IELTS Research Reports
Online Series

Investigating the development of 'grammatical range and accuracy' at different proficiency levels in the IELTS Speaking test

Hanne Roothooft and Ruth Breeze

IELTS™

●● BRITISH
●● COUNCIL

✕ idp

Cambridge Assessment
English

Investigating the development of 'grammatical range and accuracy' at different proficiency levels in the IELTS Speaking test

This project sheds light on the morphemes and grammatical structures used at different band levels, as well as on the error rates and types associated with their use. The authors provide information relevant for examiner and teacher training, and make suggestions for improving rating scales and exam tasks.

Publishing details

How to cite this article

Roothooft, H., and Breeze, R. 2019. IELTS: Investigating the development of 'grammatical range and accuracy' at different proficiency levels in the IELTS Speaking test. *IELTS Research Reports Online Series, No. 1*. British Council, Cambridge Assessment English and IDP: IELTS Australia.
Available at https://www.ielts.org/teaching-and-research/research-reports

Introduction

This study by Hanne Roothooft and Ruth Breeze 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, with over 110 empirical studies receiving 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.

The study detailed in this report concerns the skill of speaking; in particular, it seeks to establish whether it is possible to identify grammatical structures that distinguish different band levels from band 4 to band 8. Following on from earlier morpheme-order studies (e.g. Dulay & Burt, 1973; Krashen, 1977; Pica, 1983) and theories of learnability and processability (e.g. Pienemann 1998), the current researchers attempted to compare findings from second language acquisition on the order of acquisition of grammatical morphemes and complex grammatical structures with actual spoken performances of candidates performing at different band levels on the IELTS Speaking test. The study will potentially be of interest to a wide range of IELTS stakeholders including teachers, coursebook writers, examination task writers and raters.

So what were the major findings of this study? First of all, analysis of seven grammatical morphemes showed that the accuracy order varied considerably across the different IELTS band levels. Secondly, and unsurprisingly, attempts at using more complex grammatical structures increased at higher band levels, which resulted in error rates also increasing at these levels. In addition, the data also provided examples of errors committed by speakers of different first languages. Finally, the study confirmed some important similarities between the IELTS Speaking test data and earlier findings from second language acquisition research.

IELTS Speaking tasks are, of course, not designed to elicit specific grammatical structures but to prompt fluent and natural use of language. Nevertheless, the findings of this research should provide useful information for test-takers preparing to take an IELTS test, as well as for other IELTS stakeholders, especially raters, as it documents the range of structures that are generally encountered at each level and also provides examples of typical errors.

Vivien Berry
Senior Researcher English Language Assessment
British Council

Investigating the development of 'grammatical range and accuracy' at different proficiency levels in the IELTS Speaking test

Abstract

This project sheds light on the morphemes and grammatical structures used at different band levels, as well as on the error rates and types associated with their use. The authors provide information that is relevant for examiner and teacher training, and make suggestions for improving rating scales and exam tasks.

A sample of 73 IELTS Speaking tests from band levels 4 to 8 was analysed in order to identify grammatical structures that distinguish different band levels and to calculate the error rates which characterise their use.

The results showed that different features followed different paths. Although some features, such as plural -*s*, were acquired early, others, like the irregular past, seemed to follow a path of gradual improvement from bands 4 to 8, while one, namely third person -*s*, was found to undergo dramatic improvement only at bands 7 and 8.

These findings bear some interesting similarities to patterns observed in second language acquisition studies.

A second focus of this study was to characterise the range of more complex structures attempted by candidates at different levels, centring on the use of conditionals, relative clauses, indirect questions and passive structures. It was found that attempts at these structures did indeed tend to increase at higher band levels, though this was often accompanied by a high failure rate.

This study contributes to our understanding of the order in which grammatical morphemes and complex structures are acquired, while at the same time providing useful information for IELTS examiners, teachers and course book writers on what structures candidates are likely to produce and what errors are typical of candidates at different levels.

Authors' biodata

Hanne Roothooft

Dr Hanne Roothooft obtained her PhD degree from the University of Navarra, Spain, in June 2014. She is currently working as an Assistant Professor at the Public University of Navarre (Spain), teaching subjects such as English for Human Resources and English Didactics for Pre-school Teachers. Her research interests include corrective feedback and teacher cognition. She has published the following article with Ruth Breeze: A comparison of EFL teachers' and students' attitudes to oral corrective feedback in *Language Awareness*, 25(4), (pp 318–335).

Ruth Breeze

Dr Ruth Breeze (MA, PhD) is Senior Lecturer in English at the University of Navarra, Spain. She has published widely in the areas of discourse studies, specialised languages and language pedagogy. She is the author of two monographs (*Corporate Discourse*, Bloomsbury Academic, 2015, and *Rethinking Academic Writing Pedagogy for the European University*, Rodopi, 2012) and has co-edited several volumes of research, including *Essential Competencies for English-medium University Teaching* (Springer, 2016). She is Co-PI of the GradUN research group on public discourse within the *Instituto Cultura y Sociedad* at the University of Navarra.

Table of contents

List of tables

List of figures

1 Introduction

The main aim of this project was to analyse a representative sample of IELTS Speaking test data at band levels 4 to 8 in order to establish whether it is possible to identify grammatical structures that distinguish different band levels. In particular, we chose to focus on the accuracy of six grammatical morphemes for which an acquisitional order has been proposed in the bibliography on this area. Apart from looking at accuracy, we also set out to include the issue of grammatical range in our analysis, by focusing on the use of simple and more complex structures by candidates at different band levels.

The identification of the grammatical structures that are typical of different proficiency levels can contribute to our knowledge of how grammatical morphemes and complex structures are acquired, while at the same time providing useful information for IELTS examiners, teachers and course book writers on what structures candidates can produce and what errors are typical of candidates at different levels.

2 Literature review

It is generally accepted that the acquisition of a second language follows a predictable route, especially for certain grammatical features. Evidence for this route of acquisition comes from the so-called morpheme-order studies (e.g. Baily, Madden and Krashen, 1974; Dulay and Burt, 1973; Pica, 1983; Rosansky, 1976), on the one hand, and from theories of learnability and processability (e.g. Pienemann, 1998). Studies on the acquisitional order of a number of morphemes were mainly carried out in the 1970s and 1980s. Such studies were mainly cross-sectional, based on data elicited from a group of learners at a specific time, rather than following the development of the same learners over time. For instance, Dulay and Burt (1973) elicited oral data from three groups of 6 to 8-year-old Spanish-speaking children acquiring English as a second language. They calculated an acquisitional order of eight English morphemes, based on the accuracy scores for each morpheme. The finding that the order of acquisition was remarkably similar for the three groups led the researchers to propose a natural or universal order of acquisition for certain morphemes. This theory was further strengthened by subsequent studies, which found very similar orders of acquisition, for instance Dulay and Burt (1974) and Baily et al. (1974). Moreover, it appeared that the learners' first language had little or no influence on this acquisitional order. For instance, Larsen-Freeman (1976) elicited data from adults with Arabic, Japanese, Persian and Spanish as their mother tongue and concluded that the first language did not have a significant effect on the order in which English morphemes were acquired.

These studies have since been criticised on methodological grounds (e.g. Cook, 1993). One criticism states that order of accuracy cannot be interpreted as order of acquisition (e.g. Long and Sato, 1984). Another questions the selection of the morphemes under study, since a feature such as articles is very different in nature from third person -*s*, for example. There are very few more recent morpheme order studies, but some of these have also challenged the 'universal' aspect of the route of acquisition by showing significant effects of the learners' first language (Luk and Shirai, 2009; Shin and Milroy, 1999). Nonetheless, Larsen-Freeman and Long (1991) argue that the fact that all these studies have yielded similar findings cannot be ignored.

Rather than testing the order of acquisition, Goldschneider and DeKeyser (2005) attempted to discover the factors that might explain the similar findings of the morpheme-order studies. Based on a meta-analysis of 12 studies, they arrived at a common order for six functors, in order of acquisition: -*ing*, plural, articles, regular past, possessive, third person singular. The authors found that a large part of the findings of these studies can be explained by a combination of five factors, such as perceptual salience and frequency. For instance, -*ing* is much more salient than third person -*s*, since it constitutes a syllable and contains a vowel, and therefore tends to be acquired sooner.

Another theory which stresses the universal nature of second language acquisition is Pienemann's (1998) processability theory. This theory was first developed based on data from naturalistic acquisition of German by Italian and Spanish adults working in Germany (e.g. Meisel, Clahsen and Pienemann, 1981). The acquisition of English as a second language was also investigated by Pienemann and his colleagues (e.g. Johnston and Pienemann, 1986). Pienemann found clear developmental patterns in these data, which led him to propose a hierarchical order of acquisition, in which learners go through six stages and the acquisition of the features of one stage imply that the features of the previous stages have been acquired. For instance, the first stage is characterised by the use of single constituents or formulaic chunks, whereas subject-verb agreement (for instance third person -*s*) only appears at stage five, and the sixth and final stage is characterised by the acquisition of subordinate clauses, such as indirect questions. Interestingly, this theory does not only include morphological features but also syntactic structures, such as yes/no inversion. Moreover, it is pointed out that not all aspects of the language follow a predictable path of acquisition, and that some features are variational, which means they are acquired at very different points by different learners.

Given the problems with previous studies on the acquisition of grammatical features, for instance the question of the influence of the first language and the dearth of more up-to-date research on the acquisition of morphemes, the analysis of a corpus of learner English which consists of different levels of proficiency may shed more light on how English as a second language is acquired.

In combination with the morpheme order studies, another framework which can help us study the grammatical development of learners is that of CAF: complexity, accuracy and fluency. Researchers working within this framework, such as Skehan (1998), propose that there may be a trade-off effect between these three dimensions of proficiency. For example, if a learner focuses on fluency, he or she may become less accurate. Similarly, an attempt to use more complex language may negatively affect accuracy. Particularly interesting for the present study is the notion of complexity, which can be defined in two different ways: difficulty and inherent complexity (Bulté and Housen, 2012). Some studies define complexity as the difficulty with which certain language features are acquired or processed. According to psycholinguistic studies (e.g. Byrnes and Sinicrope, 2008; Diessel, 2004), subordinate structures such as relative clauses or conceptually difficult structures, such as passives, are more difficult to process and appear to be acquired later than other structures. Another interpretation of complexity, however, refers to the inherent complexity of a language feature, for instance syntactic complexity, which can be measured by calculating the number of subordinate clauses versus the number of main clauses. Inherent complexity of language features may also help explain the order of acquisition found in the morpheme studies. In the case of plural -*s*, for example, there is a one-to-one mapping of meaning onto form, which makes it less complex than third person -*s*, which expresses both present tense and third person (Bulté and Housen, 2012). Indeed, we have seen that the latter morpheme was found to be acquired late, whereas the former is acquired early (Goldschneider and DeKeyser; 2005).

With regard to syntactic complexity, Pienemann's (1998) processability theory also predicts that subordinate clauses (which are more complex) are acquired at the final stage of acquisition.

In addition to the theoretical interest inherent in studies on acquisition order, this topic also has obvious practical implications. Some attempts have been made to tie specific grammar features into the Common European Framework of Reference for Languages (CEFR), for instance by the English Profile Project (e.g. Salamoura and Saville, 2009). However, so far these studies are not conclusive and they have focused only on a few aspects of grammar. Another study by Hawkins and Buttery (2010) has attempted to establish which grammatical structures are typical of each level of the CEFR, based on examples from the Cambridge Learner Corpus. Unlike the data used in the present study, this corpus only contains written data. One study which focused on spoken data from the Cambridge English exams is Kang (2013), but this was broader in focus, looking not only at grammar but also at discourse management and pronunciation.

More information on which features are attempted at different levels, as well as which errors are typically made, would not only be useful for IELTS examiners, but would also provide more information on what learners at different CEFR levels can do with regard to grammar. In order to be able to compare our findings to studies based on the CEFR, for the purposes of our present research, we assume that IELTS band 4 more or less corresponds to B1, band 5 to B1+ (between a B1 and a B2 on the scale), band 6 to B2, band 7 to C1 and band 8 to C1+ (https://www.ielts.org/ielts-for-organisations/common-european-framework).

3　Aims

The present study was designed to contribute to second language acquisition research by gathering more evidence on the order of acquisition of grammatical morphemes, on the one hand, and on the acquisition of complex syntactic structures, on the other hand, by analysing data from the IELTS speaking tests in band levels 4 to 8. At the same time, this study was intended to benefit IELTS examiners and teachers by shedding light on the range of structures which can be encountered at each band level and providing examples of typical errors. More specifically, our aims were the following:

1. To identify the grammatical structures which distinguish different band levels of the IELTS speaking test, by studying a sample of candidates' performance at band scores 4 to 8.

2. To study the degree of accuracy with which a series of grammatical morphemes is produced by learners at band scores 4–8. These morphemes have been selected on the basis of second language acquisition studies.

3. To study the use and degree of accuracy of a number of more complex structures, in order to analyse the development of grammatical range at band scores 4–8.

4. To develop detailed descriptors of the grammatical structures that candidates are expected to produce and the degree of accuracy with which they need to produce them at band levels 4 to 8, in order to help examiners rate candidates more objectively on their grammatical range and accuracy.

5. To provide an outline of what is expected in terms of grammatical range and accuracy across the IELTS spectrum in order to help teachers, course book writers and examination task writers.

3.1   Research questions

The following three research questions underlie the present study.

RQ1.   Do the accuracy scores of six grammatical morphemes in IELTS levels 4–8 reflect the order of acquisition established in second language research?

RQ2.   Which complex grammatical structures (e.g. relative clauses, passives) are used at IELTS levels 4–8 and how does their accuracy evolve?

RQ3:   Which grammatical structures distinguish IELTS band levels 4–8?

4   Method

4.1   Participants

Since we were provided with the data we requested, it was not necessary to administer the speaking test to students at our institution (see our original proposal). We were given a sample of 15 speaking tests at levels 4 to 8, which means a total of 75 samples. Most of these samples had already been transcribed, but 23 were only provided as audio files.

From January to April 2015, we completed the transcription of the 23 audio files. As the sound quality was insufficient to allow transcription of two of the samples, we worked with a total of 73 – 15 for band levels 4 to 6 and 14 in bands 7 and 8. We were given 9 samples for band 8 and 6 samples for band 8.5, but we decided to group all of them under band 8. Table 1 gives an overview of the samples per band level and the first language of the candidates. It needs to be noted that certain first languages are more frequent in specific band levels, for instance Arabic in band 4 or Tagalog in band 7, which might potentially influence our findings.

Table 1: *Overview of transcripts per band level with L1*

Band 4	Band 5	Band 6	Band 7	Band 8
1. Spanish	1. Arabic	1. Spanish	1. Tagalog	1. Marathi
2. Arabic	2. Spanish	2. Arabic	2. Gujarati	2. Arabic
3. Arabic	3. Arabic	3. Arabic	3. Tagalog	3. Kannada
4. Arabic	4. Arabic	4. Arabic	4. Farsi	4. Malayalam
5. Arabic	5. Chinese	5. Chinese	5. Thai	5. Luo
6. Arabic	6. Chinese	6. Chinese	6. Unknown	6. Tagalog
7. Arabic	7. Chinese	7. Chinese	7. Kannada	7. Tagalog
8. Arabic	8. English	8. Unknown	8. Tagalog	8. Chinese
9. Arabic	9. Thai	9. Korean	9. Tagalog	9. Thai
10. Chinese	10. Thai	10. Korean	10. Tagalog	10. Bosnian
11. Chinese	11. Thai	11. Tagalog	11. Tagalog	11. Tagalog
12. Chinese	12. Vietnamese	12. Tagalog	12. Tagalog	12. English
13. Indonesian	13. Tagalog	13. Chinese	13. Hindi	13. Chinese
14. Japanese	14. Arabic	14. Arabic	14. Tamil	14. Chinese
15. Tagalog	15. Chinese	15. Urdu		

The IELTS speaking test takes the form of an oral interview between the examiner and the candidate. It can be divided into three parts: (1) personal questions about the candidate; (2) a monologue about a topic; and (3) a longer discussion about the topic from part 2. It normally lasts for 11 to 14 minutes.

4.2 Data analysis

4.2.1 Exploratory analysis

Apart from the grammatical morphemes which we set out to investigate, we did not start with a pre-conceived list of specific grammatical structures to analyse, since we first needed to obtain a general overview of the type of features we could expect in the speech situation of the IELTS test. To do this, we first undertook an exploratory analysis of eight randomly picked band 5 samples, for which we made an inventory of the different types of errors we came across. This exploration revealed that a number of the proposed morphemes, such as third person -s or -ing, were relatively infrequent. For instance, in three of the eight above-mentioned samples not a single context for third person -s could be identified. This might be due to the format of the speaking test, which mainly contains general questions about the candidate's likes and dislikes or about people in the candidate's country. For example: "What makes you feel happy?" or "Where do people in your country like to go on holiday?". Research on task-based learning and teaching has pointed out the difficulty of designing tasks which elicit certain language structures while at the same time being sufficiently challenging and interesting from a communicative point of view (e.g. Ellis, 2003). At first glance, the IELTS interviews are not intended to elicit specific structures, but rather to promote fluent and "natural" use of language on everyday topics. Apart from rating grammatical accuracy and range, examiners also take into account fluency, pronunciation and lexical resources. For the present study, however, it needs to be kept in mind that the results are inevitably influenced by the instrument used to elicit the data.

The exploratory analysis also identified a number of other potentially interesting features to analyse, such as comparatives, relative clauses, indirect questions and passives. The last three are especially useful to investigate the issue of grammatical range. The IELTS descriptors for grammatical range and accuracy refer to "basic" or "simple" structures versus "complex" structures. The use of more complex structures presumably merits a higher score, and complex structures can be hypothesised to be more frequent at higher levels. As discussed in the literature review, subordinate structures such as relative clauses have been found to be more complex, both from a linguistic or syntactic point of view and from a developmental point of view. In Pienemann's (1998) hierarchy, indirect questions also appear at the final stage of acquisition. Passives have also been classified as complex structures which are thought to be acquired late (Diessel, 2004).

After this first exploratory analysis, we started focusing on the six grammatical morphemes discussed in the literature review. Since past tense -ed was one of them, we also decided to include the irregular past tense. Morpheme order studies generally rank the irregular past tense before the regular past tense (e.g. Krashen, 1977).

4.2.2 Obligatory context analysis

The method of analysis used to investigate the grammatical morphemes was obligatory context analysis, since the same method was used in the majority of the morpheme order studies (Goldschneider and Dekeyser, 2005). This way, our results can be more easily compared to previous research. Obligatory context analysis, first used by Brown (1973), consists of identifying all obligatory contexts for the use of a certain grammatical functor and counting the number of times in which this functor has been correctly supplied.

Accuracy rates can then be calculated by dividing the number of correctly supplied morphemes by the total number of obligatory contexts. For instance, in the following short passage from the IELTS speaking test, four obligatory contexts for third person -s are outlined in bold type. There is only one case in which third person -s has been supplied ("it makes me relax"). This would give us an accuracy score of 1/4 or 25% for this passage. Rather than calculating accuracy scores, the present study makes use of error rates. This would give us an error rate of 75% for third person -s in this example.

> =uh .h I like to travel (.) uh: I (.) I like to travel with my husband? (0.3) because **make** me relaxed .hh and er::::: it **makes** me relax and can .hh and if I want to do anything (0.6) you know what I do (0.3) .hh and if I want to:: er: to other (0.2) to study other (0.3) er:: language (allah) he hel- he **help** us how to can (0.2) es: er:: how to:: (0.5) er studied hh and er:: (0.2) how to (0.5) make er friends with other: (.) people he **know** everythings what I I do (band 5)

A criticism of obligatory context analysis involves the fact that this method does not take into account those cases in which learners supply a form in a context where it is not obligatory. This is also known as overuse, and it is quite common at earlier stages of acquisition. It has been found, for instance, that when acquiring the past tense learners might use a different morpheme, such as -ing, in a past tense context (Lightbown and Spada, 2013).

In order to remedy this problem, Pica (1984) proposed the following formula to analyse the use of morphemes:

n correct suppliance/(n obligatory context + n non-obligatory context) X 100

However, Pica's (1984) study showed that, even when overuse was taken into account, the order of acquisition was largely the same as in previous morpheme order studies. In the present study, we therefore decided not to include non-obligatory contexts in our calculation, but we did keep a separate record of cases of overuse.

4.2.3   Statistical analysis of the data

For each grammatical item, error rates were calculated per interview and the total error rate per band level was reported. In order to determine whether there were significant differences in error rates between band levels, chi-square tests of independence were then calculated for each grammatical feature, using the free online software vassarstats (www.vassarstats.net). The chi-square tests were performed comparing the total number of errors in each band to the total number of correct usages.

4.2.4   Problems with error analysis

Since the identification of errors and obligatory contexts is a far from straightforward process, it was necessary to establish a list of guidelines for the analysis of each grammatical morpheme or structure.

First of all, there is the problem of formulaic sequences or chunks. Research on chunks tells us that they are stored as lexical items in a particular learner's mind and are often more grammatically complex than other language produced by that same learner (Schmitt, 2004). This would mean that they do not reflect the learner's real grammatical competence and therefore we decided to exclude the most obvious formulaic sequences from our calculations. For example, a chunk such as "lots of" would not be counted as an obligatory context for plural -s, in the same way that "a little bit" would not be included as an obligatory context for the indefinite article.

Second, learner language is sometimes difficult to interpret. If the meaning of a certain passage was not clear enough, it was decided to exclude it completely from the analysis. For example:

> Now the neighbours are very careful when downtown know way people is in the neighbour (band 4).

When presenting the results, the specific criteria for analysing each structure or morpheme will be discussed.

4.2.5   Analysis of the data by two raters

Due to the above-mentioned difficulties with coding the data, data analysis of all structures was carried out by both researchers separately and then compared and discussed. Although this procedure is time-consuming, it was felt to be necessary after analysing a sample of band 5 interviews for articles and noticing some discrepancies between the two raters. Finally, it was decided that the raters should work independently and then compare their results to achieve consensus scores.

5   Results

5.1   Morphemes

As mentioned in the Introduction, we proposed to analyse the use of six grammatical morphemes whose order of acquisition has been determined by a substantial number of second language acquisition studies. According to Goldscheider and DeKeyser's (2001) meta-analysis of these studies, these morphemes are acquired in the following order: *-ing*, plural *-s*; articles, *-ed*, possessive *-s*, third person singular *-s*. Since articles have traditionally been included in morpheme order studies, even though they are not technically morphemes, we decided to include them in our study. Moreover, we also decided to add the irregular past, which is generally thought to be acquired before the regular past. When discussing our results, we will thus compare them to the following order of acquisition:

1.   Present progressive (*-ing*)
2.   Plural *-s*
3.   Articles
4.   Irregular past
5.   Regular past
6.   Possessive *-s*
7.   Third person singular *-s.*

5.1.1   Present progressive

Obligatory context analysis was used to calculate error rates for the present progressive *-ing* in all the speaking tests for a given band level. In Table 2, the left-hand column displays the number of errors and the number of obligatory contexts (OCs), while the right-hand column shows the error rates, expressed in percentages. The total error rate for each band level could then be calculated. As mentioned in Section 4.2, certain features are more frequently used than others and it is possible not to encounter a single context for some of the features. These cases have been indicated as "NC" for "no context" in all tables.

Errors against the present progressive usually involve incorrect formation, as in the following examples:

I er I living in X city in X province (band 4)

I am study to be a teacher (band 5)

Like er news what going around the world (band 7).

Table 2: *Error rates for the present progressive*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	NC		0/2	0	NC		1/3	33.3	0/6	0
2	NC		0/2	0	1/3	33.3	1/10	10	0/2	0
3	0/1	0	NC		0/1	0	0/4	0	0/7	0
4	6/6	100	1/1	100	1/1	100	0/3	0	NC	
5	4/5	80	NC		1/1	100	0/2	0	0/1	0
6	NC		1/4	25	NC		0/18	0	0/3	0
7	NC		NC		1/1	100	0/3	0	0/4	0
8	NC		0/2	0	2/4	50	0/12	0	0/4	0
9	0/2	0	0/2	0	1/8	12.5	0/14	0	0/1	0
10	NC		NC		0/2	0	0/2	0	NC	
11	NC		1/1	100	0/8	0	0/4	0	0/8	0
12	1/1	100	1/2	50	0/16	0	0/1	0	0/3	0
13	0/2	0	0/2	0	NC		0/14	0	1/7	14.3
14	4/5	80	1/5	20	5/14	35.7	0/4	0	0/3	0
15	1/2	50	NC		1/4	25				
Total	**16/24**	**66.7%**	**5/23**	**21.7%**	**13/63**	**20.6%**	**2/94**	**2.1%**	**1/49**	**2%**

As can be noted in Table 2, the error rate is considerably higher at bands 4, 5 and 6 than at bands 7 and 8. A chi-square test revealed significant differences ($p<0.01$) between bands 4 and 5 (chi-square = 9.5911), and between bands 6 and 7 (chi-square = 14.9511), but not between band 5 and 6 (chi-square = 0.0124), or between band 7 and 8 (chi-square = 0.0012).

5.1.2   Plural *-s*

To calculate the error rates for plurals, irregular plurals such as "people" or "children" were counted as correct instances. However, an error such as "peoples" was coded as an error of overuse, since the morpheme *-s* is supplied in a non-obligatory context. This is why it was not counted as an error, but it was recorded as a separate category of overuse. Instances of plural *-s* which were considered to be part of chunks or formulaic sequences were not counted, for example in expressions such as "lots of" or "all kinds of".

Only cases in which we were sure the candidate had plural meaning in mind were counted, for example because they used quantifiers (a lot of, many) or numbers (two).

On the wall there's many picture and photograph (band 5)

I study three subject to be a teacher (band 6).

A typical error is not to use the plural in expressions like "one of the things", "some of the things":

> It's one of the most popular newspaper in Japan (band 4)

> Since then it is one of the leading (girl) magazine (band 7).

In some cases, because of the absence of articles or other problems, we could not be sure if plural meaning was intended, and therefore these cases were not included in the analysis:

> Many people take erm traditional erm photograph. (band 4)

> Watch English movie er read newspaper from English and er use er use the book to read English (band 4).

Table 3: *Error rates for plural -s*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	2/33	6	0/37	0	1/34	2.9	2/30	6.7	0/77	0
2	0/21	0	0/32	0	3/42	7.1	0/47	0	0/37	0
3	1/7	14.3	3/38	7.9	0/42	0	0/44	0	0/46	0
4	0/29	0	6/48	12.5	3/39	7.7	1/57	1.8	0/47	0
5	0/10	0	0/25	0	0/23	0	0/24	0	0/57	0
6	4/11	36.4	0/31	0	1/60	1.7	4/61	6.6	2/44	4.5
7	2/10	20	1/28	3.6	5/39	12.8	2/51	3.9	0/76	0
8	1/22	4.5	1/31	3.2	1/25	4	0/40	0	0/55	0
9	2/25	8	0/7	0	4/45	8.9	1/79	1.3	0/41	0
10	2/21	9.5	3/7	42.3	3/26	11.5	0/44	0	0/73	0
11	1/22	4.5	9/24	37.5	0/13	0	0/48	0	0/67	0
12	4/12	33.3	16/30	53.3	0/54	0	1/23	4.3	0/67	0
13	3/27	11.1	0/41	0	0/25	0	0/40	0	2/60	3.3
14	3/20	15	0/20	0	8/47	17	0/59	0	1/57	1.8
15	0/26	0	0/21	0	0/37	0				
Total	**25/296**	**8.4%**	**39/420**	**9.3%**	**29/551**	**5.3%**	**11/647**	**1.7%**	**5/804**	**0.6%**

Chi-square revealed that there were no significant differences at $p<0.01$ between bands 4 and 5 (chi-square = 0.0816), bands 5 and 6 (chi-square = 3.9257), or bands 7 and 8 (chi-square = 3.8221), but there was a significant difference between 4 and 7 (chi-square = 25.1698), 5 and 7 (chi square = 32.8097), and bands 6 and 7 (chi-square = 13.4777). It thus appears that there is a major dividing line in accuracy between bands 4-6 and 7-8 on this issue.

5.1.3   Articles

The following instances of article use were analysed in the transcripts:

- the indefinite article (a/an)
- the definite article (the)
- zero article.

After a first analysis of article use in band 5, the two researchers agreed on a number of guidelines on how to deal with uncertain cases or examples which were difficult to interpret.

The following criteria were established:

A. The use of articles was not counted in a number of cases:

1. If article use was unclear because of an apparent breakdown in communication, these articles were excluded, which means they were not counted. The following is an example of a passage which was judged impossible to interpret by both raters:

   The natural you can't say, the trees, mountains, no lie or not (band 5).

2. Articles which were used in the following expressions were not counted: "a lot of", "lots of", "a little bit", "for example", "sort of", "various", "a wide range", "kind of", "certain".
   In the case of "a lot of", neither the initial indefinite article "a" nor the article following the expression was included. For instance, "a lot of things" could count as two correct uses: "a" in "a lot" and zero article in "things".

   Still a lot of things that I think I can do and er:: (band 6)

3. To avoid complications, it was decided not to count gerunds as uses of the zero article, for example:

   Every once or twice a week I go to the community centre to do exercise for example swimming, I prefer swimming (band 6).

B. Apart from straightforward examples of article use, the use of articles was also counted in the following cases:

1. Article use was counted in the context of names of countries ("the USA", "India") and cities, names of languages ("English, Spanish"), days, months and years.

2. "Other" and "another" were included in the analysis.

Table 4 displays the results of the analysis of article use in all band levels. The error rates for articles for each sample are given, as well as the total error rate per band. Even though there is considerable variation within band levels, there is a decrease in error rate from band 4 to band 8, as can be expected. This decrease is clearly illustrated by Figure 1. There are different kinds of errors involving article use in the data. One of these is the absence of articles where they are required:

   I like watching movie on computer (band 4)

   If I take big picture of you maybe you smile (band 5)

   Nowadays booking is very difficult and of course there is time limit for eh like for example (band 7).

Another problem is using articles where they are not needed, for instance with names of cities or countries:

   I want to go to the South Korea (band 5).

Articles are also used wrongly with plural nouns when talking in general, or with uncountable nouns:

   I think when the some the er when the people want er too much money or er too much like them this is er bad bad will unhappiness (band 4)

   Because the parents would like the children to spend time y'know studying (band 7)

   The best way to recognise a work (band 6)

   But if you take the happiness in a positive way (band 8)

   They give you a lot of the intellectual knowledge (band 8).

Table 4: *Error rates for articles*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	17/59	28.8	10/67	14.9	10/77	13	11/84	13	5/166	3
2	8/40	20	8/48	16.7	14/104	13.5	4/107	3.8	1/100	1
3	3/23	13	22/67	32.8	6/79	7.6	3/82	3.7	6/102	5.9
4	22/63	34.9	29/105	27.6	25/116	21.6	11/139	7.9	5/100	5
5	31/80	38.8	5/55	9.1	20/78	25.6	17/133	12.8	3/124	2.4
6	18/52	34.6	12/47	25.5	19/98	19.4	16/103	15.6	11/97	11.3
7	24/66	36.4	24/97	24.7	14/75	18.7	6/134	4.5	12/137	8.8
8	24/64	37.5	3/64	4.7	14/55	25.5	8/124	6.5	7/79	8.9
9	10/44	22.7	10/76	13.2	17/96	17.7	3/75	4	4/136	2.9
10	8/43	18.6	10/46	21.7	21/120	17.5	5/106	4.7	11/174	6.3
11	18/51	35.3	22/89	24.7	13/88	14.8	10/104	9.6	7/183	3.8
12	11/38	28.9	17/65	26.2	3/64	4.7	0/81	0	4/100	4
13	12/69	17.4	9/72	12.5	30/117	25.6	8/100	8	11/111	9.9
14	28/86	32.6	20/69	29	25/119	21	13/174	7.5	1/94	1.1
15	7/55	12.7	13/53	24.5	23/91	25.3				
Total	**241/833**	**28.9 %**	**214/1020**	**21%**	**254/1377**	**18.4%**	**115/1546**	**7.4%**	**88/1703**	**5.2%**

In order to find out if this difference between band levels is statistically significant, a chi square was calculated between adjacent bands. The differences were significant at $p<0.01$ between band 4 and 5 (chi square = 15.6482), between band 6 and 7 (chi square = 80.1178), and between band 7 and 8 (chi square = 7.1103), but not between band 5 and 6.

Figure 1: *Error rates for articles per band level*
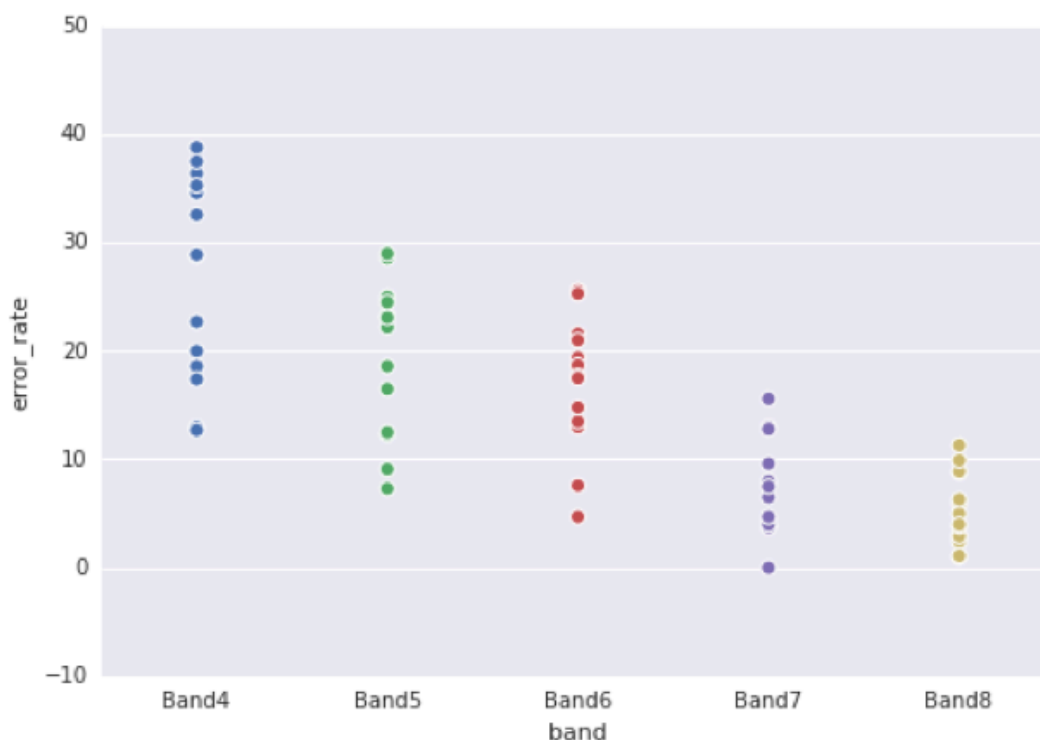
5.1.4 Irregular past

Since the regular and irregular past are thought to be acquired in different ways, they were analysed separately. Table 5 shows that the total rate decreases as the band level increases in the case of the irregular past, except from band 4 to 5, where there is a slight increase. It can be seen that there are relatively few contexts for the past tense. This is partly due to the limited number of questions that refer to past time in the exam format. If we randomly select one of the interviews, we find that it does not contain any questions referring to the past, only to the present and the future (e.g. "What do you generally use a computer for?", "Thinking about the future, do you think people in this country will travel more or less?"). Another interview only contains one question which could elicit past tense: "Are relationships different now than in the past between neighbours?".

Nonetheless, some past contexts could be identified and some examples of typical errors against the irregular past can be given. Mostly, learners simply use the base form or present tense form of the verb instead of the past tense form:

When they are young my father fell in love my mother (band 4)

Yesterday my children catch two bird in my house (band 5)

When I was er in the high school I find myself er in a biology subject (band 6).

Table 5: *Error rates for the irregular past tense*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	1/2	50	0/2	0	1/3	33.3	NC	NC	0/10	0
2	NC	NC	1/4	25	1/6	16.7	0/2	0	1/11	9
3	NC	NC	0/1	0	0/3	0	0/2	0	0/7	0
4	NC	NC	2/2	100	NC	NC	2/19	10.5	0/13	0
5	NC	NC	NC	NC	0/2	0	0/4	0	0/6	0
6	NC	NC	3/9	33.3	2/3	66.7	0/3	0	0/8	0
7	NC	NC	0/5	0	0/1	0	0/9	0	0/4	0
8	1/1	100	NC	NC	0/4	0	0/2	0	1/3	33.3
9	1/2	50	0/3	0	0/1	0	0/4	0	0/10	0
10	1/2	50	1/1	100	0/4	0	0/3	0	0/9	0
11	3/4	75	0/1	0	0/2	0	7/16	43.8	0/32	0
12	0/2	0	1/2	50	0/1	0	0/3	0	0/9	0
13	1/1	100	1/2	50	3/7	42.9	0/4	0	0/3	0
14	0/14	0	2/3	66.7	7/28	25	NC	NC	0/4	0
15	0/1	0	0/1	0	NC	NC				
Total	**8/29**	**27.6%**	**11/36**	**30.6%**	**14/65**	**21.5%**	**9/71**	**12.7%**	**2/129**	**1.6%**

The chi-square test only detected significant differences (at $p<0.01$) between adjacent bands 7 and 8 (chi-square = 10.9063) (all the other bands were significantly different from band 8). No significant differences emerged between bands 4 and 5 (chi-square = 0.0032), bands 5 and 6 (chi-square = 0.5933), or bands 6 and 7 (chi-square = 1.8967), or between bands 4 and 6 (chi square = 0.4091) or bands 4 and 7 (chi square = 3.2441).

5.1.5   Regular past

As with the irregular past, there are only a few contexts for the regular past tense. Errors also involve the use of a present or base form of the verb when referring to past time:

> When I was most, more younger also them, we go outside and play together in the big garden (band 5)

> because in the past they travel less and now they travel more (band 6)

> before they tend to buy a simple rose (…) now it has to be (…) it will cost you 200 bucks (band 7).

Table 6: *Error rates for the regular past tense*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	NC	NC	4/9	44.4	0/2	0	0/1	0	0/5	0
2	NC	NC	1/1	100	0/1	0	0/3	0	0/7	0
3	NC	NC	0/1	0	0/4	0	NC	NC	0/3	0
4	NC	NC	NC	NC	1/1	100	1/6	16.7	0/2	0
5	NC	NC	0/4	0	NC	NC	0/1	0	NC	NC
6	NC	NC	NC	NC	0/5	0	0/3	0	NC	NC
7	NC	NC	0/3	0	0/1	0	0/9	0	0/3	0
8	0/1	0	0/1	0	0/3	0	1/5	20	NC	NC
9	0/2	0	0/3	0	0/5	0	0/3	0	0/7	0
10	0/1	0	0/1	0	2/6	33.3	0/1	0	0/3	0
11	0/1	0	NC	NC	1/1	100	1/7	14.3	1/14	7.1
12	NC	NC	0/1	0	NC	NC	0/4	0	0/12	0
13	1/1	100	0/2	0	1/5	20	NC	NC	0/1	0
14	1/2	50	0/1	0	6/11	54.5	0/4	0	0/4	0
15	0/1	0	NC	NC	0/2	0				
Total	**2/9**	**22.2%**	**5/27**	**18.5%**	**11/46**	**23.9%**	**3/47**	**6.4%**	**1/61**	**1.6%**

No significant differences were found between bands 4 and 5 (chi-square = 0.0591), 5 and 6 (chi-square = 0.2415), 6 and 7 (chi-square = 5.3714) and 7 and 8 (chi-square = 1.6749) (p<0.01). In fact, the only significant difference at p<0.01 were between bands 4 and 8 (chi square = 8.1), 5 and 8 (chi square = 8.3929) and 6 and 8 (chi square = 13.0663).

5.1.6   Possessive *-s*

As can be seen in Table 7, possessive *-s* was by far the most rarely used morpheme in our data.

An example of an error against the possessive is the following:

> One of my uncle friends works in nasa (band 7).

Given the very low number of contexts, it is difficult to make any generalisations about the acquisition of this morpheme based on these data.

Table 7: *Error rates for genitive –s*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	1/1	100	NC		NC		0/2	0	NC	
2	NC		NC		NC		NC		NC	
3	0/1	0	0/1	0	0/2	0	0/2	0	NC	
4	0/2	0	NC		NC		1/3	33.3	0/3	0
5	NC		NC		NC		0/2	0	0/1	0
6	NC		NC		0/3	0	0/1	0	NC	
7	NC		0/1	0	NC		NC		NC	
8	1/1	100	NC		NC		0/1	0	NC	
9	NC		NC		NC		NC		0/2	0
10	NC		NC		NC		NC		0/1	0
11	0/1	0	NC		NC		NC		0/1	0
12	NC		0/1	0	0/1	0	NC		0/4	0
13	NC		NC		NC		0/2	0	NC	
14	NC		NC		NC		0/7	0	0/2	0
15	NC		NC		NC					
Total	**2/6**	**33.3%**	**0/3**	**0%**	**0/6**	**0%**	**1/20**	**5%**	**0/14**	**0%**

Chi-square test showed no significant differences between adjacent bands at p<0.01.

5.1.7 Third person -s

All omissions of the third person -s in affirmative sentences were counted as errors, for instance:

Because if he study well he will get more er a degree (band 4)

However, it was decided not to include mistakes against the verb "have", since it is irregular, and is involved as an auxiliary in various constructions. Errors against the third person occurring in negatives, as in the following example, are not included in this category either.

I want a little Australian parrot. Is blue and white and don't speak (band 4).

When performing the obligatory context analysis, we found only a few contexts for the use of the third person -s. In some of the samples, there was not a single context which required third person -s. As with the past tense, this is related to the fact that almost all the questions in the IELTS interview ask about the candidate's own opinion or experience (e.g. "Would you prefer to buy a picture postcard or take a photo of a new place?") or about people and things in general (e.g. "How do people in your country generally feel about birds?"). We could only find two question which could elicit the use of the third person in all of the interviews: "Describe one of your neighbours" and "Describe a building that you like".

As percentages can be misleading when there are relatively few instances, Table 8 also provides the absolute number of errors in relation to the total number of contexts. In band 4 transcript 1, for example, we identified two contexts in which third person -s needs to be used, but for both of them the morpheme in question was absent. This gives us a total of 2 out of 2 errors, or an error rate of 100 per cent (see Table 8).

Table 8: *Error rates for third person –s*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	2/2	100	NC	NC	2/4	50	0/1	0	0/9	0
2	NC	NC	1/6	16.7	8/9	88.9	0/4	0	0/7	0
3	1/1	100	3/3	100	1/3	33.3	1/7	14.3	1/8	12.5
4	NC	NC	6/9	66.7	3/3	100	1/4	25	1/9	11.1
5	1/1	100	4/5	80	0/3	0	1/5	20	0/8	0
6	8/8	100	2/2	100	4/5	80	0/8	0	0/9	0
7	1/1	100	2/3	66.7	1/3	33.3	0/12	0	0/5	0
8	4/5	80	NC	NC	1/4	25	3/16	18.8	0/7	0
9	7/7	100	2/3	66.7	NC	NC	1/3	33.3	1/7	14.3
10	2/3	66.7	2/2	100	2/2	100	0/1	0	0/14	0
11	2/2	100	5/5	100	2/3	66.7	0/7	0	1/10	10
12	2/2	100	2/3	66.7	3/3	100	0/7	0	1/7	14.3
13	4/4	100	1/8	12.5	2/2	100	0/9	0	1/8	12.5
14	2/8	25	1/1	100	4/4	100	1/21	4.8	0/7	0
15	0/3	0	3/8	37.5	3/3	100				
Total	**36/47**	**76.6%**	**34/58**	**58.6%**	**36/51**	**70.6%**	**8/105**	**7.6%**	**6/115**	**5.2%**

Looking at the total error rates, we notice that these are very high in the first three bands, after which there is a very marked decrease in band 7 and only a slight decrease after that. The chi-square test detected no significant differences at $p<0.01$ between band 4 and 5 (chi-square = 3.7748), 5 and 6 (chi-square = 1.6915), and 7 and 8 (chi-square = 05313). However, it found a significant difference between bands 4 and 7 (chi square = 75.1033), bands 6 and 7 (chi-square = 67.2155) and 6 and 8 (chi square = 79.88659). The main watershed here thus appears to lie between bands 6 and 7.

5.1.8   Overuse errors

As mentioned in the methodology section, a criticism of obligatory context analysis is that it does not take into account errors of overuse. When going through the interview scripts, we came across some overuse errors, especially for four of the grammar features: plural -*s*, irregular past, regular past and third person -*s*. Table 9 shows the total number of overuse errors for these features in each band level. As can be seen, the numbers are not very high, although these errors are quite common in the case of the plural. Overuse errors also occur in all band levels.

Table 9: *Total number of overuse errors per band level for four of the features*

	Band 4	Band 5	Band 6	Band 7	Band 8
Plural -*s*	11	16	23	17	5
Irregular past	5	2	6	6	10
Regular past	6	3	0	11	2
Third person -*s*	2	2	6	2	3

In the case of plural -*s*, a typical problem is to overgeneralise the plural ending to nouns which have an irregular plural. This still occurs in higher levels, up to band 7:

A lot of peoples got to there because it's near a shopping supermarket (band 5)

If you travel with peoples they will er have to er to know their quality (band 6)

Somehow, peoples don't have more stories (band 7)

People said parents is a first teacher of the childrens (band 6).

This is very informative especially for the working womans like us (band 7).

Some candidates also add plural -*s* to uncountable nouns:

But it is not easy because they use slangs (band 6)

Like you have to hand in the homeworks in time (band 7)

I'm using it for encoding informations (band 7).

Finally, some candidates use the plural -*s* in a singular context. This even occurs in band 8:

It is an important things because er yeah you will know that there is someone er beside you (band 6)

Because I had very good English teachers and a very good English books (band 7)

I have a bachelor degrees in science (band 8)

I like to look at a beautiful things (band 8).

In the case of the regular and irregular past tense, they are sometimes used instead of the infinitive of the verb:

If they don't know how to talk they will face many problem. (…) They have to ask for places that they want to go, and found someone can help them. (band 4)

The newspaper (…) is very quick to, to knew something (band 6)

One of the best solutions to, to dis- disentangled or just simply release that stress (band 7)

They can forgot all their troubles (band 8).

Past tense forms can also be found in present tense contexts:

Right now? I believe erm the jobs that have erm had, that had a good pay is that the jobs are of the ban- in the banks (band 6)

Then I felt something renewed in me, when I do that (question: "describe something you do that helps you feel relaxed") (band 7)

There is some parents who really are concerned about what is being shown in the television that er they tried to control them (band 8).

Finally, there are a few instances of candidates overusing third person -*s*, for example with other persons:

Because er I likes business I like computer (band 4)

They usually treats persons or people (band 8).

Third person -*s* can also be added to infinitives or to modal verbs:

Because er em travel, travel make anybody relies on er happiness (band 4)

I like to have the two floors so can moves a lot (band 5)

It cans I think erm help you to increase your confidence (band 6).

### 5.1.9   Accuracy order for the seven morphemes

If we order the seven morphemes discussed in this section according to their error rates, we can obtain an accuracy order for each band level. These orders are displayed in Table 10.

Table 10: *Accuracy orders for the seven morphemes in the different band levels*

Band 4	Band 5	Band 6	Band 7	Band 8
Plural -*s* (8.4%)	Possessive (0%)	Possessive (0%)	Plural -*s* (1.7%)	Possessive (0%)
Regular past (22.2%)	Plural -*s* (9.3%)	Plural -*s* (5.3%)	-*ing* (2.1%)	Plural -*s* (0.6%)
Irregular past (27.6%)	Regular past (18.5%)	Articles (18.4%)	Possessive (5%)	Regular past/Irregular past (1.6%)
Articles (28.9%)	Articles (21%)	-*ing* (20.6%)	Regular past (6.4%)	Regular past/Irregular past (1.6%)
Possessive (33.3%)	-*ing* (21.7%)	Irregular past (21.5%)	Articles (7.4%)	-*ing* (2%)
-*ing* (66.7%)	Irregular past (30.6%)	Regular past (23.9%)	3rd person -*s* (7.6%)	Articles/3rd person -*s* (5.2%)
3rd person -*s* (76.6%)	3rd person -*s* (58.6%)	3rd person -*s* (70.6%)	Irregular past (12.7%)	Articles/3rd person -*s* (5.2%)

As Table 10 shows, the accuracy orders differ for each band level. For instance, regular past comes before irregular past in all bands except in band 6 and band 8 (where the same error rate was obtained for both regular and irregular past). It also needs to be noted that the rates for the possessive are based on very limited data. It is interesting, however, that third person -*s* occupies the final position in nearly all bands. If we look back at the accuracy order established in previous studies (see 5.1), we note that third person -*s* is acquired the latest. The present data would confirm the difficulty of acquiring this feature. Plurals, on the other hand, tend to be acquired early and they indeed occupy the first or second position in all bands. A striking difference with previous studies is that -*ing* or the present progressive had a high error rate, especially in the lowest three levels, whereas it is said to be among the first morphemes to be acquired. Finally, articles also appeared to pose more difficulties in the present study, while they are the third morpheme to be acquired according to the morpheme-order studies. In band 8, although error rates are low for all features, articles and third person -*s* together were the features which induced the most errors of all seven morphemes. Figure 2 provides an overview of the total accuracy rates of all morphemes.

Figure 2: *Overview of total error rates for all morphemes*
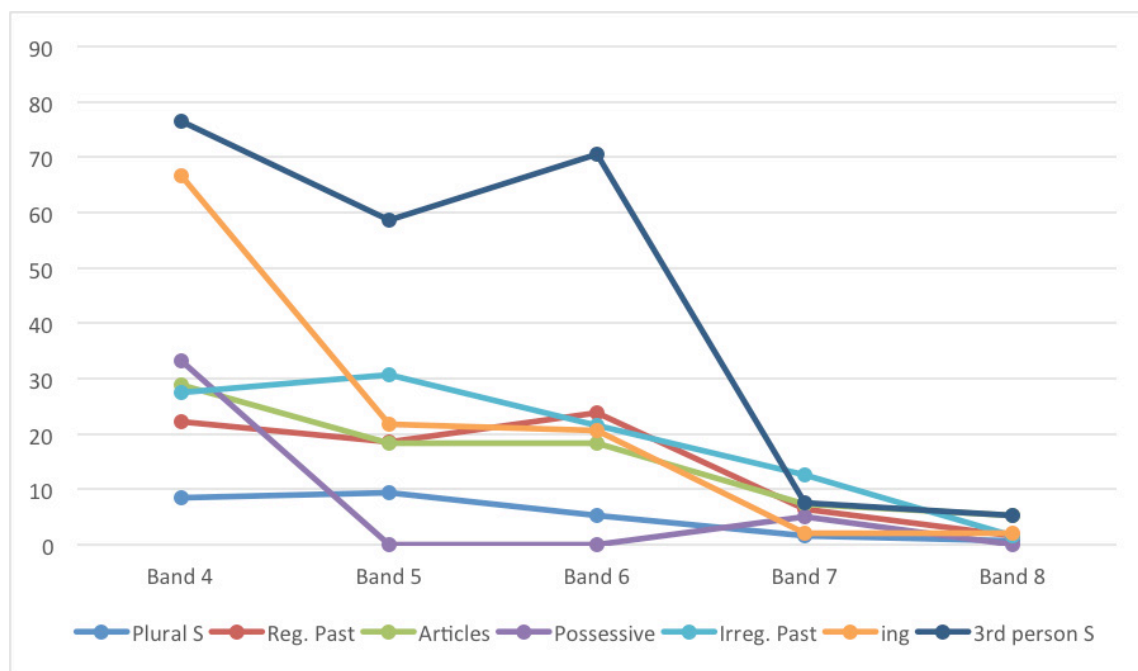
5.2   Analysis of comparatives and superlatives

Our exploration of the data revealed that some candidates made errors against comparatives and superlatives, so we decided to investigate this further. As Table 11 shows, several candidates did not use a single comparative or superlative form, while others used only a few. Most candidates who used these items did not make any errors. It needs to be noted that we counted all contexts, which means that each token is counted, even if it is repeated several times. As can be seen in Table 11, the first candidate in band 6 produced six correct instances of comparatives or superlatives. However, four of these were "the best" and two others "the most important". This could mean the candidate knows how to use the superlative of these two frequent adjectives, but we cannot say that he or she has totally acquired comparatives and superlatives. For those candidates who made errors, typical errors are:

I think like if the people were more cosier now (band 5)

And I think it's more nice to travel with (band 7)

And we live in the same locality so that makes it even more good (band 8).

Table 11: *Error rates for comparatives/superlatives*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	NC	NC	0/4	0	0/6	0	1/4	25	0/8	0
2	1/2	50	1/3	33.3	NC	NC	0/4	0	0/1	0
3	0/1	0	NC	NC	0/1	0	0/5	0	1/1	100
4	0/2	0	0/3	0	0/1	0	1/4	25	0/3	0
5	NC	NC	0/1	0	0/1	0	0/4	0	0/4	0
6	NC	NC	0/3	0	0/10	0	0/6	0	0/2	0
7	0/1	0	1/1	100	0/10	0	0/6	0	0/4	0
8	0/1	0	0/3	0	1/5	20	0/6	0	1/8	12.5
9	NC	NC	1/9	22.2	1/3	33.3	0/7	0	0/7	0
10	0/2	0	0/1	0	0/1	0	0/5	0	1/9	11.1
11	NC	NC	1/4	25	0/1	0	0/3	0	1/5	20
12	NC	NC	0/2	0	NC	NC	0/1	0	0/4	0
13	NC	NC	0/1	0	0/8	0	0/12	0	1/11	9
14	0/2	0	0/3	0	0/2	0	0/7	0	0/3	0
15	NC	NC	0/11	0	NC	NC				
Total	**1/11**	**9%**	**4/49**	**8.2%**	**2/49**	**4.1%**	**2/74**	**2.7%**	**5/70**	**7.1%**

Interestingly, the error rate does not decrease from band 4 to 8, but it increases in band 8. A chi-square test revealed no significant differences (p<0.01) between the error rates from one band to the next.

5.3   Grammatical range and complexity

5.3.1   Conditionals

Conditionals are a type of subordinate structure which can be classified as more complex (see the Literature Review) and therefore, we hypothesised that they would be used more frequently and more accurately in higher band levels. Moreover, second and third conditionals can be considered more complex than first or zero conditionals. We completed the analysis of conditional structures for all band levels. Table 12 provides error rates for all types of conditionals in the different samples. As can be seen in Table 12, few errors were made against conditionals. We therefore thought it would be more informative to provide the total number of conditionals used in each band, together with a breakdown into first and second conditional structures. There were no third conditional structures in the data.

First of all, we decided that a conditional structure is correctly used if the use of the tenses is correct. This means there may be other types of errors in the sentence, such as third person *-s* or article use. Both these examples were considered to reflect accurate usage of the first conditional:

If someone else come here too, she will post the same one (band 5)

If I want to book ticket or if I want to buy clothes (…) it's a useful tool (band 6).

If we look at the total number of conditionals used per band level in Figure 4, we see that the number increases from band 4 to 6, but it drops again in bands 7 and 8. Our hypothesis that an increase in level would mean an increase in attempted conditional structures is thus not confirmed by the data. On the other hand, if second conditional structures are considered more complex than first conditional structures, Table 12 indicates that the most second conditional structures are used in band 7. At the same time, there is an error in each of these, while two candidates in band 4 correctly used a second conditional, for example:

If I was businessman I would prefer reading newspaper (band 4).

Moreover, there were only two second conditionals in band 8 and one of these was incorrect:

If your teacher could actually, you know, should give you a little attention (…) I think that'll definitely uplift the er child (band 8).

While most of the errors made against conditionals are second conditional errors, there were also some wrongly used first conditional structures, for example: (a) in which a past tense verb is used even though the context is clearly a first conditional; and (b) in which the future tense is used in the if-clause:

(a) and if if I felt relaxed then I have to swim in in the beach (band 7)

(b) It will be very successful if I will study here in England a master (band 6).

Table 12: *Conditional structures in all band levels*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs		No. of errors/ No. of OCs		No. of errors/ No. of OCs		No. of errors/ No. of OCs		No. of errors/ No. of OCs	
1	NC		0/4		0/3		0/5		0/8	
2	0/2		1/2		0/2		0/6		0/3	
3	0/3		1/5		1/4		0/1		0/3	
4	0/1		1/5		0/10		2/4		1/2	
5	NC		0/4		4/9		0/2		0/2	
6	0/5		0/1		0/7		0/2		NC	
7	0/1		1/9		0/1		0/2		0/6	
8	NC		0/2		1/6		1/2		0/1	
9	0/2		NC		2/5		1/1		NC	
10	0/3		0/3		0/3		1/1		0/2	
11	NC		1/8		0/2		2/2		0/2	
12	0/6		0/1		0/3		1/1		0/5	
13	0/6		0/4		0/5		0/5		0/1	
14	0/6		0/1		0/6		2/2		0/2	
15	0/1		NC		0/9					
Total no. of first conditional errors	0/34		2/45		4/70		2/28		0/36	
Total no. of second conditional errors	0/2		3/4		4/5		8/8		1/2	
Total no. of conditional errors	0/36	0%	5/49	10.2%	8/75	10.7%	10/36	27.8%	1/37	2.7%

As can be seen from the error rates for conditionals in Table 12 (including all types of conditionals), the highest rate occurs in band 7. If we look at Table 12, we can see that this is mainly due to 8 errors against the second conditional, which is more complex than the first conditional. A significant difference was only detected using the Chi-square test ($p<0.01$) between bands 7 and 8 (chi-square = 8.6915). However, this result needs to be considered in the light of the surprisingly high error rate in band 7, where the rate peaks.

Figure 3: *Total number of attempts at conditionals in each band level*
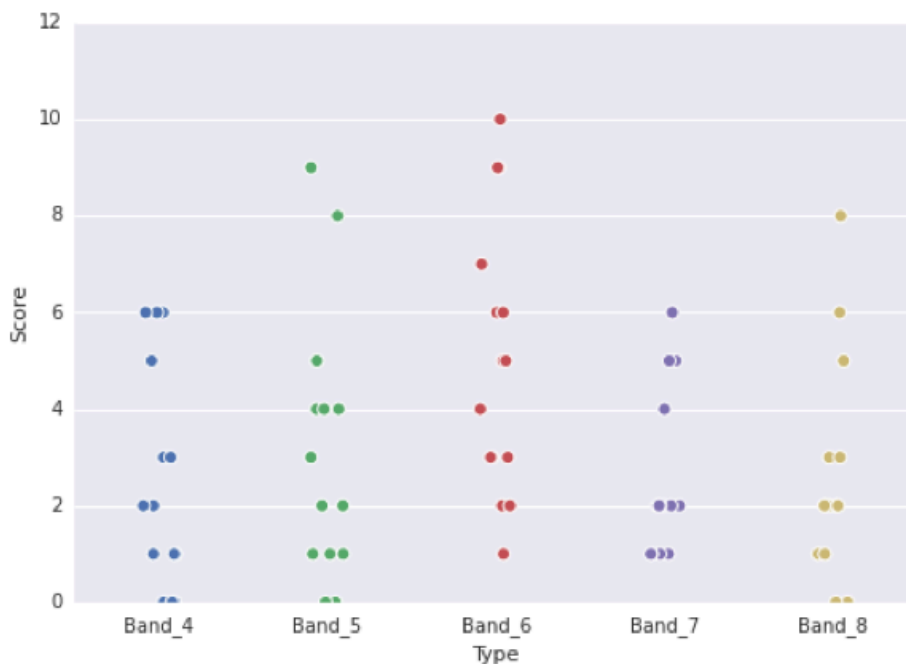
5.3.2 Relative clauses

In order to check whether subordinate syntactic structures such as relative clauses are more frequent in higher levels, we started looking at a higher level, band 7, and a low level, band 4. Table 13 shows that relative clauses are rare in band 4, with most candidates using just one or two, while the candidates in band 7 tended to use them much more frequently, between 4 and 27 times. In general, not many errors were made against this structure. Two types of errors were identified: using the wrong relative pronoun (see example a) and adding a superfluous pronoun (see example b).

(a) I can do anything what I wanna do (band 7)

(b) I'd like to talk about my neighbour who she is she is my middle school teacher (band 4).

Another error, although less frequent, is the lack of any relative pronoun:

If you have enough money you can, you can, you can do you want (band 4).

Not only does the error rate against relative clauses decrease (with a slight increase from band 7 to 8), there is also a marked difference between the number of relative clauses attempted in band 4 and those in band 8. Without taking into consideration the errors, band 4 only contains 19 relative clauses, band 5 already contains 61, band 6 has a total of 86 relative clauses, and finally bands 7 and 8 contain more than twice this number of relative clauses (185 and 189, respectively).

Table 13: *Error rates for relative clauses*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	0/1	0	0/1	0	0/15	0	0/6	0	0/13	0
2	0/1	0	0/2	0	1/6	16.7	0/4	0	0/15	0
3	0/1	0	2/9	22.2	1/2	50	0/4	0	0/10	0
4	2/3	67	2/2	100	0/5	0	0/7	0	0/14	0
5	NC	NC	0/2	0	0/3	0	0/8	0	1/20	5
6	0/1	0	0/4	0	0/6	0	1/11	9	0/10	0
7	NC	NC	0/7	0	0/5	0	0/27	0	1/21	4.8
8	0/1	0	0/9	0	0/6	0	0/20	0	1/4	25
9	1/3	33.3	3/6	50	0/4	0	0/19	0	1/20	5
10	1/1	100	0/6	0	0/2	0	2/19	10.5	0/13	0
11	1/1	100	0/2	0	1/9	11.1	1/8	12.5	0/23	0
12	1/1	100	3/3	100	0/15	0	0/13	0	1/14	7.1
13	NC	NC	0/5	0	1/5	20	0/23	0	1/6	16.7
14	0/3	0	0/1	0	0/1	0	0/16	0	0/6	0
15	0/2	0	1/2	50	0/2	0				
Total	**6/19**	**31.6%**	**11/61**	**18%**	**4/86**	**4.7%**	**4/185**	**2.2%**	**6/189**	**3.2%**

The chi-square test showed significant differences at p<0.01 between bands 5 and 6 (chi-square = 6.9743), but not between other contiguous bands. Differences were significant between bands 4 and 6 (chi square = 13.0954), and between bands 5 and 7 (chi square = 20.18). The main dividing line thus seems here to be between bands 5 and 6.

5.3.3 Indirect questions

Indirect questions were considered to be correctly used when the word order was correct and no auxiliary was included. Examples of correct indirect questions are:

I don't know what the young people think about this (band 6)

Well I don't know what actual meaning what my parents thought about (band 7).

Indirect questions which include auxiliaries or for which the word order is incorrect were counted as errors, for example:

I don't know what is these people (band 4)

I like to connect with my neighbour and er to know where is my neighbour country (band 4)

I couldn't remember what is the name of this restaurant (band 5)

So I was able to see what does the, what does a modern building offers (band 7).

Compared to relative clauses, the number of indirect questions in our data is limited. Attempts at indirect questions increase, with only 4 occurrences at band 4, 10 at bands 5 and 6, 12 at band 7 and finally 28 at band 8 (see table 14). The only significant difference between continuous bands detected by the chi-square test was between bands 7 and 8 (chi-square = 7.3557), although the number of occurrences makes statistical calculations unreliable.

Table 14: *Error rates for indirect questions*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	2/2	100	NC		NC		NC		0/1	0
2	NC		NC		NC		NC		NC	
3	NC		NC		NC		0/1	0	0/4	0
4	NC		0/4	0	NC		NC		NC	
5	NC		NC		0/3	0	NC		0/1	0
6	NC		NC		NC		0/2	0	NC	
7	NC		0/1	0	0/5	0	1/4	25	NC	
8	1/1	100	NC		NC		1/2	50	0/8	0
9	NC		1/1	100	NC		NC		0/2	0
10	0/1	0	NC		NC		NC		1/4	25
11	NC		3/3	100	0/2	0	2/2	100	0/1	0
12	NC		NC		NC		NC		NC	
13	NC		NC		NC		0/1	0	0/1	0
14	NC		0/1	0	NC		NC		0/8	0
15	NC		NC		NC					
Total	**3/4**	**75%**	**4/10**	**40%**	**0/10**	**0%**	**4/12**	**33.3%**	**1/30**	**3.3%**

5.3.4 Passive

Examples of correctly used passive structures include the following:

> Practical knowledge should be given more because it's not always the rider's fault (band 7)

> some computer firms will be opened there (band 6)

> in terms of the plastic usage that is being done (…) recycling of the human waste that is generated there (band 8).

The following are examples of errors against the passive:

> When I immigrate to Canada I have applied for the (inaudible) you see, but I haven't admitted yet (band 7)

> It's a good ah, a piece to be, to to be look (band 7).

In two cases, the passive was overused, which means it was used in an active context:

> you don't know what it will be happened (band 5)

> If there is a task that I needed to be done (band 8).

It is interesting that hardly any passive structures are used in the three lowest levels, while band 7 sees a sudden increase with 75 attempts at passive structures, more than the 66 encountered at band 8. The only passive structure used in band 4 is the following:

> the future it's might be made some rule for co communication.

More attempts also mean more errors, which may explain the seemingly contradictory increase in error rate from band 4 to band 7.

Since the number of occasions for using the passive were so small in the samples from bands 4 to 6, the chi-square calculation was only performed for bands 7 and 8. The difference was not significant at $p<0.01$ (chi-square = 3-4008).

Table 15: *Error rates for the passive*

Transcript	Band 4		Band 5		Band 6		Band 7		Band 8	
	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)	No. of errors/ No. of OCs	Error rate (%)
1	NC		NC		NC		1/2	50	0/16	0
2	NC		NC		NC		1/7	14.3	NC	
3	NC		NC		0/1	0	0/3	0	NC	
4	NC		NC		NC		1/3	33.3	0/13	0
5	NC		NC		0/1	0	NC		0/8	0
6	NC		NC		NC		1/9	11.1	0/5	0
7	NC		NC		NC		0/11	0	0/7	0
8	NC		0/3	0	NC		2/6	33.3	0/1	0
9	NC		NC		NC		0/4	0	0/2	0
10	NC		NC		NC		0/11	0	0/4	0
11	NC		NC		0/5	0	0/4	0	1/10	10
12	NC		NC		NC		0/7	0	0/3	0
13	NC		0/1	0	NC		NC		NC	
14	0/1	0	NC		NC		0/7	0	NC	
15	NC		NC		NC					
Total	**0/1**	**0%**	**0/4**	**0%**	**0/7**	**0%**	**6/74**	**8.1%**	**1/69**	**1.4%**

5.4   Complex grammatical structures in bands 4 to 8

For our analysis of grammatical range we decided to focus on four grammatical structures which can be classified as more complex: conditionals, relative clauses, indirect questions and passives. Table 16 provides an overview of the mean error rates for these four structures in bands 4 to 8.

Table 16: *Overview of mean error rates for complex structures*

	Band 4	Band 5	Band 6	Band 7	Band 8
Conditionals	0%	10.2%	10.7%	27.8%	2.7%
Relative clauses	31.6%	18%	4.7%	2.2%	3.2%
Indirect questions	75%	40%	0%	33.3%	3.3%
Passives	0%	0%	0%	8.1%	1.4%

As Table 16 shows, error rates in conditionals and passives were found to be higher in band 7 than in the lower bands. However, as we have seen this is due to the fact that fewer attempts at complex structures are made at lower levels, while higher level candidates use a wider range of structures and therefore are more likely to make errors against these complex structures. For relative clauses and indirect questions, on the other hand, a more linear decrease in error rate seems to occur. It needs to be kept in mind that the number of indirect questions and passives was limited in our data, perhaps as a consequence of the task design, and therefore these results need to be interpreted with caution.

6   Discussion

6.1   Research question 1

RQ1: Do the accuracy scores of six grammatical morphemes in IELTS levels 4–8 reflect the order of acquisition established in second language research?

Our analysis of the use of seven grammatical morphemes in obligatory contexts in the different band levels of the IELTS Speaking test showed that the accuracy order varies considerably for each level. On the basis of band 7, for example, we could say that *-ing* is acquired early, whereas in band 4 this morpheme seems to cause more difficulties, as it is the morpheme with the second-highest error rate in this band. Nevertheless it is striking that third person *-s* occupies the final position in nearly all five bands (in band 7 it occupies the penultimate position), since this would confirm the order of acquisition found in previous morpheme order studies, such as Dulay and Burt (1973, 1974) and Krashen (1977).

Even though from a structural point of view, this would be seen as a simple structure, it appears that third person *-s* is complex from a developmental point of view. According to our findings, error rates against this feature are high in bands 4, 5 and 6 and only drop considerably at bands 7 and 8, where errors are less than 10%. In terms of the CEFR, this would mean that third person *-s* errors are normal at levels B1, B1+ and B2.

Another similarity between the present study and previous morpheme-order studies concerns plural *-s*. Both in our study and in previous studies, this morpheme is found to be acquired early. Nonetheless, overuse errors against the plural (errors such as "peoples" or "informations") were found to occur at all band levels.

6.2   Research question 2

RQ2: Which complex grammatical structures (e.g. relative clauses, passives) are used at IELTS levels 4–8 and how does their accuracy evolve?

As mentioned earlier, IELTS raters do not only take into account accuracy when it comes to grammar, but they should also pay attention to the range of structures used by candidates. The present study therefore also focused on the use of four structures which are structurally and/or developmentally more complex: conditionals, relatives, indirect questions and passives.

It was found that attempts at these structures tend to increase at higher band levels, for instance, only one passive structure could be identified in band 4, four in band 5, seven in band 6, to reach a number of 72 attempted passives in band 7. At the same time, this means that error rates for such structures can be higher in the top bands than in the lower bands. The highest error rate, both for conditionals and passives, occurred in band 7.

Kang (2013) also found that error rates increased for passives, conditionals and relative clauses.The latter structure did not follow a similar pattern in the present study, as error rates for relative clauses gradually decreased from band 4 to 7, although in band 8 they again increased slightly (but not significantly). Unfortunately, indirect questions were not included in Kang's (2013) study, but these are thought to be acquired late, based on Pienemann (1998). In our data, a high rate of errors against this structure indeed occur at bands 4 and 5, but strangely there are no errors at band 6, after which the rate again increases in band 7. In any case, the highest number of attempted indirect questions occurs at band 8 (a total of 28 attempts versus 10 attempts in band 6). It also needs to be kept in mind that individual variation between candidates is high, since the 10 correct indirect questions in band 6 are only spread over three candidates, whereas the others did not use the structure at all. In band 8, on the other hand, nine different candidates attempted this structure, so it is more widespread at this level.

6.3   Research question 3

RQ3: Which grammatical structures distinguish IELTS band levels 4–8?

Previous studies have tried to identify specific features which distinguish CEFR levels (e.g. Hawkins and Buttery, 2010). Based on the present study, we can make a number of generalisations about how certain grammatical patterns evolve from band 4 to band 8, and which types of errors are distinctive of a specific band level.

If we look back at Figure 2 (see Section 5.1.9), which gives an overview of the evolution of the error rates for the seven grammatical morphemes, we note that most features follow a linear pattern, with errors decreasing from the lowest to the highest level. This type of pattern is referred to as "progressive learning errors" by Hawkins and Buttery (2010). However, in the case of the regular past, the error rate first decreases but then increases again in band 6, before it finally goes down in bands 7 and 8, while there is a slight increase for the irregular past from band 4 to 5, after which error rates again decrease. Such a pattern could be explained by theories of U-shaped learning, in which a learner goes from a low error rate to a higher one before he or she gets it right, and this has indeed been found for the acquisition of the past tense. Because of overgeneralisation of the -ed ending, it is possible that more errors occur at a higher level.

Another interesting observation based on Figure 4 is that there appears to be a clear difference between band 6, on the one hand, and bands 7 and 8 on the other, for third person *-s, -ing,* regular past, articles and irregular past, in that these features are below a 10% error rate at bands 7 and 8. Morpheme order studies such as Dulay and Burt (1973) considered a feature to be acquired if its accuracy was 90% or higher, so we could say that these features are acquired at the highest bands, but not in the first three. According to the same reasoning, plural *-s* seems to be acquired already at band 4.

With regard to the features that can be said to be typical of a given band level, statistical tests were used to establish significant differences between bands for each grammatical feature. In the case of third person *-s*, there was no significant difference between bands 4, 5 and 6, but there appeared to be a cut-off point between bands 6 and 7. For articles, there were significant differences between bands 4 and 5, 6 and 7, and 7 and 8, and at band 7 the error rate decreased to less than 10%. Thus, we could further specify the descriptors for these bands as follows:

- bands 4, 5, 6: candidates can be expected to make a large number of errors against the use of articles and third person *-s*
- bands 7, 8: although some errors against articles and third person *-s* still occur, such errors are infrequent.

A similar pattern could be observed for plural errors, even though error rates were generally much lower than for the third person or articles. For plurals, at bands 4, 5 and 6 relatively few errors occur (less than 10%), while at bands 7 and 8 these are nearly non-existent.

For relative clauses, on the other hand, a significant dividing line was detected earlier, between band 5 and 6.

For other features, there does not seem to be a significant difference between specific levels, as was the case for comparatives and indirect questions.

Finally, regarding grammatical range, an important distinction needs to be made in the band levels between attempted structures which still contain errors and successfully attempted structures. For some complex structures, the error rate is higher in the highest level, because they hardly appear in low levels. For instance, band 7 can be characterised by attempts at using passive structure, with some errors, while in band 8 passives are relatively frequent and are generally produced without errors. Also, according to the present data, band 6 contains a considerable number of attempts at conditional structures, whereas band 7 is characterised by the highest number of conditional errors (mainly due to second conditional errors).

7 Conclusion

Despite the fact that the present study is only based on 73 interviews and that certain features rarely appear in the data, the present study indicates some important similarities between the patterns of grammatical range and accuracy in the IELTS Speaking test and those observed in studies of second language acquisition.

It was possible partly to confirm research on the route of acquisition, by showing that morphemes such as plural -*s* seem to be acquired early, whereas third person -*s* is acquired late. Furthermore, it was shown that subordinate structures, such as indirect questions and relative clauses, tend to cause more problems for lower-level learners, and this confirms studies which show these grammar items are acquired relatively late.

Finally, this study made it possible to add to the research trying to identify which language features are typically used at different levels of the CEFR, by showing that errors against third person -*s* and articles are typical of B1 up to B2, or that C1 is characterised by a higher number of attempted conditionals and passives, which still contain errors.

Future research could be designed to take into account the possible effects of the candidates' first language on error types and range of grammatical features. It might be interesting to control for L1 effects by obtaining samples across all band levels from specific L1 backgrounds.

Lastly, another aspect worthy of further investigation is the impact of task type on the kinds of structures that are elicited. Use of linguistically more demanding tasks is likely to yield a wider range of attempted structures at the higher band levels, and therefore, care should be taken with both task selection and interpretation of results.

References

Baily, N., Madden, C., and Krashen, S. (1974). Is there a "natural sequence" in adult second language learning? *Language Learning,* vol 21, pp 235–243.

Brown, R. (1973). *A First Language: The Early Stages*. Harvard University Press: Cambridge, Mass.

Bulté, B., and Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken and I. Vedder (Eds), *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*, pp 21–46. John Benjamins: Amsterdam.

Byrnes, H., and Sinirope, C. (2008). Advancedness and the development of relativisation in German: A curriculum-based longitudinal study. In L. Ortega and H. Byrnes (Eds), *The longitudinal study of advanced L2 capacities*, pp 109–138. Routledge: New York,

Cook, V. (1993). *Linguistics and second language acquisition*. St. Martin's Press: New York.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge University Press: Cambridge.

Dulay, H., and Burt, M. (1973). Should we teach children syntax? *Language Learning,* vol 23, pp 245–258.

Dulay, H., and Burt, M. (1974). Natural sequences in child second language acquisition, *Language Learning*, vol 24, pp 37–53.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press: Oxford.

Goldschneider, J.M., and DeKeyser, R.M. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants, *Language Learning,* vol 51 (1), pp 1–50.

Hawkins, J., and Buttery, P. (2010). Criterial features in learner corpora: theory and illustrations, *English Profile Journal,* vol 1 (1), pp 1–23.

Johnston, M., and Pienemann, M. (1986). *Second Language Acquisition: a Classroom Perspective*. New South Wales Migrant Education Service: Sydney.

Kang, O. (2013). Linguistic analysis of speaking features distinguishing general English exams at CEFR levels, *Research Notes*, vol 52, pp 40–48.

Krashen, S. (1977). Some issues relating to the Monitor Model. In H. Brown, C. Yorio and R. Crymes (Eds), *On TESOL '77*, TESOL, Washington DC.

Larsen-Freeman, D. (1976). An explanation for the morpheme acquisition order of second language learners, *Language Learning,* vol 26, pp 125–134.

Lightbown, P. and Spada, N. (2013). *How Languages Are Learned*. Oxford University Press: Oxford.

Luk, Z., and Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural –s, articles and possessive –s, *Language Learning,* vol 59 (4), pp 721–754.

Long, M., and Sato, C. (1984). Methodological issues in interlanguage studies: an interactionist perspective. In A. Davies, C. Criper and A.P.R. Howatt (Eds), *Interlanguage,* Edinburgh University Press: Edinburgh.

Meisel, J., Clahsen, H., and Pienemann, M. (1981). On determining developmental stages in natural second language acquisition, *Studies in Second Language Acquisition*, vol 3, pp 109–135.

Pica, T. (1983). Adult acquisition of English as a second language under different conditions of exposure, *Language Learning*, vol 33, pp 456–497.

Pica, T. (1984). Methods of morpheme quantification: their effect on the interpretation of second language data, *Studies in Second Language Acquisition,* vol 6, pp 69–78.

Pienemann, M. (1998). *Language Processing and Second Language Development: Processability Theory*. John Benjamins: Amsterdam

Rosansky, E. (1976). Methods and morphemes in second language acquisition, *Language Learning*, vol 26, pp 409–425.

Salamoura, A., and Saville, N. (2009). Exemplifying the CEFR: criterial features of written learner English from the English Profile Programme. In I. Bartning, M. Maisa and I. Vedder (Eds), *Eurosla monograph series 1. Communicative proficiency and linguistic development*, pp 101–132. Eurosla. Available at:
http://eurosla.org/monographs/EM01/101-132Salamoura_Saville.pdf

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing, and use*. John Benjamins Publishing Company: Amsterdam

Shin, S., and Milroy, L. (1999). Bilingual language acquisition by Korean schoolchildren in New York City, *Bilingualism: Language and Cognition,* vol 2, pp 147–167.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford University Press: Oxford.