resource of hyponymy was favoured in the test-taker's text as shown by ‘hyp 1b’, ‘1c 2’, and ‘hyp 1d’ and ‘1e’.

7	er::] my favourite	
8	ewspaper is the nation (.) the daily nation. .hhh (.) and er	newspaper 1
9	(2) it has::: (.6) <u>since</u> it is a newspaper it has very many	rep 1a
10	Articles (b)in it. .hhhh an it has ernews (c) updates of cours::e	hyp 1b 1c2
11	an::d it has: educative areas as well sometimes? .hhhhh (.5)	
12	it also has fashion (d) it has also::: advertisements (e) all sorts of	hyp 1d 1e

(Level 7 ID201)

(Notes: rep – repetition; hyp – hyponymy)

Unlike the case of repetition, the use of synonymy in the performances did not follow such a clear-cut pattern, ie, the use of synonymy was irregular across the band levels. More frequent use of synonymy is observed at the lower level (Level 5) than higher levels (Level 7) as illustrated in the examples below.

99	.hh (1.2) first they must be honest (e) and truthful to him (.8) and err and	honest e	syn e1
100	don't err:: don't lie with him and err say the same with real problem (.5)		
101	problems and characters and help him out to err change his behaviour	behaviour f	
102	(.7) because they are concerned with their behaviour and err I think that 103		rep f1
103	err to help each other and to encourage each other (.5) like my friends I 104		

(Level 5 ID506)

(Notes: syn – synonymy; rep – repetition)

73	I'm going to describe a piece of equipment (a) that I use a lot (.7) and	piece of equipment a		
74	that is my car (b) (.5) first of all this thing is a vehicle (.7) which is	car b	hyp a	syn b
75	motorised by a petrol engine and its black (1.1) it is a four seater (1.6)		hyp	
			a1/mer b1	
76	and er (.7) most of all I use it to transport myself and my friends to			

(Level 7 ID708)

(Notes: hyp – hyponymy; mer – meronymy)

Notwithstanding this irregular distribution, it may be argued that, even though synonymy is indeed defined as a lexical relation based on identity, it is found in performances at higher levels of proficiency as well as in lower level ones. Hyponymy and meronymy are defined as extending lexical relations based on attribution. With no clear boundary between the two terms, they often work jointly in the development of text (Halliday & Matthiessen 2013). The results of the lexical cohesion analysis indicate that test-takers at higher proficiency levels make more use of hyponymy than those at lower levels. A possible explanation for this is that test-takers at higher proficiency levels perceive the elaboration of lexical relations based on attribution to be less demanding and, in consequence, favoured them in their discourse. The advantageous use of this lexical resource in discourse at higher levels of proficiency endowed these performances with a higher degree of competence.

In their analysis of lexical cohesion, Halliday and Matthiessen (2013) classified lexical relations into two types based on a criterion of identity where “one lexical item restates another” and on a criterion of attribute, ie “based on classification (specific to general)” (p 646). Based on the cohesion patterns observed in the test-taker performances above, our findings suggest that at lower levels test-takers might have chosen a device of identifying lexical relations in the text based on identity (‘repetition’ and ‘synonymy’) instead of those based on attribute (‘hyponymy’ and ‘meronymy’). This phenomenon could be explained by the fact that, as hypothesised in the literature, for less proficient test-takers a large amount of attention is devoted to the cognitive processing of language production (Levelt 1989), thus a more direct form of establishing lexical cohesion such as repetition and synonymy may be favoured (Halliday & Matthiessen 2013). This would pose less processing demand and it would thus require less linguistic processing on the part of the speaker.

5.2 Coherence analysis

In this section, the results of the analysis of the two aspects of text coherence (ie, generic structure and theme–rheme development) are presented.

5.2.1 Text generic structure

As explained in the methodology section, in order to examine the text structure observed in test-taker performance, we first identified the text structure that test-takers were expected to conform to and then, based on the degree of adherence to the suggested structure, each test-taker performance was awarded a score ranging from 0 to 1. The results are summarised in Table 13.

Level 5			Level 6			Level 7		
Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
.58	.39	.50	.75	.25	.75	.80	.25	1.0

Table 13: Descriptive statistics of generic structure compliance index scores

The mean scores of the compliance index in the three levels indicate that more proficient performances showed evidence of a higher level of compliance and thus, of familiarity, with the generic structure of the type of text being elicited in the speaking task, but the difference across the levels is not significant ($H = 3.46$, ns). For qualitative analysis, we examined all 58 texts to examine the extent the texts conformed to the generic structure identified in the methodology section.

- | | | |
|----|---|------------------|
| 73 | I'm going to describe a piece of equipment that I use a lot (.7) and | |
| 74 | that is my car (a) (.5) first of all this thing is a vehicle (.7) which is | id a |
| 75 | motorised by a petrol engine (a) and its black (b) (1.1) it is a four seater (c) (1.6) | de a/ de b/ de c |
| 76 | and er (.7) most of all I use it to transport myself and my friends to | |
| 77 | leisure things (d) (.4) I (.7) yes (1.3) I use it for everything (e) (.8) I first | de d/ de e |

(Level 7 ID708)

(Notes: id – identification, de – description)

In the example above, the test-taker was asked to describe a piece of equipment. It can be observed that the test-taker fulfilled the task requirement by producing a piece of descriptive discourse in compliance with the schematic features associated with this type of discourse. As explained earlier, the descriptive discourse generic structure requires an element of identification and one of description. Accordingly, the test-taker first identified the item they would talk about, ie, their car (line 74, in bold). The item was coded as 'id a' (ie, the first element in the schematic structure) and 'a' was provided as a cue to spot the identification element in the test-taker's performance. After identifying the item for the talk, the test-taker moved on to describe the car's properties and characteristics as shown in lines 75, 76 and 77 (under codes 'de a', 'de b', 'de c', 'de d', and 'de e', respectively).

The qualitative analysis revealed that when comparing the performances across the three band levels, the Level 5 performance showed a lower degree of compliance to the descriptive discourse structure than the Levels 6 and 7 performances. That is, some test-takers at Level 5 failed to fully comply with the descriptive discourse generic structure by either including or developing the description of the item identified. The example below illustrates one such case.

- | | | |
|-----|--|------|
| 147 | iser my favourite news- (.) newspaper is er::::m (.) the new | |
| 148 | er thailand (a) (.6) and the parts of (.) i like to read in the:: | id a |
| 149 | (4) education .hhhh (.) because (.2) in the future i would | |
| 150 | like (.3) to be a 2.teacher::? (.3) and if i have a free time | |
| 151 | (.) i would like to stud- (.4) dy (.4) i would like to know | |
| 152 | ifer:: what the matter in the:: (.2) education (.5) yes | |
| 153 | (.8) an:: (.9) there's (opposite) people read it (.6) who | |
| 154 | is a:: (.5) 3.educator or who is interest- (.6) interested in | |
| 155 | the:: education (.5) yes (2) an:: (.2) h- i::: (.8) i | |
| 156 | prefer:: (.) in the:: (.2) newspaper maybe it's because (1.4) | |
| 157 | it'ser:: cheaper than magazine. (.5) yes and it have a:: | |
| 158 | day- daily (.3) daily newspaper (.3) yeah (.5) and it's er | |
| 159 | have many kind of er:: (.8) newspaper (.5) it have an::: | |
| 160 | ((inaudible)) (.6) to:: (.5) read (.3) in them (.) | |
| 161 | newspaper (.5) yes= | |

(Level 5 ID216)

(Note: id – identification)

In this example, the test-taker was asked to describe their favourite newspaper, and so mentioned 'new thailand' (in line 148, coded as 'id a'). By doing this, their discourse partially conformed to descriptive discourse generic structure as it contained an identifying element (ie, the name of their favourite newspaper). However, the test-taker did not describe the newspaper 'new thailand' in the rest of their speech as required. Instead they mostly talked about why they like to read the education section and what they would like to do in future (in lines 149–155), and then eventually they moved on to a comparison of the cost of newspapers and magazines in lines 156 and 157. The test-taker failed to develop the descriptive element and, therefore, this performance was assessed as partially complying with descriptive discourse schematic structure.

In contrast to this example above, the extract below, taken from a Level 7 performance displayed a higher degree of compliance with the schematic structure of descriptive text.

199	magazine called ((inaudible)) 'Friday' (a) .hhh so (.3) (win:) this	id a
200	is actually m::: (.2) it covers everything (.3) y'know it's	
201	where a according to the (emirates wife) the news has been	
202	given in this magazine (.3) and (.) m: starting from <i>the first</i>	de a
203	<i>page all news about the emirates (a)</i> (.4) the dubai because the	
204	printing is done in dubai .hhh then the <i>the remaining pages one</i>	de b
205	<i>by one it covers with the other (.2) emirates it is there</i>	
206	<i>are seven emirates anyway so it is covers by (.2)</i>	
207	<i>according to the emirates (b)</i> .hh and most of the time the	
208	<i>political issues are given in the first page (c)</i> and er .hh it	de c
209	<i>describes about any changes especially the immigrant rules (d)</i>	de d
210	whatever is changing y'know (.3) er::m according to the m:	
211	<i>fresh news what is called hot news (e)</i> is being explained .hhh and	de e
212	erm:: (.4) er second thing is er one m er m: the- there are	
213	some important categories which they're they are explaining	
214	About <i>the health column is a .hherm (.2) the thethe editor</i>	
215	<i>writes about the health column (f)</i> and er some of the .hh special	de f

(Level 7 ID227)

(Notes: id – identification, de – description)

The test-taker was asked to describe their favourite magazine to which they responded by identifying the publication as 'Friday' (line 199, coded as 'id a'). After introducing the element of identification and, in agreement with descriptive discourse structure, the test-taker proceeded to appropriately describe the publication. Accordingly, the test-taker provided a description of the sections of the magazine (political news, hot news, health advice), the type of information published in each section, and the pages this information is published on (text in italics in lines 202–203, 204–207, 208, 209, 211, 214–215, coded as 'de a', 'de b', 'de c', 'de d', 'de e', and 'de f' respectively). This test-taker performance was awarded score of 1.0 as fully complying with descriptive discourse generic structure. The quantitative and qualitative results revealed that higher-level test-takers are better able to structure their texts by complying with the expected text structure than lower level test-takers. It should be noted that, as explained in the method section, the task card given to the test-takers explicitly stated that test-takers were expected not only to provide description, but also to explain (for the question about something they own which is very important to them) "where you got it from; how long you have had it; what you use it for; and explain why it is important to you". Therefore, there is a tendency for lower level test-takers to focus on the second part of the instruction more than the first part (ie, description).

5.2.2 Theme–rheme development

As explained in the methodology section, the performance of the 58 test-takers was analysed first in light of identifying theme and rheme, and then a theme pattern (ie, theme–reiteration/constant theme, zigzag/linear theme, split/multiple theme patterns, or a combination of one or two theme patterns) observed in the study was identified. The descriptive statistics of main clauses and the categorisation of theme patterns according to the levels is summarised in Tables 14 and 15 respectively. The higher the level, the larger the number of mean clauses produced, but the difference in the mean number of main clauses across the three levels was not significant ($F(2, 55) = 2.391$, ns). Similarly, patterns of thematic progressions were found as expected. That is, while more complex patterns were observed in Level 7 test-taker performances, with no theme–reiteration/constant theme patterns, the thematic progression of one third of Level 5 test-takers followed the theme–reiteration/constant theme pattern.

Level	Mean no of main clauses	SD	Min	Max
5 (N=18)	10.11	4.26	3	18
6 (N=20)	12.40	4.28	6	21
7 (N=20)	13.40	5.46	6	25

Table 14: Descriptive statistics of quantity of text in each proficiency group

Level	Theme–reiteration (constant theme)	Zigzag/ linear theme	Combination of theme–reiteration/ constant and zigzag/ linear theme patterns
5 (N=18)	6	0	12
6 (N=20)	2	0	18
7 (N=20)	0	0	20

Table 15: Theme–rheme patterns in each proficiency group

The following section presents the findings of the detailed examination of the theme–rheme progression. On the whole, across the proficiency levels, test-takers were able to structure the text by selecting an element to act as a point of departure in the message and subsequently develop this first element in the remainder of the message in order to achieve information flow. As shown in Table 15 above, the thematic progression observed in the test-taker performance in the study consisted mostly of patterns of theme–reiteration/constant theme or zigzag/linear theme and the combination of the two patterns across the levels. There was no distinctive difference in the frequency of the patterns across the levels. Nevertheless, simple theme–rheme development (thematic reiteration/constant theme pattern), shown in the example below (Figure 8) was more frequently observed in Level 5 test-takers than Level 7 performances. This example contains five main clauses. Out of the five rhemes identified, no element in the five rhemes was taken up in the subsequent themes. This indicates that the new information introduced in the rhemes was not further followed up. As noted at the end of the previous section, the test-taker focused more on the second part of the task instruction than the first part (ie, description).

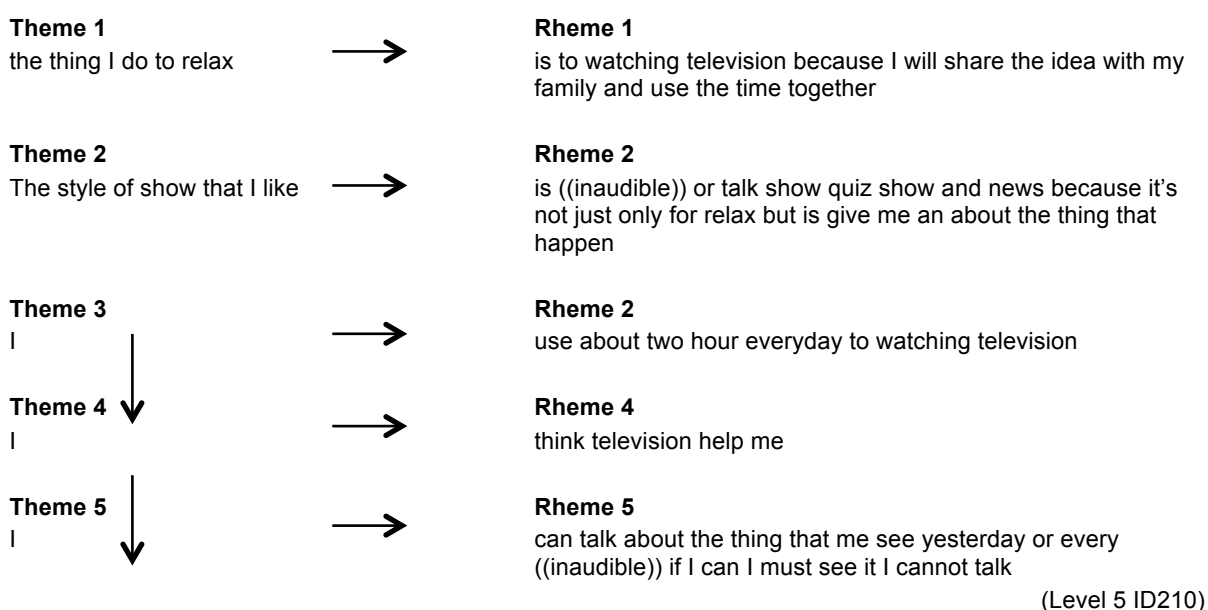
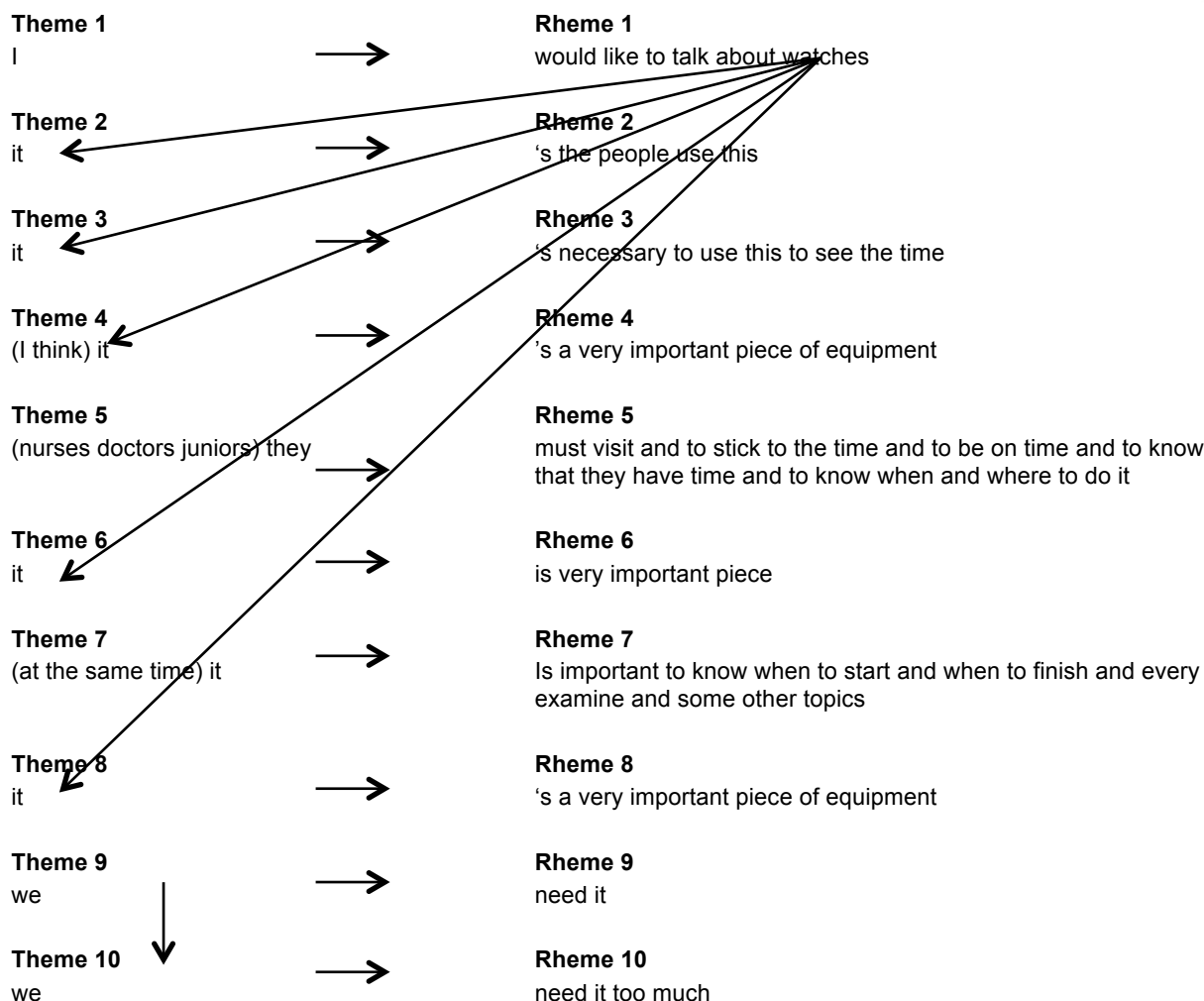


Figure 8: Theme–rheme development (Reiteration/constant theme pattern) (Level 5-1)

Compared with the example above, the theme–rheme development in Figure 9 taken from Level 6 seems more complex as a combination of a zigzag/linear theme pattern and a theme–reiteration/constant theme pattern is observed. In this example, the rheme identified in the first main clause was taken up as a theme five times in the subsequent clauses (ie, themes 2, 3, 4, 6 and 8). Through the zigzag/linear theme pattern cohesion was achieved and the theme introduced earlier was further developed, but the test-taker repeated ‘very important’ in rhemes 4, 6 and 8, and instead of providing further description of the item chosen as a useful piece of equipment, the test-taker provided reasons why they thought it was useful.



Note: In the analysis of this extract, the element 'I think' in theme 4 is considered a conversation filler and thus not considered in the thematic progression analysis.

Figure 9: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 6-1)

A similar thematic progression to the Level 6 performance in Figure 9 is observed in the Level 7 performance shown in Figure 10. In this example, a clear thematic development is observed in a zigzag/linear theme pattern. That is, the item chosen for the talk was further taken up in the subsequent clauses to achieve cohesion as shown in themes 3, 4 and 5. In this example, unlike the example above (Figure 9), the theme 'it' in themes 3, 4 and 5 followed up from the previous rheme (rheme 2) with additional information about the test-taker's car shown in rhemes 3, 4 and 5.

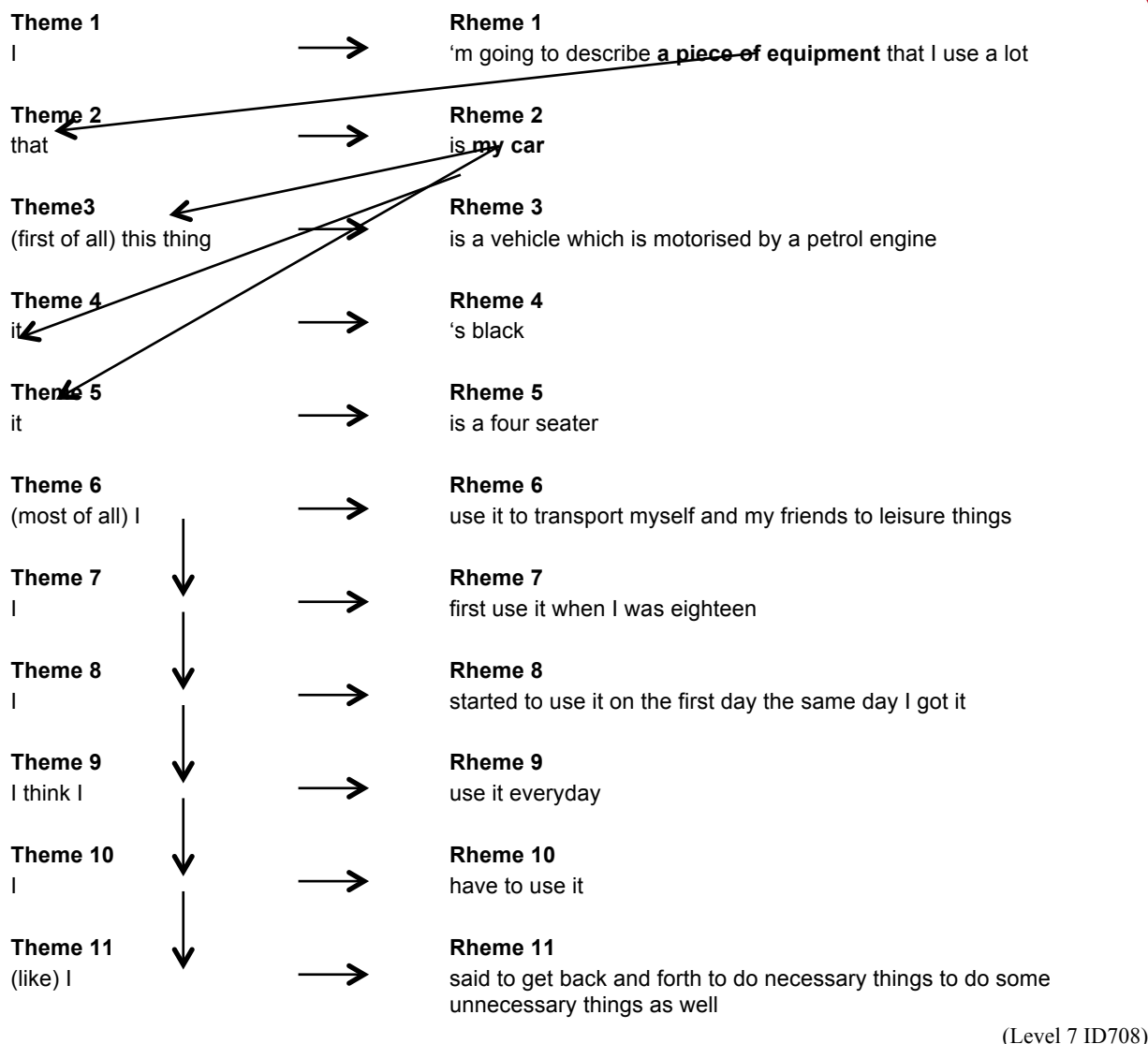


Figure 10: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 7-1)

The three examples above present a text with a relatively smaller number of main clauses resulting in simpler theme–rheme progression. As shown in Table 2, there is considerable variation in the number of main clauses in each level. In the following examples, thematic progression is observed in texts with a relatively larger number of main clauses (Figures 11, 12, and 13).

The example shown in Figure 11 below corresponds to performance at Level 5. Compared with the previous example taken from the same level, as shown in Figure 8, more thematic development is seen in this example featuring a theme–reiteration/constant theme pattern with two sets of constant themes (ie, ‘I’ as themes 1, 2, 5, 6, 7, 10, 12, 13, 14, as well as ‘it’ in themes 3, 4 and 9) and a zigzag-linear theme pattern observed four times in themes 3, 4, 9 and 11. The test-taker’s first response to the examiner’s question about how to relax was to discuss going to the poultry house (rheme 2). This is taken up in themes 3 and 4, but instead of further describing this, the test-taker explained why going to the poultry houses is relaxing (in themes 3, 4 and 9). Though the information flow is observed through this thematic progression, further information about relaxation is not given. However, the element in rheme 9 is repeated in theme 11 and further information is given about the poultry house mentioned earlier.

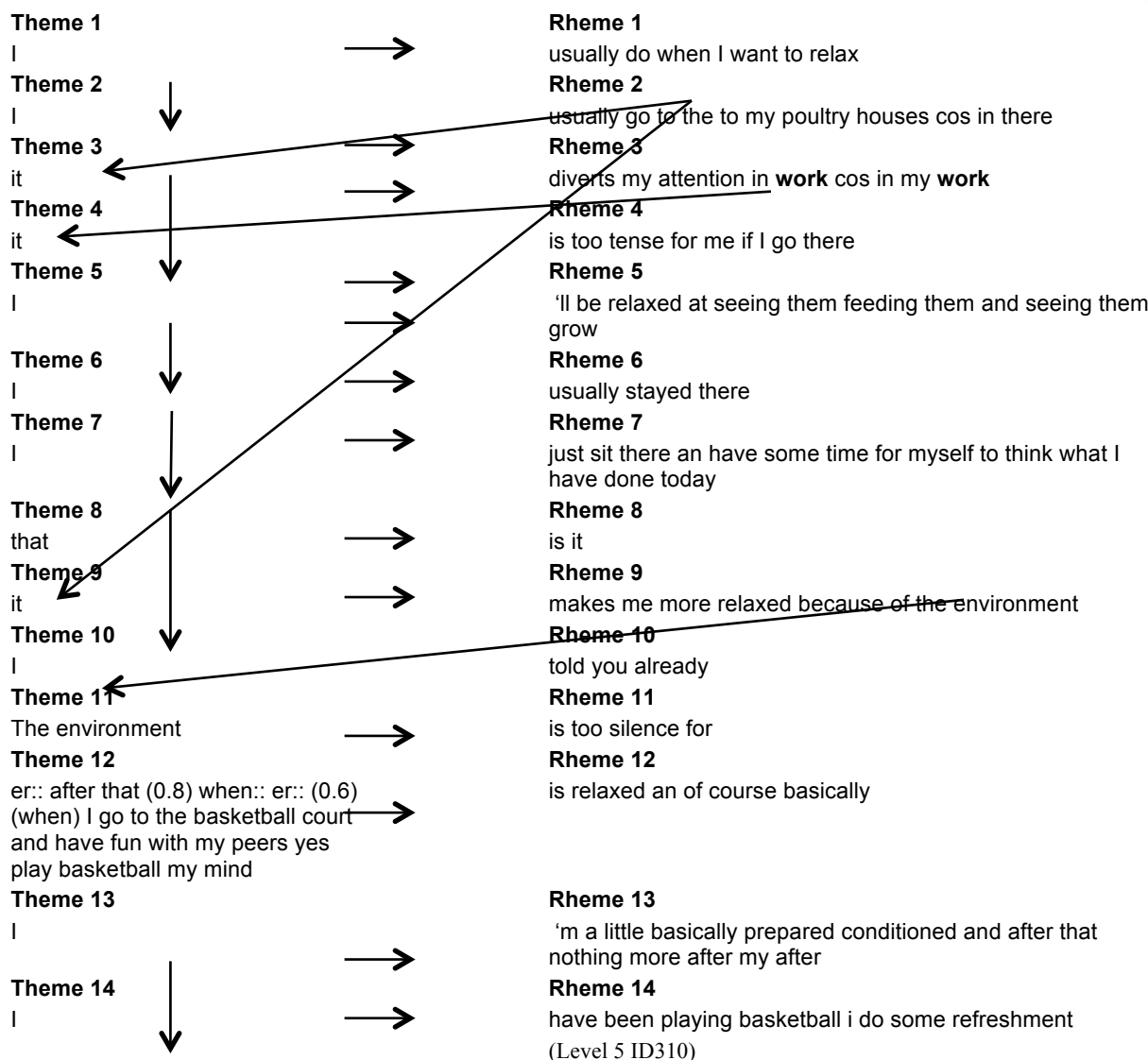
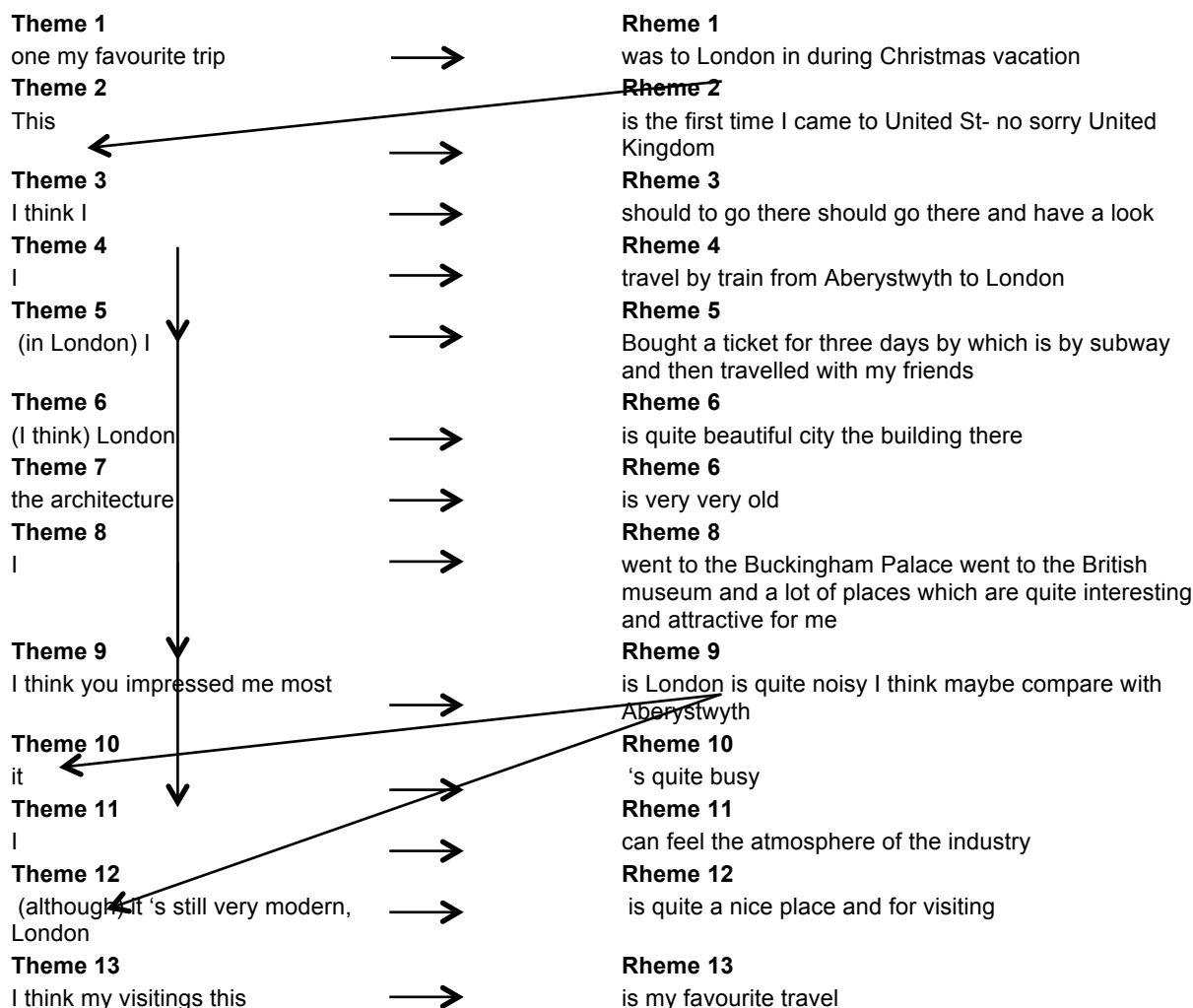


Figure 11: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 5-2)

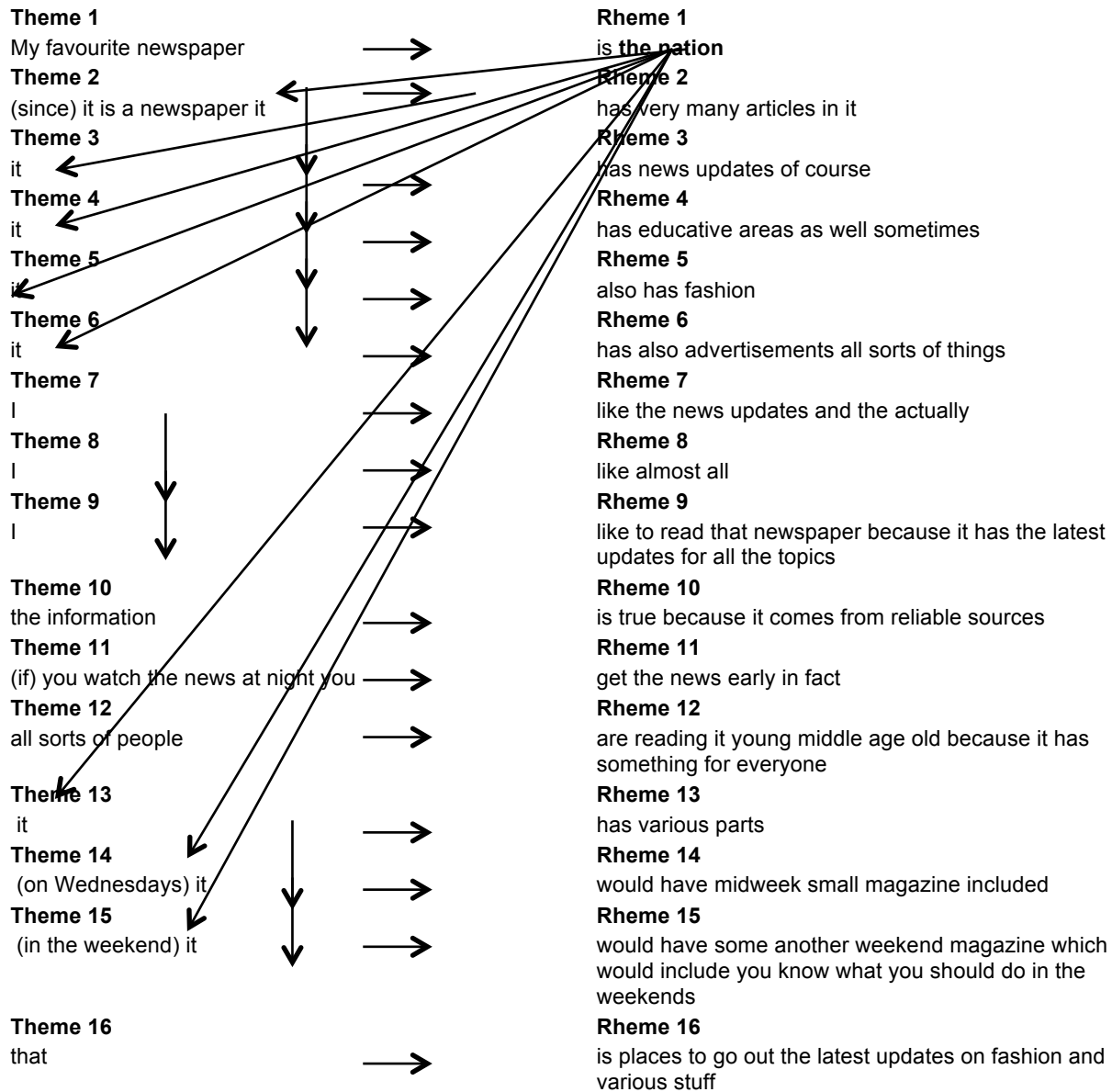
In Figure 12 below, the test-taker describes their favourite trip in rheme 1, and this is taken up in theme 2 though it is not clear what they are referring to with ‘this’. The test-taker’s favourite trip is further described in theme–reiteration/constant theme patterns in themes 3–6, and 8. The same theme ‘I’ maintains the information flow. ‘London’ in rheme 9 is taken up in ‘theme 11’ as ‘it’ and followed up again in theme 12.



(Level 6 ID601)

Figure 12: Theme–rheme development (combination of theme-reiteration/constant theme and zigzag/linear theme patterns) (Level 6-2)

Figure 13 below is taken from a performance at Level 7. This sample displays a thematic progression structure consisting of two theme–rheme patterns. A theme–reiteration/constant theme pattern with two sets of constant themes (ie, ‘it’ as themes 2, 3, 4, 5, 6, 13, 14 and 15; ‘I’ in themes 7, 8 and 9) is observed. In the thematic progression seen in themes 7, 8 and 9, the test-taker correctly followed the instruction given in the task card by adding an explanation of the reason why they like this newspaper. Compared with the examples above, taken from the lower levels, a zigzag/linear theme pattern where the element ‘the nation’ in rheme 1 was taken up in the subsequent themes (theme 3, 4, 5, 6, 7, 13, 14 and 15) achieved a high level of cohesion and results in smooth information flow. The item chosen for the talk was well developed through the text resulting in rich content. It should be noted that the element ‘it’ in theme 2 is a reference for the referent ‘the nation’ in theme 1, and themes 3, 4, 5 and 6 are picked up from ‘it’ in rheme 2 in a strict sense.



(Level 7 ID201)

Figure 13: Theme–rheme development (combination of theme–reiteration/constant theme and zigzag/linear theme patterns) (Level 7-1)

To summarise the findings above, identifying thematic progression patterns and counting the main clauses does not necessarily contribute to characterising the thematic development observed in test-taker performance at the three levels, but close examination of the data revealed differences attributable to variations in proficiency levels. As shown in the previous examples (ie, at Levels 5, 6, and 7), the performances did vary in the number of constant theme sets displayed in the theme–reiteration patterns where Levels 6 and 7, in particular, included more sets. Additionally, zigzag/linear patterns in more proficient performances tend to display a more complex configuration of the pattern itself. The dynamics of a rheme element being picked up as a subsequent theme in the flow of discourse achieves a high level of cohesion and develops richer content. These findings were observed in both shorter and longer speech segments in higher proficiency test-taker performance. As noted earlier, no split/multiple theme pattern was observed in the text produced by the 58 test-takers of this study. This is partly attributable to the nature of oral discourse in an interview setting, which is shorter than written texts.

5.3 Lexical richness

Table 16 summarises descriptive statistics of the lexical richness measures using *VocabProfile* (Cobb 2013). Type and token measures are as expected: as the level increases, more words and more different types of words are produced. One-way ANOVA analysis shows significant differences across the levels for these two measures (see Table 17 below). However, the effect size is marginal ($\eta^2 = .143$) for token, and small for type ($\eta^2 = .300$). Post hoc analysis (Bonferroni correction) shows the difference is found for token between Levels 5 and 7 ($p = .011$), and for type between Levels 5 and 6 ($p = .005$) and Levels 5 and 7 ($p = .0001$).

The three ratio measures (type-token, token per type, and lexical density) in the three groups are not very different from each other, except that lexical density shows a significant difference across the levels. The difference lies between Levels 6 and 7 ($p = .005$) and Levels 5 and 7 ($p = .034$). The percentages of the four word lists are only found to be significantly different for K1 and Off List.

Post-hoc analysis shows that the significant differences lie between Levels 5 and 6 ($p = .034$) and Levels 6 and 7 ($p = .008$) for K1 and between Levels 6 and 7 ($p = .012$) for Off List.

	Level 5		Level 6		Level 7	
	Mean	SD	Mean	SD	Mean	SD
Token	580.22	23.74	737.10	283.78	887.45	392.69
Type	145.17	37.30	171.20	45.33	226.95	7.27
Type token	.27	.07	.25	.05	.27	.06
Token per type	3.91	.95	4.18	.78	3.77	.69
Lexical density	.64	.06	.65	.06	.59	.05
K1	86.02	3.70	82.39	5.53	86.65	3.14
K2	2.81	1.56	2.42	1.15	2.56	.87
AWL	2.20	1.18	2.60	1.33	2.56	1.16
Off List	9.00	4.50	12.59	5.56	8.24	3.44

Table 16: Descriptive statistics of lexical richness measures

	Type III Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
Token	894515.26	2	447257.63	4.59	.014	.143
Type	67156.45	2	33578.23	11.80	.000	.300
Type token	.008	2	.004	1.15	.324	.040
Token per type	1.74	2	.87	1.33	.274	.046
Lexical density	.041	2	.02	6.08	.004	.181
K1	209.09	2	104.54	5.75	.005	.173
K2	1.50	2	.75	.51	.605	.018
AWL	1.87	2	.94	.62	.541	.022
Off List	214.18	2	107.09	5.10	.009	.156

Table 17: Results of one-way ANOVA analysis (lexical richness)

Figure 14 visually presents the distribution of the four lexical categories summarised in Table 16 (ie, K1, K2, AWL and Off List). Surprisingly, the distribution is very similar in all three levels' performance. The majority of the words used in the test performance are classified as 1000 level. A very small portion of the words is classified as above 2000 (K2) and as Academic Word List.

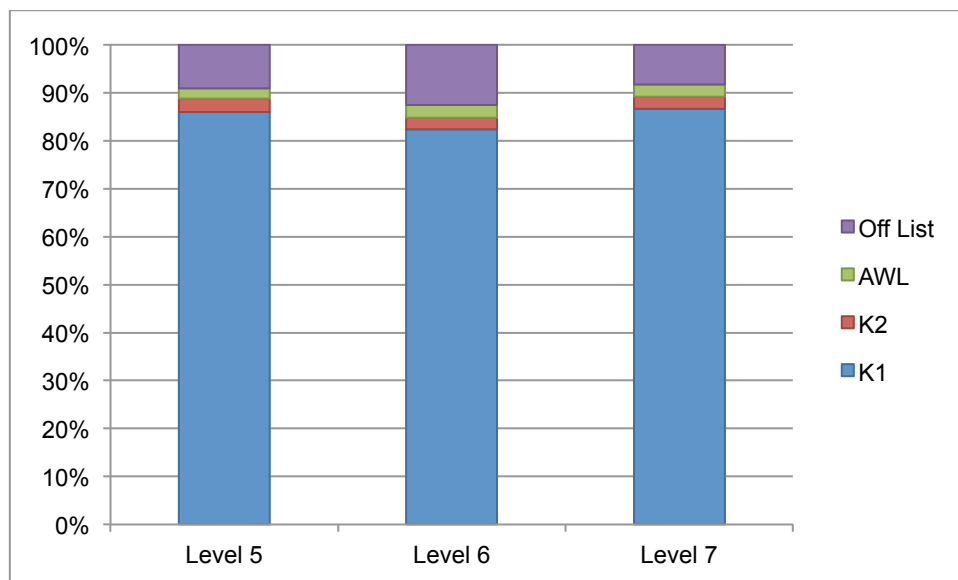


Figure 14: Distribution of the four word categories

5.4 Summary of the results

In this section, a summary of the results reported in the previous section is presented first, followed by an explanation of how these findings are correlated to the two items of the IELTS Speaking Descriptors (Public Version). Firstly, summarising the results of the findings above, a number of coherence and cohesive devices and features of lexical richness vary according to their assessed proficiency levels. Significant differences across the levels in the expected directions are found for the following features, though the effect sizes of the significant differences are all marginal or small.

- Cohesive devices:
 - Comparative conjunction
 - Accuracy of use of referential expressions
 - Lexical cohesions – Hyponymy and repetitions
- Lexical richness:
 - Word-token
 - Word type
 - Lexical density
 - K1
 - Off List

It should be noted that SDs of most features are very large, which indicates large individual variations in the test-taker performances. Table 18 provides a summary of the findings of the statistical analysis. Furthermore, the qualitative analysis provides further insights into the test-takers' use of the various cohesive and coherence devices under study. Some features such as the compliance index of the text generic structure show clear differences according to the band levels, even though the statistical analysis reveals no significant difference across the levels.

On the whole a clear difference is observed between Levels 5 and 7, but the difference between the adjacent levels (ie, Levels 5 and 6, and Levels 6 and 7) is not very clear. These findings are further examined in the Discussion section.

Category	Sub-category	Difference across the three levels	Effect size	Post-hoc analysis
Cohesion				
Conjunction				
	Additive			
	Comparative	✓		Level 5 ≠ Level 6; Level 5 ≠ Level 7
	Temporal			
	Consequential			
	Total			
Reference				
	No. of referential expressions			
	Accuracy (%)	✓		Level 6 ≠ Level 7; Level 5 ≠ Level 7
Lexical cohesion				
	Synonymy			
	Antonymy			
	Hyponymy	✓		Level 5 ≠ Level 7
	Repetition	✓		Level 6 ≠ Level 7; Level 5 ≠ Level 7
	Meronymy			
Coherence				
	Text generic structure			
	No. of main clauses			
Lexical richness				
	Token	✓	.143	Level 5 ≠ Level 7
	Type	✓	.300	Level 5 ≠ Level 6; Level 5 ≠ Level 7
	Type token			
	Token per type			
	Lexical density	✓	.181	Level 6 ≠ Level 7; Level 5 ≠ Level 7
	K1	✓	.173	Level 5 ≠ Level 6; Level 6 ≠ Level 7
	K2			
	AWL			
	Off List	✓	.156	Level 6 ≠ Level 7

Table 18: Summary of the results of quantitative analysis

5.5 Co-reference with the IELTS Speaking Band Descriptor

The discussion below explains how these findings may correlate to the public version of the IELTS Speaking Band Descriptors (fluency and coherence) based on the results summarised above. The full descriptors of all four aspects can be found online at http://www.ielts.org/researchers/score_processing_and_reporting.aspx.

5.5.1 Level 7

- *Speaks at length without noticeable effort or loss of coherence*
- *May demonstrate language-related hesitation at times, or some repetition and/or self-correction*
- *Uses a range of connectives and discourse markers with some flexibility*

As reported above, the Level 7 test-takers' mean score for one of the two features of coherence (ie, text generic structure, number of main clauses in thematic progression) in the study is highest. That means test-takers at this level are able to produce a text complying with the text type identified in the examiner's instructions. Moreover, as shown in the conjunction devices (see Table 2 and Figure 1), although the proportion of additive and consequential conjunctions is far larger than temporal and comparative conjunctions across the levels, the proportion of each of the four conjunctions is less unbalanced than Level 5 and 6 performances. In fact, the use of comparative conjunctions is significantly higher than the lower levels.

On the whole, the test-takers were able to use a wider variety of cohesive devices accurately, as shown in the lexical cohesion scores and the percentage of accurate use of referential expressions. As shown in the analysis of theme–rheme development, the texts produced by test-takers at this level achieved a higher level of cohesiveness through more complex thematic progression and rich information flow, resulting in comprehensible texts with rich content.

5.5.2 Level 6

- *Is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation*
- *Uses a range of connectives and discourse markers but not always appropriately*

The test-takers' compliance with the generic structure at this level was not as accurate as Level 7 test-takers, but their score was substantially higher than the Level 5 test-takers. Furthermore, Level 6 test-takers' performance was less balanced in the proportion of the use of the four conjunctions compared with the Level 7 test-takers. However, significantly higher frequency of comparative conjunctions is observed in Level 6 test-taker performance than in Level 5 test-takers and the frequency of additive conjunctions is the highest of the three levels. Although the frequency of the referential expression of Level 6 test-takers is the lowest of the three levels, the percentage of accurate use of the referential expression is higher than Level 5 test-takers.

5.5.3 Level 5

- *Usually maintains flow of speech but uses repetition, self correction and/or slow speech to keep going*
- *May over-use certain connectives and discourse markers*
- *Produces simple speech fluently, but more complex communication causes fluency problems*

The results of the current study reveal that test-takers at this level are able to use connectives and discourse markers but, compared with the test-takers at higher levels, approximately 85% of the total number of conjunctions observed in their performances were either consequential or additive conjunctions. Lexical expressions were frequently observed in their speech, but the percentage of accurate use of the lexical expressions was the lowest of the three levels. Level 5 test-takers mostly used repetition devices to achieve lexical cohesion. For compliance with the generic structure, the test-takers were able to use the expected structure.

Test-takers at this level tended to focus on explaining reasons for their choice of the item rather than describing it. This was evidenced in the thematic progression pattern observed in the test-taker performance at this level.

6 DISCUSSION AND CONCLUSION

The current study examined various features of discourse competence observed in the performance of 58 test-takers in IELTS Speaking Part 2. In particular, the data analysis focused on the three cohesive devices and two aspects of coherence. In addition, lexical richness observed in the performances was evaluated. In order to identify the distinctive features in test-taker performance at each of the three levels and to compare the performances across the levels, we first quantified the results of the data analysis of the three features of cohesive devices for statistical analysis. Then detailed qualitative analysis followed to validate the results of the statistical analysis and also to provide further insights into the characteristics of test-taker performances.

The descriptive statistics of the frequency of various cohesive devices, including one conjunction (comparative), accuracy of referential expressions, and aspects of coherence (ie, text generic structure, number of main clauses in thematic progression) was in the expected direction. Higher proficiency test-takers more frequently used a variety of cohesive devices, and their referential expression was more accurate than lower proficiency test-takers. The structure of the text produced by higher-level test-takers closely conformed to the expected text structure (ie, description) and contained rich content. In addition, the speech produced by Level 6 and 7 test-takers complied with the expected structure (ie, description) more than Level 5 test-takers' speech.

As summarised in Table 17 at the end of the Results section, the findings showed a statistical difference for some of the features. However, individual variations were very large across the features and levels. As expected, more distinctive differences were observed between Level 5 and Level 7 than in the adjoining levels.

These findings indicate that in many aspects of discourse competence, Level 7 test-takers tend to demonstrate better control of cohesive and coherence devices, which results in more comprehensive texts with rich content; but the picture is not clear-cut. As noted above, individual differences were large and significant differences were observed in a few aspects only, not in all aspects under study. Close examination of the transcribed data revealed that the length of test-taker speech measured by word-token varied across all three levels.

On the whole, Level 7 test-takers spoke longer than Level 5 or 6 as shown in the descriptive statistics of word-token in Table 16 (see also Appendix 2 for detailed results of descriptive statistics of lexical richness), but a wide range of word-tokens in all levels was observed (ie, Level 5, 258–947; Level 6, 205–1293; Level 7, 319–1552). In fact, the lowest number of word-tokens was not very different across the levels. This might be partly due to the required length of speech (ie, 1–2 mins) and test-takers' strategies to achieve the task performance. Some test-takers regardless of level, preferred to give a short description, but others took time and produced a lengthy speech to describe the item they chose for their talk. Since we did not examine the speed of test-takers' speech, it is not known if there was a difference in speech rate across the levels. However, based on previous studies (eg, Brown et al. 2005; Iwashita, Brown, McNamara & O'Hagan 2008) and frequent pauses observed in the transcribed data of lower level test-takers, it is possible that higher level test-takers spoke faster and more fluently, resulting in a greater amount of speech. This shows Level 7 test-takers had more opportunities to use a variety of cohesive and coherence devices to make their speech structured and comprehensive.

The non-significant difference in the features of cohesion and coherence could also be attributed to the very low frequency of the devices used by test-takers. Similar findings have also been reported in previous studies of discourse competence (eg, Brown et al. 2005; Banerjee et al. 2004). As shown in the descriptive statistics of frequency (per 100 words) of conjunctions (Table 3), referential expressions (Table 6), and lexical cohesions (Table 11), the frequency of the use of three cohesion devices was 3.75–4.15 for conjunction (total of four types of conjunction use), 2.14–2.60 for lexical expressions, and 1.70–1.94 for lexical cohesion (total of five types of lexical cohesion devices) respectively.

In addition, as reported in the Methodology section, ellipsis and substitution devices were rarely observed in the test-taker performances in the current study and were therefore excluded from the analysis. These findings are not very surprising considering the length of speech (ie, 1–2 minutes), but as shown in the qualitative analysis, there are some clear differences between levels in various features, including the types of conjunctions, lexical cohesion devices used in each level, text structure, and thematic progression.

The few instances of cohesion and coherence devices observed in the current study might also be explained in terms of test-takers' awareness of these devices. In Speaking Part 2, test-takers were given two minutes to prepare for their talk after the examiner's instructions. The findings of planning studies in pedagogic contexts have shown that participants mostly spend the planning time thinking about the strategies to approach the task, vocabulary, and forms to use in task performance (eg, Ortega 1999; Sangarun 2005; Tajima 2003), but no study has reported that coherence and cohesive devices were taken into account. Considering these findings, it is possible that the test-takers in the current study were also concerned with features reported in the planning studies.

Furthermore, in other planning studies in which participants were given instructions about what to focus on during planning time (referred to as 'guided planning'), the instructions were mostly about specific features of forms (eg, Foster & Skehan 1996; Mochizuki & Ortega 2008; Skehan & Foster 2005; Yuan & Ellis 2003). These planning studies showed that learners considered the structure of the talk during planning time, but were not expected to attend to the structure of their speech or to linking devices that would make the text more comprehensible. Even if researchers and educators alike stress the importance of discourse competence, discourse markers are not always taught (Hellerman & Vergen 2007), so the way learners/test-takers structure the text using various cohesive devices seems to largely depend on learners. Lee reported that explicit teaching of the concept of coherence enhanced students' awareness and the coherence of their writing (2002).

The test-takers who attended an IELTS preparation course prior to taking the test might have been instructed to attend to coherence and cohesion, as they are clearly stated in the band descriptors and, therefore, they might have considered these features during the two minute preparation time. Nevertheless, without information about the test-takers' possible attendance at an IELTS test preparation course or what they did during preparation, this issue remains speculation.

As reported above, we found cohesion and coherence devices are infrequent in test-taker performance and that there are some, but not distinctive, differences in the use of discourse devices. These findings may indicate that the use of cohesion and cohesive devices is not a serious concern for test-takers in making the text more comprehensible in the short speech required for IELTS Speaking Part 2. However, disregarding these discursive features in the assessment may threaten the predictive validity of the test as many test-takers who take IELTS in order to study at university will be required to produce long stretches of speech (eg, oral presentations) during their study. As reported in the studies examining the quality of oral production by non-native speaking teaching assistants in universities in the US, lack of discourse markers by second language speakers causes some comprehension problems (eg, Tyler 1992; Williams 1992). Furthermore, the research findings revealed that native speakers use different discourse devices according to text types (eg, Geva 1992).

Therefore, it is possible that the cognitive demand imposed on task performance may result in the limited use of discourse devices in the text that the test-takers produce. But, as the current study examined the test-taker performance of one type of task (ie, description), it is not known how test-takers might employ different discourse devices in performing different types of tasks, which produce different types of texts.

Future studies investigating test-takers' use of discourse devices in a variety of texts will provide further insights into the characteristics of discourse competence in test-taker performance. Also use of think-aloud protocols will reveal test-takers' strategies for approaching a task and constructing a text during preparation time.

In addition, analysis of the task performance collected from a wider range of proficiency levels than the three levels examined in the current study will assist further understanding of the impact of proficiency levels on the aspects of performance under study.

Despite some limitations explained above, the current study has implications for classroom teachers. Explicit teaching of discursial features might help learners to raise their awareness of discourse competence. The findings also provide useful information about the use of various discourse devices observed in the current format of IELTS Speaking Part 2 for future development of tasks and revisions.

REFERENCES

- Alderson, JC & Banerjee, J, 2002, 'State of the art review: language testing and assessment (part two)', *Language Teaching*, vol. 35, no. 2, pp. 79–113
- Bachman, L, 1990, *Fundamental considerations in language testing*, Oxford University Press, Oxford
- Bachman, L & Palmer, A, 1982, 'The construct validation of some components of language proficiency', *TESOL Quarterly*, vol. 16, pp. 449–465
- Bachman, L & Palmer, A, 1996, *Language testing in practice*, Oxford University Press, Oxford
- Bachman, L & Palmer, A, 2010, *Language assessment in practice*, Oxford University Press, Oxford
- Banerjee, J, Franceschina, F & Smith, AM, 2007, *Documenting features of written language production typical at different IELTS band score levels*, IELTS Research Report No. 7, the British Council/University of Cambridge Local Examinations Syndicate
- Brooks, L, 2009, 'Interacting in pairs in a test of oral proficiency: Co-constructing a better performance', *Language Testing*, vol. 26, no. 3, pp. 341–366
- Brown, A, Iwashita, N & McNamara, T, 2005, *An examination of rater orientations and test taker performance on English-for-academic-purposes speaking tasks*, Monograph Series MS-29, Educational Testing Service, Princeton, NJ
- Cameron, CA, Lee, K, Webster, S, Munro, K, Hunt, AK & Linton, MJ, 1995, 'Text cohesion in children's narrative writing', *Applied Psycholinguistics*, vol. 16, no. 2, pp. 257–269
- Canale, M, 1983, 'On some dimensions of language proficiency' in J Oller (ed.), *Issues in language testing research*, Newbury House, Rowley, MA, pp. 333–342
- Canale, M & Swain, M, 1980, 'Theoretical bases of communicative approaches to second language teaching and testing', *Applied Linguistics*, vol. 1, no. 1, pp. 1–47
- Celc -Murcia, M, 2008, 'Rethinking the role of communicative competence in language teaching', in E Alcon Soler & MP Safont Jorda (eds), *Intercultural language use and language learning*, Springer, The Netherlands, pp. 41–57
- Celc -Murcia, M, Dornyei, Z & Thurrel, S, 1995, 'Communicative competence: A pedagogically motivated model with content specifications', *Issues in Applied Linguistics*, vol. 6, pp. 5–35
- Chalhoub-Deville, M, 1997, 'Theoretical models, assessment frameworks and test construction', *Language Testing*, vol. 14, no. 1, pp. 3–22
- Chalhoub-Deville, M, 2003, 'Second language interaction: Current perspectives and future trends', *Language Testing*, vol. 20, pp. 369–383
- Cobb, T, 2013, *VP Classic (Version 4)* [Computer program]. University of Qu bec, Montr al, 12 January 2013, <http://www.lexutor.ca/vp/>
- Coxhead, A, 2000, 'A new academic word list', *TESOL Quarterly*, vol. 34, no. 2, pp. 213–238
- De Beaugrande, R & Dressler, WU, 1981, *Introduction to text linguistics*, Longman, New York
- Douglas, D, 1994 'Quantity and quality in speaking test performance', *Language Testing*, vol. 11, no. 2, pp. 125–144
- Douglas, D & Selinker, L, 1992, 'Analysing oral proficiency test performance in general and specific purpose contexts', *System*, vol. 20, pp. 317–328
- Douglas, D & Selinker, L, 1993, 'Performance on a general versus a field-specific test of speaking proficiency by international teaching assistants' in D Douglas & C Chapelle (eds), *A new decade of language testing research*, TESOL Publications, Alexandria, VA, pp. 235–256
- Eggs, S, 2012, *An introduction to systemic functional linguistics*. Continuing International Publishing Group, Manchester
- Foster, P & Skehan, P, 1996, 'The influence of planning on performance in task-based learning', *Studies in Second Language Acquisition*, vol. 18, no. 3, pp. 299–324
- Fuller, JM, 2003, 'Discourse marker use across speech contexts: A comparison of native and non-native speaker performance', *Multilingua*, vol. 22, pp. 185–208
- Fulcher, G, 1996, 'Does thick description lead to smart tests? A data-based approach to rating scale construction', *Language Testing*, vol. 13, no. 2, pp. 208–238
- Fung, L & Carter, R, 2007, 'Discourse markers and spoken English: Native and learner use in pedagogic settings', *Applied Linguistics*, vol. 28, no. 3, pp. 410–439
- Gee, JP, 1998, 'What is literacy?', in V Zamel & R Spack (eds), *Negotiating academic literacies: Teaching and learning across languages and cultures*, Routledge, New York, pp. 51–60
- Geva, E, 1992, 'The role of conjunctions in L2 text comprehension', *TESOL Quarterly*, vol. 26, no. 4, pp. 731–747
- Gonzalez, M, 2005, 'Pragmatic markers and discourse coherence relations in English and Catalan oral narrative', *Discourse Studies*, vol. 7, no. 1, pp. 53–86
- Guara-Tavares, M, 2008, 'Pre-task planning, working memory capacity and L2 speech performance', Unpublished doctoral thesis, Universidade Federal de Santa Catarina, Brazil.
- Hall, JK & Doehler, PS, 2011, 'L2 interactional competence and development', in JK Hall, J Hellermann & PS Doehler, (eds), *L2 interactional competence and development*, Multilingual Matters, Clevedon, pp. 1–18
- Halliday, MAK & Hasan, R, 1976, *Cohesion in English*, Longman, London
- Halliday, MAK & Matthiessen, CMIM, 2013, *Halliday's introduction to functional grammar*. 4th edn, Routledge, New York

- Halliday, MAK & Webster, J, 2009, *Continuum companion to systemic functional linguistics*. Continuum, London
- Hellermann, J & Vergun, A, 2007, 'Language which is not taught: the discourse marker use of beginning adult learners of English', *Journal of Pragmatics*, vol. 39, no. 2, pp. 157–179
- Higgs, T & Clifford, R, 1982, 'The push towards communication' in TV Higgs (ed.), *Curriculum, competence, and the foreign language teacher*, National Textbook Company, Lincolnwood, IL, pp. 57–79
- Hymes, D, 1974, 'Ways of speaking', in R Bauman & J Sherzer (eds), *Explorations in the ethnography of speaking*. Cambridge: Cambridge University Press, pp. 433–452
- Hymes, D, 1972, 'On communicative competence' in J Gumperz & D Hymes (eds), *Directions in sociolinguistics*, New York: Holt, Reinhart & Winston, pp. 35–71
- Iwashita, N, Brown, A, McNamara, T & O'Hagan, S, 2008, 'Assessed levels of second language speaking proficiency: How distinct?' *Applied Linguistics*, vol. 29, no. 1, pp. 24–29
- Jacoby, S & Ochs, E, 1995, 'Co-construction: An introduction', *Research on language and social interaction*, vol. 28, pp. 171–183
- Kang, JY, 2005, 'Written narratives as an index of L2 competence in Korean EFL learners', *Journal of Second Language Writing*, vol. 14, pp. 259–279
- Kormos, J, 2011, 'Task complexity and linguistic and discourse features of narrative writing performance', *Journal of Second Language Writing*, vol. 20, pp. 148–161
- Laufer, B & Nation, P, 1995, 'Vocabulary size and use: Lexical richness in L2 written production', *Applied Linguistics*, vol. 16, no. 3, pp. 307–322
- Lee, I, 2002, 'Teaching coherence to ESL students: a classroom inquiry', *Journal of Second Language Writing*, vol. 11, no. 2, pp. 135–159
- McNamara, T, 1996, *Measuring Second Language Performance*, Longman, London
- Martin, J, 1992, *English Text: system and structure*. Amsterdam: Benjamins
- May, L, 2009, 'Co-constructed interaction in a paired speaking test: The rater's perspective', *Language Testing*, vol. 26, pp. 397–422
- 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), *Studies in language testing, 19: IELTS collected papers, 2: Research in speaking and writing assessment*, Cambridge University Press, Cambridge, pp. 250–315
- Mochizuki, N & Ortega, L, 2008, 'Balancing communication and grammar in beginning-level foreign language classrooms: A study of guided planning and relativisation', *Language Teaching Research*, vol. 12, no. 1, pp. 11–37
- Ortega, L, 2005, 'Planning and focus on form in L2 oral performance', *Studies in Second Language Acquisition*, vol. 21, no. 1, pp. 109–148
- Paltridge, B, 2000, *Making sense of discourse analysis*, Gerd Stabler, Gold Coast, QLD
- Plonsky, L, Egbert, J & Laflair, GT, in press, 'Bootstrapping in applied linguistics: Assessing its potential using shared data', *Applied Linguistics*
- Purpura, J, 2008, 'Assessing communicative language ability: Models and their components' in E Shonany & N Hornberger (eds), *Encyclopedia of language and education*, (2nd edn), vol. 7, pp. 5–68
- Roach, P, 2000, *English phonetics and phonology: A practical course*, Cambridge University Press, Cambridge
- Sangarun, J, 2005, 'The effects of focusing on meaning and form in strategic planning', in R Ellis (ed.), *Planning and task-performance in a second language*, John Benjamins, Amsterdam, pp. 111–141
- Schiffrin, D, 1987, 'Discovering the context of an utterance', *Linguistics*, vol. 25, no. 1, pp. 11–32
- Skehan, P, 1998, *A cognitive approach to language learning*, Oxford University Press, Oxford
- Skehan, P & Foster, P, 2005, 'Strategic and online planning: the influence of surprise information and task time on second language performance', in R Ellis (ed.), *Planning and task-performance in a second language*, John Benjamins, Amsterdam, pp. 193–218
- Tajima, M, 2003, 'The effects of planning on oral performance of Japanese as a foreign language', unpublished dissertation, Purdue University
- Tyler, A, 1992, 'Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse', *TESOL Quarterly*, vol. 26, no. 4, pp. 713–729
- van Lier, L, 1989, 'Reeling, writhing, drawling, stretching and fainting in coils: Oral proficiency interviews as conversations', *TESOL Quarterly*, vol. 23, no. 3, pp. 489–508
- Widdowson, HG, 1978, *Teaching language as a communication*, Oxford University Press, Oxford
- Williams, J, 1992, 'Planning, discourse marking, and the comprehensibility of international teaching assistants', *TESOL Quarterly*, vol. 26, no. 4, pp. 693–711
- Young, R, 2011, 'Interactional competence in language learning, teaching, and testing' in E Hinkel (ed.), *Handbook of research in language learning and teaching*, Routledge, New York, pp. 426–443
- Yuan, F & Ellis, R, 2003, 'The effects of pre-task and online planning on fluency, complexity and accuracy in L2 monologic oral production', *Applied Linguistics*, vol. 24, no. 1, pp. 1–27

APPENDIX 1: TEST-TAKERS' L1 AND LEVEL

	Level 5	Level 6	Level 7
L1	N=18 (F = 5, M = 13)	N=20 (F = 11, M = 9)	N=20 (F = 6, M = 14)
Albanian	1		
Arabic	1	2	2
Chinese	3	4	1
English	1		
Farsi	1	1	1
French		1	
Greek	2	1	
Gujarati			1
Hindi			1
Indonesian		2	
Japanese	1		
Kannada			1
Korean		2	
Norwegian			2
other			1
Pashtu			1
Portuguese			1
Punjabi		1	1
Tagalog	1	2	3
Thai	3		1
Urdu		3	3
Vietnamese	4	1	

APPENDIX 2: DETAILED RESULTS OF LEXICAL RICHNESS

	Level	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Token	5	580.22	230.740	54.386	465.48	694.97	258	942
	6	737.10	283.775	63.454	604.29	869.91	205	1293
	7	887.45	392.686	87.807	703.67	1071.23	319	1552
Type	5	145.17	37.296	8.791	126.62	163.71	100	231
	6	171.20	45.331	10.136	149.98	192.42	66	253
	7	226.95	70.274	15.714	194.06	259.84	114	329
K1	5	86.02	3.700	.872	84.18	87.86	78	91
	6	82.39	5.526	1.236	79.80	84.98	70	90
	7	86.65	3.144	.703	85.18	88.12	81	91
K2	5	2.81	1.560	.368	2.03	3.59	1	7
	6	2.42	1.152	.258	1.88	2.96	1	6
	7	2.56	.870	.195	2.15	2.97	1	4
AWL	5	2.20	1.176	.277	1.61	2.78		4
	6	2.60	1.331	.298	1.98	3.23		6
	7	2.56	1.160	.259	2.02	3.11	1	5
Off-List	5	9.00	4.504	1.062	6.76	11.24	2	18
	6	12.59	5.556	1.242	9.99	15.19	4	23
	7	8.24	3.443	.770	6.63	9.85	3	14
Type-token ratio	5	.271	.068	.016	.237	.305	.180	.420
	6	.248	.054	.012	.223	.273	.190	.400
	7	.275	.056	.012	.248	.301	.200	.410
Token per type	5	3.914	0.955	0.225	3.439	4.389	2.370	5.710
	6	4.179	0.779	0.174	3.814	4.543	2.520	5.310
	7	3.767	0.690	0.154	3.444	4.089	2.440	4.900
Lexical density	5	.635	.059	.014	.605	.665	.52	.76
	6	.646	.061	.014	.617	.675	.55	.77
	7	.586	.054	.012	.560	.611	.52	.72