What kind of interaction receives high and low ratings in Oral Proficiency Interviews?

Paul Seedhouse School of Education, Communication and Language Sciences, Newcastle University
Email: paul.seedhouse@newcastle.ac.uk

Abstract

Based on a Conversation Analysis (CA) of a corpus of Oral Proficiency Interviews (OPI), the study asks what kind of interaction receives high and low ratings in OPIs. The discussion focuses on issues of interactional organisation, considering turn-taking, sequence, repair and topic development in relation to candidate scores. The study presents findings of two funded studies of the International English Language Testing System (IELTS) Speaking Test (IST), which is one part of IELTS, a major international English proficiency test.

The article explains how interaction in the IST is organised in interactional terms and how this organisation generates opportunities to differentiate high- and low-scoring interaction. The study then lists the interactional characteristics of high-scoring and low-scoring tests, based on an inductive search through the database and analysis of the micro-interaction. Extracts are presented to support characterisations. Differences in score correlate to the following interactional differences in Parts 1 and 3 of the IST: ability to answer the question, engage with and develop a topic coherently, amount of trouble and repair, lexical choice, and identity construction. In Part 2 of the IST, length of turn may also be related to score.

Keywords: oral proficiency interviews, language assessment, spoken interaction, IELTS, Conversation Analysis, Scoring

1. Introduction

The English Profile project aims to develop the Common European Framework of Reference (CEFR) in relation to learners of English, and research has tended to focus on grammatical, functional and lexical features of learner English in relation to the six proficiency bands of CEFR (English Profile 2011). This article examines a question which complements English Profile research, namely: What kind of interaction receives high and low ratings in oral proficiency interviews (OPIs)? This is an area in which very little research has been done and the study aims to contribute to the development of learner profiles of oral production at different proficiency levels.

The approach adopted is an empirical and inductive search of a corpus of IELTS Speaking Tests (ISTs) to compare interaction in high- and low-scoring tests. In order to examine the micro-detail of the interaction and to understand how the interaction is organised to generate differential candidate performance, the study adopts a Conversation Analysis (CA) approach. It is argued that this is complementary to the quantitative, corpus-linguistics approach adopted in much English Profile research.

The discussion is mainly restricted to issues of interactional organisation, which is considered in relation to candidate scores. It also considers how interactional issues interconnect with lexis and syntax and how learners construct an identity through the details of their talk. This is an issue of interest to those wishing to develop profiles of learners at different levels, to teachers preparing students for OPIs, as well as to those involved in developing and validating tests.

2. Literature review

The study builds on existing research in two areas. First, it builds on existing work on interaction in oral proficiency interviews in general and on the IST in particular. Second, it builds on existing research into the specific issue of the relationship between features of candidate discourse and the scores allocated to candidates.

Studies of OPI talk have revealed the complex, multi-dimensional nature of the interaction and its relationship to test design. Young and He (1998) demonstrate that OPIs are examples of goal-oriented institutional discourse and constitute cross-cultural communication in which the participants may have very different understandings of the nature and purpose of the interaction. Their collection demonstrates that the systems of turn-taking and repair in OPIs differ from ordinary conversation. Lazaraton (2002) demonstrates that OPIs are a type of institutional interaction that shares properties with interviews (Drew & Heritage 1992) in the predetermination of interview actions and demonstrate interactional asymmetry (Drew & Heritage 1992). This asymmetry manifests the power difference in speaking rights between participants in the talk. McNamara and Roever (2006: 46) critique the traditional view of performance as an unproblematic display of individual competence and call for analysis of OPIs as a social event.

Several studies have examined the participation of individual learners in the event. He's (1998) micro-analysis reveals how a student's failure in an OPI is due to interactional as well as linguistic problems. Kasper and Ross (2001: 10) point out that their CA analysis of OPIs portrays candidates as "eminently skilful interlocutors", which contrasts with the general SLA view that clarification and confirmation checks are indices of non-native speaker (NNS) incompetence, whilst their (2003) paper analyses how repetition can be a source of miscommunication in OPIs. Wigglesworth (2001: 206) points out the need to ensure that learners obtain similar input across similar tasks. A number of studies have also been undertaken (Brooks 2009; Davies 2009; Lazaraton & Davies 2008; May 2009) of the interaction produced in paired format testing. These revealed some similarities and some differences (e.g. the interlocutor effect) with the interaction found in the examiner–candidate format of the IST.

The relationship between examiner and candidate has been the subject of research interest. In relation to the IST, Taylor (2000) identifies the nature of the candidate's spoken discourse and the language and behaviour of the oral examiner as issues of current research interest. Brown (2003) analyses two IELTS tests[1] involving the same candidate taking the same test with two interviewers with different interactional styles. The candidate's communicative ability in the two interviews was rated differently by four raters. The study emphasises the need for interviewer training and standardisation of practices; this was subsequently implemented in the design of the IST (see Section 3 below, and Taylor 2001). O'Sullivan and Lu (2006) examined cases where examiners deviated from the interlocutor frame in the IST and found that this had limited impact on the candidate's talk.

According to Lazaraton (2002: 161) "there has been very little published work on the empirical relationship between candidate speech output and assigned ratings". It is important to know how candidate talk is related to scores for a number of reasons. Test developers may use discourse analysis of candidate data as an empirical basis to develop rating scales (Fulcher 1996, 2003). Similarly, candidate talk may be used for validation processes. An empirical description of the architecture of an OPI can be useful in verifying validity and in determining whether the interaction is as was envisaged or not. Lazaraton (2002) presents a CA approach to the validation of OPIs, suggesting that qualitative methods may illuminate the process of assessment, rather than just its outcomes. It cannot be taken for granted that OPIs will generate clearly distinguishable candidate discourse. For example, Douglas's (1994) study of the AGSPEAK test related candidate scores to the categories of grammar, vocabulary, fluency, content and rhetorical organisation and very little relationship was found between the scores given and candidate discourse produced. Douglas suggests this may have been due to inconsistent rating or raters attending to aspects of discourse which were not on the rating scale.

There have been two previous studies into the relationship between features of candidate discourse in the IST and how these relate to scores allocated to candidates. Brown (2006) developed analytic categories for three out of the four rating categories employed in the IST and undertook quantitative analysis of twenty ISTs in relation to these analytic categories. Whilst she found that, in general, features of test-takers' discourse varied according to their proficiency level, there was only one measure which exhibited significant differences across levels, which was the total amount of speech. Her overall finding (2006: 71) was that "while all the measures relating to one scale contribute in some way to the assessment on that scale, no one measure drives the rating; rather a range of performance features contribute to the overall impression of the candidate's proficiency". Lazaraton's (1998) study of the previous version of the IELTS spoken test examined twenty tests and compared the relationship between candidate talk and ratings. Findings were that: there are fewer instances of repair at higher levels; higher-scoring candidates use a broader range of expressions to speculate; grammatical errors are more common in lower bands and complex structures in higher bands; appropriate responses are more common in higher bands, as is conversational discourse. The current study aims to build on the previous work and is the first micro-analytic study of the relationship between candidate talk and scores in the current version of the IST.

---

[1] The spoken tests involved were the predecessor to the IST.

3. Background information on the IELTS Speaking Test (IST)

The IELTS Speaking Test is one part of IELTS, the most widely used English proficiency test for overseas applicants to British and Australian universities. The IST is designed to assess how effectively candidates can communicate in English and is predominantly used to predict whether a candidate has the ability to communicate effectively on programmes in English-speaking universities. Over 4,000 certified examiners administer over 1.5 million ISTs annually at more than 500 centres, in over 135 countries around the world (http://www.ielts.org). ISTs are encounters between one candidate and one examiner and are designed to last between eleven and fourteen minutes. There are three main parts, each of which fulfils a specific function in terms of interaction pattern, task input and candidate output. In Part 1 (Introduction) candidates answer general questions about themselves and a range of familiar topic areas. In Part 2 (Individual long turn) the candidate is given a verbal prompt on a card and is asked to talk on a particular topic. The candidate has one minute to prepare before speaking for between one and two minutes. The examiner then asks one or two rounding-off questions. In Part 3 (Two-way discussion) the examiner and candidate engage in a discussion of more abstract issues and concepts which are thematically linked to the topic prompt in Part 2, for example "Can you compare ideas that architects and the general public have about buildings?"

Examiners receive detailed directives in order to maximise test reliability and validity. The most relevant and important are that standardisation plays a crucial role in the successful management of the IST:

> The IST involves the use of an examiner frame which is *a script that must be followed* [original emphasis] . . . Stick to the rubrics – do not deviate in any way . . . If asked to repeat rubrics, do not rephrase in any way . . . Do not make any unsolicited comments or offer comments on performance.
>
> (IELTS Examiner Training Material 2001: 5)

The degree of control over the phrasing in Parts 1 and 2 is carefully controlled to ensure that all candidates receive similar input delivered in the same manner. In Part 3, the frame is less controlled so that the examiner's language can accommodate the candidate's level. Detailed performance descriptors describe spoken performance at the nine IELTS bands, based on the following criteria: Fluency and Coherence; Lexical Resource; Grammatical Range and Accuracy; Pronunciation. Key indicators for the criteria are available to examiners. Scores were reported only as whole bands (e.g. 6.0 or 7.0) at the time these recordings were made.

3.1 The IELTS test and the CFER

A number of studies have discussed the relationship between the IELTS test and the CFER (Davidson & Fulcher 2007; Milanovic 2009; Taylor 2004a, 2004b; Weir 2005). A mapping of the IELTS scale onto the CEFR is available on http://www.ielts.org/researchers/common_european_framework.aspx, but comparisons between the two are problematic for a number of reasons, as detailed in the publications cited above. There

are six CFER levels and IELTS has a 9-point scaling system. It is important to note that the IELTS scores mentioned in this article are for the IST only and not the overall band scores, which are compared to the CEFR in the mapping referenced above.

However, for the purposes of this article, and purely as a point of reference, the examples of high-scoring ISTs employed are of IELTS 9.0 and 8.0, which very roughly correspond to CEFR levels C2 to borderline C2/C1. The examples of low-scoring ISTs employed are of IELTS 5.0 and below, which very roughly corresponds to CEFR borderline B1/B2, moving downwards into B1.

4. Data sampling

The primary raw data consist of audio recordings in cassette format of ISTs, with some centres providing digital recordings. The dataset for this study was drawn from recordings made during 2003 and 2004. Secondary data included paper materials relevant to the ISTs recorded on cassette, including examiners' briefs, marking criteria, examiner induction, training, standardisation and certification packs (Taylor 2001). A sample of 606 recordings was selected to cover four different tasks among the many used for the test. Following a quality check of the recordings, 197 of the best recordings were transcribed in total using the CA conventions available in the Appendix. Transcription being a very lengthy process, the number of transcripts was determined by the budget and time available. The aim of the sampling was to ensure that there was variety in the transcripts in terms of gender, region of the world, task/topic number and IST band score. The test centre countries covered by the data are: Albania, Bangladesh, Brazil, Cameroon, China, United Kingdom, Greece, Indonesia, India, Iran, Jamaica, Kenya, Lebanon, Malaysia, Mozambique, Netherlands, Norway, New Zealand, Oman, Pakistan, Philippines, Syria, Thailand, Vietnam and Zimbabwe. However, there is no information on nationality, L1s or ethnicity. Overall test scores ranged from IELTS 9.0 to 3.0. For the purposes of this article, the comparison was between ISTs with high scores of 9.0 and 8.0 (32 total) and low scores of 5.0 or below (44 total). The rationale for comparing tests at both ends of the scoring continuum was that differences should be most evident at the extremes. The 197 transcribed ISTs now form part of the English Profile corpus available to consortium members.

5. Methodology

The methodology employed is Conversation Analysis (Lazaraton 2002; Seedhouse 2004). CA studies the organisation and order of social action in interaction. This organisation and order is one produced by the interactants by means of their talk and is oriented to by them. The analyst's task is to develop an 'emic' or participants' perspective, to uncover and describe the underlying system which enables interactants to achieve this organisation and order. CA analysts aim to provide a 'holistic' portrayal of language use which reveals the reflexive relationships between form, function, sequence and social identity and social/institutional context. That is, the organisation of the talk is seen to relate directly and reflexively to the social

goals of the participants, whether institutional or otherwise. This methodology is appropriate to the research question as it is an inductive approach which examines the micro-detail of the interaction and can therefore uncover differences between high- and low-scoring ISTs.

As with other forms of qualitative research, the principles are not to be considered as a formula or to be applied in a mechanistic fashion. Analysis is bottom-up and data driven; we should not approach the data with any prior theoretical assumptions or assume that any background or contextual details are relevant. Another way of presenting the principles of CA is in relation to the questions which it asks. The essential question which we must ask at all stages of CA analysis of data is: "Why that, in that way, right now?" This encapsulates the perspective of interaction as action (why that) which is expressed by means of linguistic forms (in that way) in a developing sequence (right now). Talk is conceived of as social action, which is delivered in particular linguistic formatting, as part of an unfolding sequence. The first stage of CA analysis has been described as *unmotivated looking* or being open to discovering patterns or phenomena. Having identified a candidate phenomenon, the next phase is normally an inductive search through a database to establish a collection of instances of the phenomenon. After an inductive database search has been carried out, the next step is to establish regularities and patterns in relation to occurrences of the phenomenon. The specific features of individual cases are investigated in depth and are used to build a general account of a phenomenon or interactional organisation (Heritage 1984). We can only understand the organisation of the interaction and its emic logic by detailed analysis of individual instances. In the current study, when a potential characteristic of high- or low-scoring interaction was identified, a search was made amongst other ISTs to determine how systematic and widespread it was; all characteristics presented here are a result of that process. The aim is, then, to produce an account of the data which is both particularised and generalised. This involves a constant, reflexive interaction between the specific instance and the interactional system being studied (Seedhouse 2004).

I have described the process of data analysis, but the presentation of findings below is rather different. Due to limitations of space, it is necessary to select single examples to illustrate and compare specific interactional features of high- and low-scoring tests; the examples have been selected to be representative of such features within the corpus. All extracts are of different candidates.

6. The interactional organisation of the IST

In order to understand the relationship between interactional features and IST scores, it is important to have (a) an overview of how the interaction is organised and (b) an understanding of how the interactional organisation is able to generate differences in high- and low-scoring behaviours.

The organisation of interaction in the IST may be summarised as follows (Seedhouse & Egbert 1996; Seedhouse & Harris 2011). The organization of turn-taking and sequence closely follows the examiner instructions. Part 1 is a succession of teacher question–candidate answer adjacency pairs. Part 2 is a long turn by the student, started off by a prompt from the examiner and sometimes rounded off with examiner questions. Part 3 is another succession of

teacher question–candidate answer adjacency pairs, but these are intended to have a slightly less rigid organisation than Part 1. The topic of the talk is predetermined, written out in advance in scripts and is introduced by the examiner. Trouble generally arises for candidates when they do not understand questions posed by examiners. In these cases, candidates usually initiate repair. Examiner instructions are to repeat the question verbatim, once only. Examiners very rarely initiate repair in relation to candidate utterances. This is because the institutional goal is not to offer formative feedback; it is to assess the candidate's utterances in terms of IELTS bands. Overall, the organisation of repair has a number of distinctive characteristics which may be typical of OPIs in general. These are the lack of requirement to achieve inter-subjectivity, and an absence of verbally expressed evaluation or correction of errors by the examiner. The interaction is rationally organised in relation to the goal of ensuring valid assessment of oral proficiency, with standardisation being the key concept.

It is necessary to understand how the organisation of the interaction generates opportunities to differentiate high- and low-scoring tests. In Parts 1 and 3 of the IST, there is an archetypal organisation which combines turn-taking, adjacency pair and topic, as follows. All examiner questions (with the exception of the administrative questions) contain two components: (a) an adjacency pair component, which requires the candidate to provide an answer, and (b) a topic component, which requires the candidate to develop a specific topic. This organisation may be called a 'topic-scripted question–answer (Q–A) adjacency pair'. So in the IST, unlike conversation, topic is always introduced by means of a question. In order to obtain a high score, candidates need to do the following: (a) understand the question they have been asked; (b) provide an answer to the question; (c) identify the topic inherent in the question; and (d) develop the topic inherent in the question. So this core interactional structure generates multiple means of differentiating high- and low-scoring responses, as illustrated below. In order to fully understand the relationship between topic and score, it is necessary to break down the unitary concept of 'topic'. At this stage I therefore introduce the concepts of *topic-as-script* and *topic-as-action* in relation to interaction in the IST. Topic-as-script is the statement of topic on the examiner's cards prior to the interaction, whereas topic-as-action is how the topic is developed or talked into being during the course of the interaction. Whether and how candidates develop topic-as-action is consequential for the grades they received and therefore of direct relevance to the institutional business; this is illustrated in the analyses of extracts below.

7. Interactional characteristics of high-scoring and low-scoring ISTs

7.1 The topic-scripted question–answer (Q–A) adjacency pair

This section identifies empirical differences between candidates with the highest (9.0 and 8.0) and lowest scores (5.0 and below). There is no simple relationship between the candidate's score and features of their interactions, since a multitude of factors affect the examiner's ratings (Brown 2006: 71; Douglas 1994: 134). However, as broad generalisations, it is reasonable to provide this list of the interactional characteristics of high-scoring and low-scoring tests, based on an inductive search through the current corpus. We focus first on talk in the dialogic Parts 1 and 3 and then on Part 2, the individual long turn. Individual extracts are presented

to support characterisations. A Part 1 sequence (score 9.0) is shown in extract (1):

(1)
```
35   E:   okay (0.7) so uh:: what qualifications or certificates (0.8) do you hope to
36        get (1.3)
37   C:   well (1.1) after I: (0.5) get my degree in May I'm hoping to:: (1.3) uh:m
38        >probably work in England for a while and in order to do that I have to
39        do further exams< hh (0.5) unfortunately bu:t uh:m (1.1) ·hh then I just
40        hope to: (0.6) progress further i- in my field ((inaudible)) (0.2)
41   E:   okay okay (0.7) ↑let's uh move on to talk about some of the activities
42        you (0.6) enjoy in your free time (0.7) when do you have free time? (1.3)
43   C:   rarely hh heh (0.3) ·hh uh::m (0.5) I try to pace myself generally (.) in
44        terms of: getting a lot of work done during the week so I ca:n at least
45        relax a bit at the weekends (0.5) I like to:: look at movies go shopping:
46        hh heh (0.5) uhm have a chat with friends and (0.6)
47   E:   okay and uh::m (1.5) what free time activities are most popular where
48        you live? (1.6)
          (0389)
```

In extract (1), the 'topic-as-script' has been predetermined and is read by E in lines 35 and 41. Let us examine how C develops 'topic-as-action'. The question "what qualifications or certificates do you hope to get?" (lines 35–36) could be answered quite directly and drily as "I hope to get an MBA". However, C constructs a four-phase narrative presenting a vision of his/her future. The topic is developed as an action which develops a personal identity and which has the potential to engage the listener with him/her on a personal level. In lines 41 and 42, E introduces two topics-as-scripts, namely what the activities are that C enjoys in his/her free time and the question of when s/he has free time. Of particular note is the way in which C engages in lines 43–46 with these two scripted topics in reverse order and connects the two. In line 43 C's utterance "rarely" answers the second question, but the answer to the first question about free-time activities does not come until line 45. C very skilfully manages a stepwise transition of topic in lines 43–45 to move seamlessly from the second topic-as-script to the first. The topical action required is to move from 'lack of free time' to 'free-time activities' and this is accomplished by the explanation of pacing him/herself to work hard during the week to free up relaxation time at the weekends. C's development of topic-as-action works on a number of levels simultaneously. It projects an image of C as someone who is ambitious and hard-working, internationally mobile, gaining a number of qualifications, someone who plans their time carefully and has a clear vision of their life and their future. If we ask how C has taken a topic-as-script and developed it into a topic-as-action in this case, it is predominantly that C has developed a narrative of his/her personal life which projects a certain identity and enables a listener to engage with this. Moreover, the narrative is carefully structured in relation to temporal sequence; lines 37–40 portray four different time phases, one of which is not in linear order, whereas lines 43–46 show the ability to present generalisations about time.

In extract (1) we saw how C developed a topic-as-action and this simultaneously provided action on other levels: (a) it answered the questions; (b) it projected C's identity; (c) it displayed

C's level of linguistic and interactional competence; and (d) it displayed C's competence in engaging in the testing activity. So although topic-as-script in this setting is static, monolithic and predetermined, topic-as-action can be complex, dynamic and entwined with multiple actions on multiple levels.

However, not all candidates develop topic-as-action so successfully. In extract (1), we saw that C answered the questions and developed the topics. High-scoring candidates appear to be able to develop a topic-as-action concisely and without carrying on for too long, bearing in mind limitations of time. This demonstrates their competence in the assessment activity as well as their linguistic and interactional competence. In contrast to the successful example above, a candidate response may in principle (a) answer the question but fail to develop a topic-as-action; (b) fail to answer the question, but say something which bears some tangential relationship to the general topic-as-script; and (c) fail to answer the question or develop the topic-as-action. In cases (a) and (b), candidates will not achieve the highest scores for their responses and will receive the lowest ratings for case (c).

An example of a candidate answering questions without developing the topic is provided in extract (2):

```
(2)
142   E:       do you think that you will travel mor::e in the future
143            (0.4)
144   C: →     y[eah ]
145   E:        [whe]n you're older.
146            (0.3)
147   C: →     yeah
148            (0.5)
149   E:       because you enjoy it (0.9) okay now ((name omitted)) in this
150            part (0.7) I'm: going to give you a topic
               (002381t132)
```

In extract (2), the candidate (score 4.0) provides minimal answers to the questions but does not engage with the topic in any way. Answers may also fail to answer the question, but say something which bears some tangential relationship to the general topic, as in extract (3):

```
(3)
40   E:       okay (0.6) let's talk about public transport (0.5) what kinds of public
41            transport are there (0.3) where you live (2.0)
42   C:       it's eh (0.5) I (0.4) as eh (0.4) a (0.3) person of eh (0.4) ka- Karachi, I
43            (1.1) we have many (0.8) public transport problems and (0.7) many eh
44            we use eh (0.4) eh buses (0.4) there are private cars and eh (.) there are
45            some (0.3) eh (0.4) children (0.4) buses (0.8) and eh (1.9) abou- (0.2)
46            about the main problems in is the (0.4) the number one is the over eh
47            speeding (0.5) they are the oh eh (0.5) the roads (0.8) and eh (.) they are
48            [on]
49   E: →     [I  ] didn't ask you about the problems (0.6) my question was (0.6) what
```

50   kinds of public transport are <u>there</u> (.) where you live (0.7)
            (0219: Part 1)

In line 49 the examiner explicitly treats the candidate's answer as trouble in that it did not provide a direct answer to his/her question, even though it was on the general topic of public transport.

(4)
9   E: ((name omitted)) (0.2) alright (0.3) .hhh and can you tell me
10   where you're <u>from</u>
11   (0.7)
12   C: I'm <u>originally</u> from the <u>philippines</u> (.) but for the la::st erm
13   (0.3) it's like this before I came to canada in the year two
14   <u>thousand</u> I was working in <u>thailand</u>. (0.5) for three years for
15   an international organisation that <u>dealt</u> [with]
16   E: →   [okay] >sorry sorry-<
17   sorry to stop you I don't mean to [(inaudible)] where are you =
18   C: [ha ha okay]
19   E: from (0.3) the philippines right?
20   (.)
            (00198T130)

Some IELTS teaching materials (e.g. Jakerman & McDowell 2001) suggest that a good strategy in Parts 1 and 3 is to provide an answer plus one extra piece of topic-relevant information. The data suggest that this is indeed a feature of high-scoring interaction. In extract (4), we see a candidate who goes beyond this 'one extra' principle. In line 12, the answer is provided, then in 13 and 14 a single extra piece of information is added. However, when in line 16 further information is added on the business of the organisation, this is seen by E as going off-topic and repair is initiated. A rational explanation for this is that the test has a limited duration and the examiner must get through a set number of questions and keep to a timescale. Topics cannot, therefore, be allowed to be developed indefinitely. We also noted above that it was possible for candidates to fail to answer the question or develop the topic, and an example of this is provided in extract (5) below.

From a testing perspective, the archetypal organisation of topic-scripted Q–A adjacency pair in the IST appears to be very successful in generating differential performance between candidates. From a CA perspective, the topic-scripted Q–A adjacency pair is a remarkably economical instrument for carrying out the institutional business; a single examiner move requires a candidate move in response, which can be used by raters to distinguish levels of performance in relation to multiple issues, such as topic development and answering the question.

7.2 Trouble and repair

We now focus on the connection between repair and test score. There does appear to be a correlation between test score and occurrence of trouble and repair: in interviews with high

test scores, fewer examples of interactional trouble requiring repair are observable. Lazaraton (1998) also notes this in relation to the previous version of the IELTS spoken test. As noted above, examiner questions contain two components: (a) an adjacency pair component, which requires the candidate to provide an answer, and (b) a topic component, which requires the candidate to develop a specific topic. Trouble may occur in relation to the question or the topic inherent in the question, or both.

(5)
```
38   E:       and would you recommend this area as a place to live
39            (1.2)
40   C:       in my::: (0.4) in my house.
41            (0.3)
42   E:       would you recommend (.) the area (.) as a plac:e (0.2) to live
43            (0.3)
44   C: →     erm:: (0.3) .hhh (0.6) ↑recommend.
45            (.)
46   E:       m ↑hm
47            (.)
48   C:       m:::: (1) in my house is (1.9) have a- (0.6) in my h- (0.3) in
49            my hou have fours (.) roo[::m    ]
50   E:                                [m hm] =
51   C:       = bedroom (0.5) and is (0.2) two toilets
52            (.)
53   E:       m [hm]
54   C:         [I   ] live with (0.3) I (0.2) I (.) live with mother[::].
             (00181t130)
```

In extract (5) (score 5.0) the candidate is, despite repetition, unable to (a) understand the question they have been asked; (b) provide an answer to the question; (c) identify the topic inherent in the question; and (d) develop the topic inherent in the question, although some attempt is made to develop the topic 'place to live'. The candidate repeats a specific lexical item (*recommend*) using a specific repair initiation technique. This implies that their trouble in comprehension relates to that specific lexical item.

7.3 Display of engagement with topic

As noted above, to achieve a high score, candidates must develop the topic inherent in the question.

(6)
```
139   E:      °sh t-° ((could be L1)) (0.6) why do you choose certain types of
140           foo:d. (1.9) why do you choose to (0.7) certain types of
141           foo::d.
142           (0.2)
```

143   C: →   because I like it (0.2) hhhe .hhh he .hhh
144   (0.2)
          (002144t130)

The candidate in extract (6) (score 4.0) does provide an answer in line 143, and it is topic-relevant. However, the answer provided is not one which develops the topic substantially or would enable the examiner to continue to develop the topic further. In this sense, the response is topic-closing rather than topic-engaging, a minimal response to a topic. Although the response is linguistically correct, it does not provide a linguistic display of higher-level competence and uses a basic level of syntactic construction and lexical choice.

(7)
165   E: (0.4) <u>can</u> you explain <u>why</u> mo:st people buy <u>newspapers</u>.
166   (0.6)
167   C: →   erm <u>basically</u> to <u>update</u> themselves (0.4) to::: (.) er::m (0.3)
168   <u>give</u> themselves <u>give</u> themselves information they (.) <u>might</u> need
169   (0.4) for:: (0.4) their <u>everyday</u> lives for their <u>work</u> for their
170   erm (0.6) for their <u>studies</u>
          (007150t130)

In extract (7) (score 8.0) the candidate engages with the topic by expanding beyond minimal information and by providing multiple examples. The response would enable the examiner to develop the topic further, with the examples providing possible 'branches' to follow.

7.4 Topical coherence

Another aspect of topic in relation to score differentiation is whether candidates develop it coherently or not.

(8)
145   E: what would you <u>like</u> to <u>change</u> about your eating <u>habits</u>:.
146   (1.4)
147   C: my <u>eating</u> habit uh
148   (0.2)
149   E: m::.
150   (2.3)
151   C: →   I don't know <u>how</u> to say ah (0.2) because (0.7) I'm not (5.5)
152   maybe I will- hhh .hhh I (rarely) think (0.5) so:: (1.7) I
153   know:: (2.4) ((inaudible)) for me (.) I just (0.7) (I need for
154   other paper) er::: (0.3) I just eat only
155   (1.1)
          (002144t130)

Candidates with low scores sometimes struggle to develop a topic and provide a coherent answer, as in extract (8) (score 4.0). Turns often feature lengthy pauses. Given the short duration of the test, candidates would be better advised to say that they cannot answer a question and move to one which enables them to display their linguistic ability, rather than struggle for a long time with topics they cannot develop. See also extracts (5), (11) and (13) for similar examples.

(9)
128   E: why did you choose to learn english and not another language
129   (0.6)
130   C: erm:: (1.8) in hong kong I think commercially (0.3) erm english
131   is very useful (0.9) for business and also erm (0.3) in hong kong
132   there is a lot of (.) people coming from different countries
133   (0.2) so if you can speak a good english you can communicate with
134   those people (0.2) who do not know cantonese or speak mandarin
135   (0.2)
        (5756t130)

By contrast, the candidate in extract (9) (score 8.0) develops the topic coherently and economically, using the markers *also* (line 131) and *so* (line 133) to connect clauses, together with a defining relative clause in line 134.

7.5 Syntax

The ability to develop topic if often linked to the ability to construct syntax.

(10)
101   E: fashion (.) are you interested in fashion?
102   (0.3)
103   C: not interested in that no[::]
104   E: = why not?
105   (0.5)
106   C: →   erm (0.2) I don't fee::l (0.6) I don't think it's necessary to
107   be fashionable as long as you are comfortable in what you are
108   wearing?
          (2001t134)

In extract (10) (score 9.0) lines 106–108, the candidate provides an answer which engages with the topic and which also displays the ability to construct a sentence with a subordinate clause.

(11)
120   E: m hm (1.2) do you think children should be encouraged to take
121   an interest in sports?

```
122        (.)
123   C:   ye::s:: (0.3) [because] (0.4) m:: (0.3) it's not (stronger) =
124   E:               [why::?  ]
125   C:   = (1) the- er (1.2) if our children? (1.1) don't like sport
126        (0.3) they will no (0.6) there (0.2) maybe er::: (1) no:: (.)
127        stronger (0.7) oh:. (2.3) the happy (0.2) maybe (0.2) was (0.5)
128        was then er sport
129        (0.5)
           (002144t130)
```

By contrast, the candidate in extract (11) (score 4.0) not only fails to develop the topic in a meaningful way, but also fails to construct anything resembling a syntactically complete turn. Note that in lines 125–128 C produces a fairly long turn (19 seconds), but this will not receive a high score.

7.6 Lexical choice

Candidates with a high score may develop a topic using lexical items which are less common and which portray them as having a higher level of education and social status.

```
(12)
77                 E:    m hm: (0.3) what do you enjoy about using computer
78                       (1.2)
79                 C: →  erm::: (0.6) I like it because it's erm (0.5) it sort of gives
80                       you a sense of exclusivity not many people know how to use the
81                       computers well these days
(41 lines omitted)
123                E:    he he he he (.).hhhh what kinda clothes do you li::ke
124                      (0.4)
(5 lines omitted)
130                C: →  more formal attire (0.7) makes me look
131                      better anyway (0.2) °yeah°
                         (2011t134)
```

The choice of relatively uncommon lexical items by the candidate (score 9.0) features *exclusivity* (line 80) and *attire* (line 129). This combines with the development of topic in both extracts to construct the identity of a top-rank professional. In extract (12), the claim of an 'elite' mastery of computers in lines 80–81 seems to reinforce the use of 'elite' vocabulary.

```
(13)
204   E:   okay I'll ask you another
205        question (0.5) er:m (0.6) can you compare ideas that architects
206        (0.2) and the general public (.) have (0.3) about buildings
```

```
207              (7.4)
208  C: →   m:: (1.2) ((inaudible)) (2.7) erm (1.6) I think (0.3) architect
209              (0.4) erm (4.3) is person (1.3) °who° (0.3) who (1.7) create the
210              building (1.3) ra- erm (1.3) it h- (0.4) I think (4.3) they
211              (0.6) they ch- erm (4.2) they should (0.8) keep (0.6) erm (1.9)
212              the traditional (0.4) erm (0.2) the way (1.4) they come fro::m
213              (10.2) °yes erm°
214              (3.1)
                 (00128t133)
```

The response in extract (13) (candidate score 5.0) features very lengthy pauses as well as an answer which is lacking in coherence and direction. The candidate recycles lexical items used by the examiner in the questions. The candidate's turn is very long (1 minute 7 seconds) but will not rate highly.

7.7 Identity construction

Through the way they develop the topic in their answers, candidates construct an identity which may relate to their score band in some way. Candidates who achieved a very high score typically developed topics which projected the identity of an intellectual and a self-confident (future) high-achiever on the international stage. The candidate in extract (14), for example, achieved a score of 9.0:

```
(14)
19                  E:      right ↑so (0.9) can you tell me about your s:tudy first
20                          wha'd'ya- (.) what'd'you st[udy  ].
21                  C:                                 [e::rm] my major is british american
22                          studies? at (deleted) university? .hh (0.4) [a:nd] basically =
23                  E:                                                   [right]
24                  C: →    = what we do is we study:: (0.8) the essential core of british
25                          and american:: (0.7) society? culture politics literature?
(41 lines omitted)
67                  E:      m::. (0.3) okay .hhh al↑right (0.3) and erm::: (0.4) what- what
68                          kind of (0.5) work or what kind of career d'ya (.) d'ya hope to
69                          do in the future.
70                          (0.3)
71                  C: →    um what I hope to do is (0.3) either (.) between diplomatic
72                          studies? (0.2) o::r (.) something like (.) er policy::. (0.6)
73                          po- policy formulation in terms of development studies or
74                          environmental studies.
                            (000132t135)
```

Another candidate with a score of 9.0 explains in extract (15) that s/he is not interested in sports:

(15)
102 C: →   it's always been <u>ar:ts:</u> (0.4) erm (0.8) <u>arts</u> an literature
103   mostly (.) performing arts sometimes (0.8) er:: and when I
104   (0.3) when I <u>graduated</u> from <u>high</u> school <I <u>knew</u> (0.3) that>
105   (0.7) <u>eventually</u> I'm gonna (0.5) end up in the <u>writing</u> field
             (000198t130)

Candidates with low scores, by contrast, tended to develop topics in a way which portrayed them as somebody with modest and often localised aspirations. The candidate in extract (16) had a score of 5.0:

(16)
119 E: = m hm (0.3) and has learning english been <u>difficult</u> for you.
120   (0.4)
121 C: →   m::: (.) actually it (1.2) <u>difficult</u> for me because the <u>second</u>
122   (.) <u>language</u> (0.3) <u>and</u> normally I never <u>speak</u> (0.7) <u>english</u> (1)
123   just <u>thai</u>:: [((inaudible))] (.) [yeah]
             (00181t130)

The response portrays the candidate as a weak language learner who rarely communicates on an international level and is lacking in self-confidence.

Lazaraton and Davis (2008) introduced the concept of language proficiency identity (LPID), identifying ways in which a LPID can be talked into being by candidates through their talk. Lazaraton and Davis (p. 318) suggest asking the question: "What are the means by which test takers can position themselves as proficient and competent in a speaking test?" The current study concurs that this is an area worthy of further investigation, particularly in relation to examiner reaction to the LPID which candidates construct through their talk.

7.8 Characteristics of low-scoring interactions

By examining extract (17) (candidate score 5.0), we can pull together a number of the characteristics already identified above.

(17)
21   E: and can you tell me the <u>good</u> things about your <u>hous:::e</u>
22   (12.5)
23   E: tell me the <u>good</u> things about your <u>hous::e</u>
24   (1)
25   C: →   °°it's local°°
26   (0.6)
27   E: °°m (0.3) hm:°°

```
28            (7.5)
29    E:      <what is the area (.) like> where you live
              (00022t132)
```

In line 22 there is a very long pause of 12.5 seconds; the candidate does not request question repetition and this is eventually supplied by the examiner. When the answer is provided in line 25, it does (partly) answer the question and is linguistically correct. However, it does not engage with or develop the topic or enable the examiner to continue to develop the topic further and so is topic-closing. The response is minimal and actually only identifies one good thing about the house, whereas the question specifies 'things'. The examiner back-channels in line 27, a strategy commonly used to encourage candidates to develop topics further, but since the candidate fails to do so in the 7.5-second pause, the examiner moves on to the next question. The syntax employed is at a minimal, basic level and the lexical choice is very limited. Furthermore, the identity created by the candidate is one of someone with very local interests (line 25). Extract (17), then, typifies a number of profile features of a low-scoring IST candidate.

7.9 Length of turn

Brown's (2006: 84) quantitative study of discourse features in relation to proficiency level in the IST concluded that "the only (measure) to exhibit significant differences across levels was the total amount of speech. This is in many ways surprising, because amount of speech is not specifically referred to in the scales . . . it is not closely related to the length of response measures, which showed trends in the expected direction but were not significant." As a generalisation, candidates at the higher end of the scoring scale tend to have more instances of extended turns in which the topic is developed and questions answered in Parts 1 and 3. However, there are some weak candidates who take relatively long turns in Parts 1 and 3, but these do not always develop the topic or answer the questions, e.g. extracts (8) and (13) above. So in the IST Parts 1 and 3, length of turn needs to be related to ability to answer questions and to the topic in relation to score.

By contrast, there is clear evidence that very weak candidates produce short turns in Part 2. These often contain lengthy pauses. In Part 2, score may therefore be linked with the duration of the candidate's talk, as in extract (18):

```
(18)
109   E:     alright? (1.5) remember you have one to two minutes for this::
110          s:o don't worry if I stop you (0.7) I'll tell you when the time
111          is up? (0.4) can you start speaking no::w pleas::e.
112          (0.7)
113   C:     the job I'd like to do is sports management (0.6) then what I
114          need to bring the job is to manage the different kinds of
115          sports (0.5) and share advice with different sports people and
116          note their (0.7) feelings or attitude towards the sport they
117          ↑do (0.8) then (0.7) why- wa- they skills I need to the job
```

```
118        (0.2) can be:: (0.8) I have to be active in the sports I
119        ↑manage (0.5) and I have to be able to:: (0.8) know how the
120        sports:: is played and
121        (9.5)
122   E:   thank you.
           (000022t132)
```

The Part 2 talk in extract (18) (score 5.0) lasts only 39 seconds. No causal connection can be established in this area at present, although this is anticipated in the Instructions to IELTS Examiners (p. 6): "Weaker candidates may have difficulty in speaking for the full two minutes." This phenomenon would require further quantitative investigation of the corpus. For the moment, it appears that the relationship between length of turn and score varies in the different parts of the IST.

7.10 Scoring in the Part 2 individual long turn

In the candidate's individual long turn in Part 2, the expectation is that the candidate will (a) follow the instructions and (b) provide an extended development of a topic. Topic development in a true monologue is not analysable using CA techniques, since there is no interaction as such. However, in the case of Part 2, examiners occasionally produce back-channelling and generally ask 'rounding-off' questions at the end of Part 2. Furthermore, the candidate is speaking to the examiner (and an audio recorder) in orientation to the institutional goal, so I treat this as a dialogue and hence analysable using CA methods. I will now examine individual long turns from Part 2, one with a high score and one with a low score, to identify some of the key features of topic development which impact on score.

```
(19)
258                 E:   okay (1.6) erm ((clears throat)) I'd like you to describe a job
259                      (0.3) you think (0.4) would be interesting.
260                      (45.0: candidate prepares talk)
(4 lines omitted)
265                 C:   okay I think this job is er:: (1.3) being a doctor (0.3)
266                      medical practitioner (1.6) especially those who::. (0.5) erm
267                      (1.6) treat er (0.5) very (0.6) common diseases (0.4) not
268                      specialist (2.2) I think it's (0.3) quite interesting because
269                      er you- (0.3) you meet a lot of people (1.4) different kinds of
270                      people (1.1) and er (0.7) people who are having problems (0.3)
271                      people who are- people who are in pain (0.6) and er (0.7)
272                      you have to try as much as you can (0.7) to:: make them feel
273                      better (1.4) or try to solve your problems you cannot solve
274                      you- you put them in a state that they will feel more at ease
275                      (0.4) °with with him° (0.9) for example eh somebody (0.3)
276                      selling ah:: (0.6) let's say::: (0.8) erm:::. (2) selling er::
```

```
277                    let's say:: (1.7) MALAria (0.4) the illness is common in my
278                    pla::ce (1) okay (0.4) the patients who have this ((inaudible))
279                    condition erm:: (0.3) vomiting and all (1) this is very:::
280                    (0.2) ah:: troublesome (0.4) it (makes (0.2) makes you feel
281                    (mistaken) (0.7) (it's all) (.) it's up to the doctor to make
282                    you relax (0.4) even though your sick nearly won't go away
283                    (0.8) right after you see the doctor? but (0.7) the doctor must
284                    try to (0.3) er:: talk you to:: (1.2) er:: feel at ease (0.5)
285                    jus- (.) just to let you know that okay these are the
286                    medications I want to give you = these are the things you must do
287                    (0.8) and er (0.7) hopefully (.) it will work for you (0.8)
288                    (it's) all:: (0.3) I think that is a very interesting job
289                    (0.2)
290            E:     alright. (0.3) thank you
                      (2150t132)
```

The first point to note is that the candidate (score 9.0) clearly completes the instruction "describe a job you think would be interesting" (lines 258–259). Having identified the job of medical practitioner, all of the subsequent talk is related to this job and why it would be interesting. In line 265, the candidate identifies the generic term *doctor* and then narrows the focus to the sub-topic of types of doctor, moving to medical practitioner and non-specialists who treat common diseases (lines 266–268). In line 268 the topic shifts in stepwise fashion to why the job is interesting, the reason being because you meet many people and have to try to cure them. There is then an exemplification of an illness (malaria) which then leads stepwise to a description of symptoms and how patients feel and how the doctor should communicate with patients. So within the overall topic of the doctor's job, there is a good deal of stepwise, flowing development of sub-topics as well as shifts of perspective, presenting illness from the patient's and doctor's perspectives. Note also that an image of professional high-achievement and aspiration is created by the candidate's choice of profession.

```
(20)
176                    here's your topic? = I
177                    would like you to describe your favourite newspaper or
178                    magazine.
(10 lines omitted)
189            C:     m:::: okay. (2) m::. (2) about newspaper or magazine (0.3)
190                    there are many (0.4) pub- (0.7) of (.) no (.) many kind of
191                    article (0.6) er such as (0.3) programs (0.5) (fishing)
192                    entertainment (0.2) in the national news (0.2) local news (0.5)
193                    an:::: (1) other information about (0.3) er (0.7) television er
194                    (0.3) er:: or radio (1.2) .hhhh (0.3) um: (0.7) the:: (0.3)
195                    (the part er most I read is) er ((drops pen)) hhhe (1) is (the
196                    fashion) (1.7) m:: (0.2) maybe I- (0.2) maybe I like to read
```

197	(0.8) the (fishing) so:: (1.3) er (0.2) every- (0.2) every time
198	I:. (.) read the newspaper the first <u>part</u> (0.6) I read is the
199	(<u>fishing</u>) (0.7) hhhh (1.4) and what sort (0.3) of (paper) is
200	given. (0.7) to er (.) there's may::be:: (0.7) er some business
201	man (0.2) an::: (0.7) .hhhh other:: (1.2) <u>because</u> malaysians
202	like to (0.5) <u>stay</u> in the <u>restaurant</u> on a <u>saturday</u> (mama chows)
203	an or <u>other</u> chinese restaurant (0.5) they have a- the- new-
204	newspaper to (0.3) they have <u>time</u> to (0.5) er spending (0.5) er
205	(0.5) <u>reading</u> newspaper or magazine (0.7) so (1.3) m:: (1.5)
206	but for me (0.5) I like go to the- (0.3) I <u>like</u> go to:: (0.2)
207	coffee shops such as (0.3) ((inaudible)) (0.5) or starbuck
208	(1.3) or (0.4) another coffee <u>shop</u> (0.9) I read the <u>magazine</u>
209	(0.3) er (1.2) in the <u>shop</u> and enjoy:: (.) the <u>coffee</u> (1.4)
210	m:: (0.6) but (0.5) erm (1.8) how (0.5) how do I. (0.3) despite
211	a vision (0.7) why I (0.5) why I ((inaudible)) (0.4) because
212	(1) because (0.2) maybe I- (0.4) I am a- (0.) I am Chinese
213	(1) we are (0.2) we are a lot of (fishing) (0.4) er for
214	pr- (0.2) probably tuna fish an: (0.2) we ha- (0.7) I have
215	(0.2) read (0.3) a- a- already (0.7) <u>so</u>: (1.8) this may- (0.2)
216	this maybe a::: (3.4) maybe a::::. (0.7) maybe (0.6) maybe
217	(0.7) my hobby:: (0.2) so (0.5) I (reckon I read) (0.3) er
218	(0.2) I <u>read</u> everyday °and er°
219	(1.0)
	(002144t130)

The first point to note is that the candidate (score 4.0) never actually responds to "describe your favourite newspaper or magazine" in that one is never actually named or described. The general overall topic to which the candidate is talking seems to be 'say something about newspapers and magazines' and no sharp topical focus is achieved. Lines 190–199 mention the type of articles and information contained in newspapers and magazines in general. In line 201 there is a shift to where Malaysian people read newspapers. However, this does not appear to be a motivated, stepwise topic shift. In line 206 there is a stepwise transition to where the candidate likes to read a magazine. From line 210 onwards the topic seems to drift to other ways that the candidate and other Chinese Malaysians spend their time, including tuna fishing! This has drifted well away from the original question. Note also that the candidate does not develop an intellectual image through choice of reading matter.

From examining the data, a number of differences are evident between these high-scoring and low-scoring Part 2s. These include the extent to which the question is answered and the topic focus is clear, the extent to which topic shift is clearly marked, stepwise, motivated and flowing or not, and the length of time for which the topic is developed.

8. Conclusions

This study started from an analysis of the organisation of the interaction and how candidates oriented to this; the topic-scripted Q–A adjacency pair was found to be crucial in differentiating between candidate responses and scores. In contrast to Douglas's (1994) study of a different OPI, it was found that the IST generates clear and specifiable differences in interaction between high- and low-scoring candidates. The study has shown that it is possible to create a broad learner profile of interactional characteristics and differences between high- and low-scoring ISTs. Differences in score correlate to the following interactional differences in Parts 1 and 3 of the IST: ability to answer the question, engage with and develop a topic coherently, amount of trouble and repair, lexical choice, and identity construction. In Part 2 of the IST, length of turn may also be related to score. This suggests the possibility of developing performance descriptors for tests based on test data.

This approach may be able to make some contribution towards the English Profile project's aim of providing learner profiles for the CEFR levels, but a number of problems are evident. The approach in this study was to compare the highest- and lowest-scoring ISTs, but differences may not be so evident between interaction in the six CEFR levels. Furthermore, the IST produces a tightly regulated, homogenised variety of interaction, which means it is possible to conduct valid comparisons of individual ISTs. However, it would be very difficult to compare instances of interaction from a heterogeneous corpus featuring many different varieties of interaction.

It has been suggested (McCarthy & Carter 2002; Walsh & O'Keeffe 2007) that CA and corpus linguistics may be employed together in a complementary way. CA can been used to uncover differences in micro-interactional detail in OPI data, as in the current study. Uncovered differences could then be followed up by quantitative treatment using a large corpus and could be employed by test designers in developing and validating OPI ratings scales and profiles.

One limitation of the research it that it is not possible to confirm at present that the features identified in this study are in fact the same ones oriented to by the raters who scored these tests. A possible future direction for CA research in OPIs would therefore be to combine the approach of the current study with the research design of Brown's (2003) study, in which raters provided retrospective verbal reports on their reasons for awarding ratings. Further research might also be undertaken to relate generic types of assessment tasks to the differentiated discourse features noted in this study. Which features of task design might be most successful in generating differentiated discourse?

Acknowledgements

Egbert for her contribution to the first project, and to Andrew Harris for his contribution to the second project.

Appendix: Transcription conventions

A full discussion of CA transcription notation is available in Atkinson and Heritage (1984). Punctuation marks are used to capture characteristics of speech delivery, *not* to mark grammatical units.

[	indicates the point of overlap onset
]	indicates the point of overlap termination
=	(a) turn continues below, at the next identical symbol
	(b) if inserted at the end of one speaker's turn and at the beginning of the next speaker's adjacent turn, it indicates that there is no gap at all between the two turns
(3.2)	an interval between utterances (3 seconds and 2 tenths in this case)
(.)	a very short untimed pause
<u>word</u>	underlining indicates speaker emphasis
e:r the:::	indicates lengthening of the preceding sound
-	a single dash indicates an abrupt cut-off
?	rising intonation, not necessarily a question
!	an animated or emphatic tone
,	a comma indicates low–rising intonation, suggesting continuation
.	a full stop (period) indicates falling (final) intonation
CAPITALS	especially loud sounds relative to surrounding talk
º º	utterances between degree signs are noticeably quieter than surrounding talk
↑ ↓	indicate marked shifts into higher or lower pitch in the utterance following the arrow
> <	indicate that the talk they surround is produced more quickly than neighbouring talk
()	a stretch of unclear or unintelligible speech
((inaudible 3.2))	a timed stretch of unintelligible speech
(guess)	indicates transcriber doubt about a word
.hh	speaker in-breath
hh	speaker out-breath
hhHA HA heh heh	laughter transcribed as it sounds
→	arrows in the left margin pick out features of especial interest

Additional symbols

ja ((tr: yes))	non-English words are italicised, and are followed by an English translation in double brackets.

[gibee]	in the case of inaccurate pronunciation of an English word, an approximation of the sound is given in square brackets
[æ]	phonetic transcriptions of sounds are given in square brackets
< >	indicate that the talk they surround is produced slowly and deliberately (typical of teachers modelling forms)
C:	Candidate
E:	Examiner

References

NOTE: The following publications are not referenced as they are not publicly available:
Instructions to IELTS Examiners.
IELTS Examiner Training Material, 2001.
Examiner script, January 2003.

Atkinson, J. & Heritage, J. (eds.) (1984). *Structures of social action*. Cambridge: Cambridge University Press.

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing 26*, 341–366.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing 20*.1, 1–25.

Brown, A. (2006). Candidate discourse in the revised IELTS Speaking Test. *IELTS Research Reports 6*, 71–89.

Davidson, F. & Fulcher, G. (2007). The Common European Framework of Reference and the design of language tests: A matter of effect. *Language Teaching 40*, 231–241.

Davies, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing 26*, 367–396.

Douglas, D. (1994). Quantity and quality in speaking test performance. *Language Testing 11*, 125–144.

Drew, P. & Heritage, J. (eds.) (1992). *Talk at work: Interaction in institutional settings*. Cambridge: Cambridge University Press.

English Profile (2011). Introducing the CEFR for English. http://www.englishprofile.org/images/pdf/theenglishprofilebooklet.pdf (accessed 20/11/2011).

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing 13*, 208–238.

Fulcher, G. (2003). *Testing second language speaking*. Harlow: Pearson.

He, A. (1998). Answering questions in language proficiency interviews: A case study. In Young, R. & He, A. (eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam: Benjamins, 101–115.

Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge: Polity Press.

Jakerman, V. & McDowell, C. (2001). *IELTS Practice Tests Plus 1*. Harlow: Pearson Longman.

Kasper, G. & Ross, S. (2001). *'Is drinking a hobby, I wonder': Other-initiated repair in language proficiency interviews*. Paper presented at AAAL meeting, St. Louis, Missouri.

Kasper, G. & Ross, S. (2003). Repetition as a source of miscommunication in oral proficiency interviews. In House, J., Kasper, G. & Ross, S. (eds.), *Misunderstanding in social life. Discourse approaches to problematic talk*. Harlow: Longman/Pearson Education, 82–106.

Lazaraton, A. (1998). *An analysis of differences in linguistic features of candidates at different levels of the IELTS Speaking Test*. Report prepared for the EFL Division, University of Cambridge Local Examinations Syndicate, Cambridge.

Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge: Cambridge University Press.

Lazaraton, A. & Davies, L. (2008). A microanalytic perspective on discourse, proficiency, and identity in paired oral assessment. *Language Assessment Quarterly 5*, 313–335.

May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing 26*, 397–421.

McCarthy, M. J. & Carter, R. A. (2002). 'This, that and the other': Multiword clusters in spoken English as visible patterns of interaction. *Teanga 21*, 30–52.

McNamara, T. & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell.

Milanovic, M. (2009) Cambridge ESOL and the CEFR. *Research Notes 37*, 2–5.

O'Sullivan, B. & Lu, Y. (2006). The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. IELTS Joint Funded research report. http://www.ielts.org/pdf/Vol6_Report4.pdf (accessed 27/03/2012).

Seedhouse, P. (2004). *The interactional architecture of the language classroom: A Conversation Analysis perspective*. Malden, MA: Blackwell.

Seedhouse, P. & Egbert, M. (2006). The interactional organisation of the IELTS Speaking Test. *IELTS Research Reports 6*, 161–206.

Seedhouse, P. & Harris, A. (2011). Topic development in the IELTS Speaking Test. *IELTS Research Reports 12*, 69–124.

Taylor, L. (2000). Issues in speaking assessment research. *Research Notes 1*, 8–9.

Taylor, L. (2001). Revising the IELTS Speaking Test: Retraining IELTS examiners worldwide. *Research Notes 6*, 9–11.

Taylor, L. (2004a). Issues of test comparability. *Research Notes 15*, 2–5.

Taylor, L. (2004b). IELTS, Cambridge ESOL examinations and the Common EuropeanFramework. *Research Notes 18*, 2–3.

Walsh, S. & O'Keeffe, A. (2007). Applying CA to a modes analysis of higher education spoken academic discourse. In Bowles, H. & Seedhouse, P. (eds.), *Conversation Analysis and Languages for Specific Purposes*. Bern: Peter Lang, 101–140.

Weir, C. J. (2005). Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing 22*, 281–300.

Wigglesworth, G. (2001). Influences on performance in task-based oral assessments. In Bygate, M., Skehan, P. & Swain, M. (eds.), *Researching pedagogic tasks: Second language learning, teaching and testing*. Harlow: Pearson, 186–209.

Young, R. F. & He, A. (eds.) (1998). *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam: Benjamins.