

# The Pilot Speaking Examinations

Csépes Ildikó, Sulyok Andrea and Öveges Enikő

## Introduction

### The current oral examination

The current school-leaving oral examination consists of two tasks. The first is to read aloud and summarise a text of 10-15 lines that the candidate is not familiar with. The summary may be given either in Hungarian or in English. The teacher may ask the candidate to translate one or two sentences, or to give synonyms or definitions of certain words. The second task is a 5-10-minute conversation with the teacher on a topic of everyday life or culture. There is no preparation time between the two tasks. In the second task candidates pick a slip of paper with a topic on it. 'Marking is done by the candidates' own teacher, and the oral score forms 50% of the final result.' (Fekete et al., 1999:27) There are serious concerns as to the validity and reliability of this form of assessment, and so one of the main aims of the piloting process was to experiment with alternative ways of testing students' speaking abilities, and to explore ways in which the marking of oral performances could be made more reliable.

### The new speaking examination procedure

The speaking examination is designed to take 15-20 minutes according to the Test Specifications. The examination procedure is made up of four phases. First, there is a one- or two-minute-long warmer, aiming to put the candidates at ease. This part of the exam is not assessed. The candidates then are given three different Speaking tasks (each taking 5-6 minutes on average) that engage them in performing a variety of language functions in different situational contexts. The guidelines in the Specifications also suggest that some of the tasks may be conducted in either the traditional, individual mode or in the paired mode, where two candidates are expected to complete a task together. It is suggested that the examination should preferably end on a positive note by the examiner thanking the candidate(s) for their participation.

Based on the Test Specifications, both oral and written instructions for candidates must be given in English, except for Advanced level mediation tasks, but experimentation was envisaged with instructions in the mother tongue instead of English. Indeed, when piloting test tasks, there is an excellent opportunity to experiment with different test formats. In the Speaking test trials this meant that the instructions for some tasks were worded both in English and in Hungarian, but candidates were always given only one version, and some tasks were conducted in both the individual and the paired mode. The main purpose of the oral exam piloting in December 1998 was to find out how the examination procedure as a whole, its different formats as well as the selected Speaking tasks worked with the target population. The lessons that could be learnt from the piloting would be very important in terms of refining the Speaking test Specifications and working out guidelines for appropriate interlocutor behaviour. In addition, it was intended to develop and trial a rating scale for marking candidates' performance, as detailed in Chapter 6.

## The Speaking tasks

According to the Test Specifications, the tasks have to engage candidates in language use contexts that match their age and interest and do not require them to assume unfamiliar roles. The tasks can require candidates to give descriptions, make a summary, bridge information and opinion gaps, role play, argue, give and follow instructions/directions, etc. It is also suggested in the Specifications that the tasks can be accompanied by prompts, preferably authentic ones, including notices, advertisements, pictures, menus, instructions, charts, timetables, etc.

In order to ensure that a wide range of Speaking tasks as well as formats could be piloted, four different Speaking test booklets were developed for the December speaking pilots in 1998. The booklets included 10 different Speaking tasks altogether with some of them being piloted in two versions: with instructions given in either English or Hungarian and in both the individual and paired mode. Taking into account certain other aspects – such as prompts used, the direction in which information was meant to flow and language functions in focus –, the ten tasks can be described in the following way, shown in Table 11.1 below:

Table 11.1. A description of the pilot Speaking tasks

<b>task description</b> ( <i>language function in brackets</i> )	<b>one-way info flow</b>	<b>two-way info flow</b>	<b>prompt</b>	<b>instruction</b>	<b>mode</b>
Selecting 1 or 2 posters from a pool of 7 to decorate one's bedroom (giving reasons)	√		visual	in English	individual
Phoning a restaurant to reserve a table for three (making arrangements)		√	a restaurant advertisement	in English	individual
Planning a holiday in Spain (arguing, giving reasons)		√	holiday advertisement in Hungarian	in English & in Hungarian	individual & paired
Persuading your son not to go on a hiking trip abroad (persuading & arguing)		√	walking tour advertisement	in English	individual
Describing a room in order to find four differences in comparison to another picture (giving a description)	√		picture	in English	individual
Organising an afternoon & evening programme for a British penfriend (arguing & giving reasons)		√	verbal	in English	individual & paired
Describing a picture: a parachutist landing on the ground / a child being weighed in a doctor's surgery (giving a description, making deductions)	√		picture	in English	individual
Exchanging holiday tips for Spain (giving information, making suggestions, comparing & contrasting)		√	holiday advertisement in Hungarian	in English	individual & paired

Giving directions using a street map	√		2 different versions of a street map	in English & in Hungarian	individual & paired
Giving a recipe of a Hungarian soup	√		a recipe in Hungarian with a picture	in English & in Hungarian	individual & paired

As can be seen from Table 11.1, the tasks aimed to tap into candidates' ability to use English in a variety of different contexts, all of which had been previously accepted as authentic language use tasks for the target population.

### Test Booklet / Combination One

#### Task One

**Intended level:** Basic and Intermediate

**Task type:** taking part in a conversation based on visual prompts

**Prompt:** visual (seven posters to choose from)

**Task description and requirements:** giving description, making comparisons, expressing opinions, likes and dislikes, giving reasons

**Mode:** individual

**Instruction for Candidate:**

*On this page you can see some posters. Look at them and choose one or two you would put in your bedroom. Tell your reasons why you would like that one / those ones and where you would put it / them in your bedroom. While doing so, describe your bedroom (or the bedroom you would like to have).*

Unfortunately, due to the poor quality of the photocopying as well as the reduction of the original size of the pictures, it was a bit difficult for candidates to make out what some of the posters showed or what the writing said in them. Although the task seemed to work well in general, the selection of the posters should be made more carefully in the future in order to avoid this problem.

#### Task Two

**Intended level:** Intermediate

**Task type:** taking part in a conversation based on a verbal prompt

**Prompt:** advertisement

**Task description and requirements:** taking part in a telephone conversation and making arrangements for reserving a table

**Mode:** individual

**Instruction for Candidate:**

*You would like to eat out with two of your friends in a restaurant. Your friends are busy and asked you to call the restaurant and make arrangements. Use the advertisement below.*

<p><b>MICHAELS RESTAURANT</b>  An excellent varied menu.  A totally unique dining experience.  <b>3 Crown Street, Bolton,</b>  <b>Tel: (01204) 373325</b></p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

**Instruction for Interlocutor:**

- Inquire about:*
- number of people
  - time
  - what table (by window?)
- Give info:*
- no window table available between 6 p.m. and 8p.m.
  - no chicken dishes available tonight

This task seemed to require a greater contribution from the examiner-interlocutor, especially in the case of less talkative candidates. Trying to maximise co-operation, some examiners asked more questions from the candidates than what was suggested in the instruction. In order to make sure that the examiner-interlocutor can really keep a low profile in this exchange while remaining adequately polite, his or her contribution should be strictly observed.

**Task Three**

**Intended level:** Intermediate

**Task type:** taking part in a role-play

**Prompt:** advertisement (verbal and visual) in Hungarian

**Task description and requirements:** planning a holiday in Spain, persuading the partner to go on a round trip instead of staying in one place, arguing, giving reasons

**Mode:** individual and paired

**Instruction for Candidate A:**

*You are planning to go on holiday with your friend and his/her family. You would like to go on a round trip to see as much of the country as possible, but your friend would like to stay in one hotel.*

*Sum up the following advertisement to persuade him to choose the round trip.*

<p>COSTA DORADA – COSTA BRAVA KÖRUTAZÁS (9 nap)</p> <ol style="list-style-type: none"> <li>1. nap: Elutazás repülőgéppel Geronába, utána autóbusszos transzfer Salouba (250 km)</li> <li>2. nap: Pihenés az üdülőhelyen, fakultatív kirándulási lehetőség: Tarragona</li> <li>3. nap: Port Aventura. Kirándulás Európa legnagyobb látvány- és vidámparkjába.</li> <li>4. nap: Barcelona – Montserrat kirándulás után este érkezés a Costa Bravara.</li> <li>5. nap: Costa Barava. Pihenés a tengerparton, fakultatív kirándulási lehetőség.</li> <li>6. nap: Costa Barava. Pihenés a tengerparton, fakultatív kirándulási lehetőség.</li> <li>7. nap: Costa Barava. Pihenés a tengerparton, fakultatív kirándulási lehetőség.</li> <li>8. nap Hazautazás autóbusszal, útközben Monte.Carlo megtekintése.</li> <li>9. nap: Érkezés Budapestre az esti órákban .</li> </ol> <p>UTAZÁS: Oda repülővel, vissza nyugati típusú, légkondicionált autóbusszal.</p> <p>RÉSZVÉTELI DÍJ: 52.900 Ft / fő. (apartmanban, önellátással. Félpanzió befizethető.) 68.900 Ft – fő. (1-2 csillagos szállodában, félpanzióval.)</p> <p>AZ ÁR TARTALMAZZA: 7 éjszakai szállást a választott elhelyezésre és ellátással, idegenvezetőt</p> <p>AZ ÁR NEM TARTALMAZZA: Biztosítást, fakultatív kirándulást, múzeumi és egyéb belépőket.</p> <p>IDŐPONT: Szeptember 19 – 27</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



Note: This table is only an approximation of the original advertisement used in the pilot exam.

**Instruction for Candidate B:**

*You are talking to your friend about your holiday plans. Your friend would like to take a round trip in Spain. Ask him/her about the tour s/he has found advertised and tell him/her about your doubts (too much travelling, hard to relax, packing all the time, etc.)*

*Suggest going to the following hotel instead and argue for this choice:*

HOTEL REYMAR (T.: 72/34-03-12)

ELHELYEZKEDÉS: Tossa legszebb partján fekvő hotel, közvetlenül a tengerparton, csodálatos kilátással az öbölre.

KOMFORT: 156 szobás szálloda, úszómedence, teázó, disco, társalgók, bárók, étterem, kávézó, napozóterasz ágyakkal, gyermekjátsszótér, szauna, parkoló állnak a vendégek rendelkezésére.

AZ ÖN SZOBÁJA: Kényelmesen berendezett, légkondicionált, fürdőszoba (hajszáritóval), telefon, TV, minibár, erkély várja a vendégeket. Széf bérelhető.

ELLÁTÁS: Félpanzió. Svédasztalos reggeli és vacsora.

SPORT ÉS SZÓRAKOZÁS: Tenisz, asztalitenisz. Hetente kétszer szórakoztató műsorok.

CLUB ESPANA TIPP: Tengerre néző szobák külön igényelhetők.

Turnusváltás napja: kedd, szombat.

Note: This table is only an approximation of the original advertisement used in the pilot exam.

Some candidates seemed to get lost because of the amount of information given as prompt for the task. The suggestion is that the text be adapted, turned into a semi-authentic text.

## **Test Booklet / Combination Two Task One**

**Intended level:** Intermediate

**Task type:** taking part in a role-play

**Prompt:** advertisement (verbal)

**Task description and requirements:** persuading one's son not to take part in a hiking trip, arguing, giving reasons

**Mode:** individual

**Instruction for Candidate:**

*Your son, Bob, 16 shows you this advertisement saying that he wants to join one of the trips advertised. (Your son's role is taken by the examiner.) You are worried about him and want to persuade him not to go. Remind him of the dangers / the costs / his young age / lack of enough training / lack of experience, etc.*

WALKING WORLDWIDE

Small group of treks and hiking trips in Nepal, India, European Alps, Greece, Morocco, Tibet, East Africa, Peru, Bolivia, Ecuador, New Guinea, more. Free 36 pg. full color catalog. Himalayan Travel, Inc. Box 481-WKG.

It was strange that the candidate was required to take the role of his own father, however, the task seemed to work well.

## **Task Two**

**Intended level:** Basic and Intermediate

**Task type:** taking part in a conversation based on a visual prompt

**Prompt:** pictures

**Task description and requirements:** giving a description of a room which shows a living-room with modern furniture

**Mode:** individual

**Instruction for Candidate:**

*In this part of the exam there are two pictures. You have one, the examiner has another. There are lots of similarities and also differences in the two pictures.*

*Please describe your pictures, and the examiner will have to collect 4 differences between the two pictures. Do not show the examiner your picture.*

There is no real information gap, in most cases candidates merely described the pictures with very simple sentence structures. A way around the problem might be using pictures with some sort of action which would elicit a description of both objects and actions.

### **Task Three**

**Intended level:** Intermediate and Advanced

**Task type:** discussion

**Prompt:** none

**Task description and requirements:** organizing an afternoon / evening programme with a Hungarian friend for a British penfriend, expressing opinion, giving arguments

**Mode:** individual and paired

**Instruction for Candidate A:**

*You and your friend (the other candidate) would like to take your penfriend from Britain somewhere in the afternoon / evening. You like listening to pop music, and there is a band in the town playing at 9 p.m. tonight. Your penfriend likes sports, music and watching TV. With your Hungarian friend try and make a programme for the three of you for this afternoon and evening.*

**Instruction for Candidate B:**

*You and your friend (the other candidate) would like to take your penfriend from Britain somewhere in the afternoon or evening. You play football, and your favourite football team is having a match this evening at 8 p.m. Your penfriend likes sports, music and watching TV. With your Hungarian friend try and make a programme for the three of you this afternoon and evening.*

The instruction was not clear, candidates often misunderstood who is who, whose friend is who, what they must do. In a few cases candidates did not take the British boy into consideration, they ignored him when planning the programme. On the other hand, it is odd that two Hungarian friends should speak English when planning where to take out their English penfriend. The situation should be modified, e.g. one of the candidates is the visiting student from abroad. Also, more options should be given on what to do so that candidates would have more scope for negotiation.

### **Test Booklet / Combination Three**

#### **Task One**

**Intended level:** Intermediate

**Task type:** describing a picture / taking part in a conversation based on a visual prompt

**Prompt:** pictures (a choice of two)

**Task description and requirements:** describing a picture of a parachutist landing on the ground / a child being weighed in a doctor's surgery, finding out who might have taken the pictures and why, giving a description, making deductions

**Mode:** individual

The picture of the parachutist needs special vocabulary, it is recommended to be used at the Advanced level only. There is no instruction on the sheet.

## Task Two

**Intended level:** Intermediate

**Task type:** taking part in a role-play

**Prompt:** advertisement (verbal and visual) in Hungarian

**Task description and requirements:** students exchanging holiday tips for Spain, one of them has just returned, the other intends to leave shortly, looking up, collecting and passing on information, making suggestions, comparing and contrasting

**Mode:** individual and paired

### Instruction for Candidate A:

*You meet your friend in the street, who tells you that he has just come home after a 2-week holiday in Spain. Tell him/her that you are also planning to go there soon. Tell him/her what your hotel will be like and ask him/her about his/her experience (weather, prices, eating, etc.)*

APARTMAN LIDO LLOBET

ELHELYEZKEDÉS: Figueretas központjában, közvetlenül a homokos parton található ez a hangulatos ház, ahonnan gyönyörű kilátás nyílik az öbölre.

KOMFORT: Lift, felnőtt- gyermekmedence, napozóterasz, légkondicionált kávézó, garázs, kert. A tengerparton napernyők és napozóágyak.

AZ ÖN SZOBÁJA: 2-4 fős stúdió jellegű, szép apartmanok, felszerelt amerikai típusú konyhával, fürdőszobával, terasszal.

ELLÁTÁS: Önellátás. Félpanzió befizethető

SPORT ÉS SZÓRAKOZÁS: A tengerparton vízisport lehetőség. Hetente többször szórakoztató programokat szerveznek.

CLUB ESPANA TIPP: Itt minden adva van egy csodálatos nyaraláshoz.

Note: This table is only an approximation of the original advertisement used in the pilot exam.

### Instruction for Candidate B:

*You meet your friend in the street and now you are telling him about the two beautiful weeks you spent in Spain. As s/he is also going there soon, give him/her advice.*

*You spent your holiday in the following hotel:*

HOTEL CLUB GOLETA (T.: 71/30-26-62)

ELHELYEZKEDÉS: A szép szállodaegyüttes Plaza d'en Bossa központjában fekszik, közvetlenül a tengerparton, 2 km-re Ibiza városától.

KOMFORT: Légkondicionált étterem, bár, kávézó, társalgó, TV-szoba, diszkó, szauna, jacuzzi, üzlet, 2 felnőtt- és gyermekmedence, napozóterasz ágyakkal.

AZ ÖN SZOBÁJA: Részben tengerre néző, barátságosan berendezett, fürdőszobás, erkélyes, telefonos, ventilátorral felszerelt. Széf bérelhető.

ELLÁTÁS: Félpanzió. Svédasztalos reggeli és vacsora.

SPORT ÉS SZÓRAKOZÁS: A Goleta sportcentrum kínálata: 4 tenispálya, minigolf, asztalitenisz, biliárd, röplabda, futball, fitness-szoba.

CLUB ESPANA TIPP: A sziget egyik legattraktívabb szállodája.

Note: This table is only an approximation of the original advertisement used in the pilot exam.

The input is lengthy, so the task lends itself to a translation task easily. Candidates often only compare hotels, instead of discussing other aspects as well. Roles are not equally

balanced. It is unrealistic that two Hungarians talk about their holiday experiences in English. Playing a foreign friend's role would give a more realistic background to the exchange.

### Task Three

The same as Task Three in Test Booklet Combination Two.

### Test Booklet / Combination Four

#### Task One

**Intended level:** Basic and Intermediate

**Task type:** taking part in a role-play

**Prompt:** two different maps of the same part of London

**Task description and requirements:** as part of a telephone conversation giving/following directions using a street map

**Mode:** individual and paired

#### Instruction for Candidate A:

*You would like to visit the International Student Centre of London but you have no idea where it is. This is your first time in London, you don't know the town well. This is the map you have with you. You are calling a friend from a telephone box near a place called Hard Rock Cafe. Your friend has a map with the ISC on it. Ask him/her how you can get there.*

#### Instruction for Candidate B:

*You are staying in London. You have a map with the International Student Centre of London indicated by an arrow. A friend is calling you to find out from you how to get to the ISC. Find out where s/he is and explain how s/he can get there.*

The maps were complicated and of poor quality. The mark indicating Hard Rock Cafe on Student A's map should be made more visible. Also, International Student House should replace ISH in the instruction. Unfamiliar names of streets resulted in pronunciation problems. Map-reading in general appears to be a problem. Roles are not equally balanced. Candidates were liable to ignore that it was a telephone conversation.

### Task Two

**Intended level:** Advanced

**Task type:** mediation

**Prompt:** verbal and visual, a picture and the recipe of a Hungarian soup speciality

**Task description and requirements:** as part of a telephone conversation mediating the recipe of a Hungarian soup

**Mode:** individual and paired

#### Instruction for Candidate A:

*Your friend who you treated to dinner last night is phoning you to say thank you. He/She asks you to give him the recipe of the dish you had. The recipe is the following:*

#### TOKAJI SZŐLŐLEVES

Hozzávalók 4 személyre

• 6 dl tokaji bor • 3 dl víz • 2dl tejszín • 2 tojássárgája • 2 fürt szőlő • ízlés szerint cukor • darabka fahéj • 5 szegfűszeg • 1/2 citrom héja

A bort a vízzel és a kimagozott szőlő 2/3-ával feltesszük főni. A fűszereket kevés vízben felfőzzük, leszűrjük és levét a leveshez adjuk, melyet hagyunk felforni. Ekkor a tejszínt a

tojássárgákkal elkeverjük, és kevés forró leves hozzáadásával felmelegítjük. (Erre azért van szükség, hogy a tejszín a leveshez adva a hirtelen hő hatására ne csapódjon ki.) Ezután a leveshez öntjük és felforrósítjuk. Ha kissé kihűlt, turmixoljuk, átszűrjük és teljesen lehűtjük. A maradék szőlőszemeket beletéve tálaljuk.

Note: This table is only an approximation of the original recipe used in the pilot exam.

### **Instruction for Candidate B:**

*You are calling your friend to say thank you for the dinner you had with them last night. You ask him/her to give you the recipe of the soup s/he made, because everybody loved it.*

A task with such vocabulary load is likely to fail unless the candidate is really proficient. The roles are not equally balanced, so the task is not suitable for the paired format.

### **Task Three**

The same as Task Three in Test Booklet Combination One.

## **Lessons learnt from the piloting**

The most important thing about piloting test tasks is that test designers can reflect critically on what happens in the exam so that the procedure or the tasks themselves could be improved. In this section, some of the insights gained after the oral piloting will be summarised in order to shed light on the unforeseen weaknesses as well as positive features of the pilot oral examinations. The lessons learnt will relate to the warmer, the tasks, the paired mode, the rubrics, the timing of tasks, and the marking criteria. Markers were specifically asked to comment on any aspect of the tasks or the rating scales, and their comments are cited or summarised as appropriate. Candidates were also asked to fill out a feedback questionnaire immediately after taking the oral test, and when drawing conclusions about the different aspects of the oral exam, their feedback will also be discussed.

### **The warmer**

No interlocutor frame was available for examiners at the time of conducting the pilots, and so the warmer took an unnecessarily long time in some cases (5 to as many as 10 minutes). As this part of the exam is not assessed anyway, the examiners should strictly observe the maximum three-minute limit allotted for this phase and should be provided with a list of pre-specified questions, which they should adhere to as much as possible. It is not recommended that the pre-specified questions focus on topics of special interest, and questions relating to candidates' future should be restricted to school-leavers. It was also concluded after the piloting that in order to make the warmer sound more natural, personal remarks and positive verbal backchannelling from the examiner should be encouraged. In addition, especially in view of the length of the warmer, it was felt that an opportunity to assess the candidate was being wasted and therefore it was suggested that, in future, performance during the warmer should also be assessed.

### **The tasks**

Seeing the Speaking tasks in action provoked mixed feelings, especially on the part of the team that assessed the video-taped performances.

The assessors' team found that most of the tasks were not motivating enough. Where the prompts contained unfamiliar proper names (in Spanish or English) or too much information was crammed into them, candidates had to tackle task-irrelevant features,

which in turn seemed to have a negative impact on their performances. Furthermore, in some role plays candidates had to take rather unnatural roles. While we can all agree that pretending to be the father / mother of the examiner is the kind of role that should be avoided in the future – in fact, such roles were explicitly excluded in the Test Specifications – there does not seem to be a solution to the problem of role-playing two Hungarian friends planning a holiday together in English. The latter situation may seem to be somewhat unnatural, but the range of authentic contexts in which two Hungarians have to speak English to each other is extremely limited. If only authentic contexts were accepted, the testing of candidates' oral competence would be made rather limited too.

In the individual mode, examiner-interlocutors' contributions to the tasks varied considerably: sometimes they diverged from the prompts, i.e. went beyond the prompts to demonstrate their own creativity. In order to standardise interlocutor behaviour, it is important to develop an interlocutor frame for all the role play tasks in which the examiner takes part. In other words, interlocutors should be given detailed prompts for all the tasks in terms of how they are expected to contribute to the exchange and they also need to be given guidance on how a smooth transition between tasks could best be facilitated.

No matter how dissatisfied some of the assessors may have been with some of the tasks, it should be emphasised that candidates' feedback on all the tasks was overwhelmingly positive: in the follow-up questionnaire 81% of them said that they liked the tasks and 78% of them claimed that they would be happy to see similar tasks in the school leaving examination in the future. Some of the candidates also commented on the Speaking test as a whole that it

- was very interesting
- reflected language used in everyday contexts and topics
- emphasised one's ability to use English for real communication.

As for task difficulty, 44% of the test takers said that the tasks were of middling difficulty while nearly the same per cent of them claimed that the tasks were either difficult or easy (24% vs. 22%). By piloting the individual tasks on the target population, the Speaking trials in December were hoped to provide some insights into the difficulty level of the tasks, too. Unfortunately, due to the unsatisfactory performance of some of the tasks, the issue of task difficulty needs to be further investigated, i.e. more piloting is required in order to make the calibration of tasks possible.

### **Test format**

As was pointed out in the description of the ten tasks used in the Speaking trials, half of them were conducted in both the individual and paired modes. Unfortunately, some of the paired exam performances were problematic. Problems arose due to the imbalance between candidates' roles, especially in those cases where the information flow was one-way ('giving directions' and 'giving a recipe'). Therefore, item writers should bear in mind that only two-way information flow tasks should be designed for the paired mode, and the tasks should be constructed in such a way that the two interactants' potential contributions are comparable.

One of the most interesting outcomes of the speaking trials is that the team was able to collect evidence from the test takers that the paired mode was thought to be a welcome testing format by the students. Most of the candidates were very happy to do the tasks with a peer candidate. They commented on their experiences very favourably, although the quality of their test performances on the video did not necessarily support the view

that the paired format would generate better oral performances. What is important to note here is that the face validity of the paired exam seems to be high. According to the questionnaire responses, candidates found that

- the paired mode was 'good', 'better', 'easier', 'more lifelike'
- it helped them to relax because they had to talk to a peer partner
- mutual acquaintance helped to ease the tension and they knew what to expect from their partner
- they could help each other, e.g. while one of them was talking the other had time to think about what to say next.

Some candidates, however, were more critical of the paired mode and considered potential problems as well. A few remarked that performance in the paired mode very much depended on who one had as a partner. For instance, negative impact might be generated if one's partner is incomprehensible or if there is a big proficiency gap between the two peer interactants.

### **Rubrics**

One of the main concerns behind using instructions in English and Hungarian in different versions of the Speaking examination was to see whether the language in which the rubrics were worded made any difference. In general, we may conclude that it did not seem to influence candidates' performance very much. However, instructions (regardless of language) did cause problems for some of the candidates. This was obvious from their puzzled looks on the video as well as from their questionnaire feedback. Some test takers said that they did not always fully understand what was expected of them or the situation was not described clearly enough.

The team members were also concerned about the quality of the instructions. It was suggested that the examiner should always check briefly that the instructions accompanying each task had been properly understood by asking questions about various details of the task such as 'Where are you now?', 'Who are you now?' or 'What are you planning to do?', etc. The all-too-often-asked question of 'Is everything clear?' did not seem to be as effective as the aforementioned ones could be since the latter might provide misleading evidence on the candidate's part in relation to an adequate comprehension of the rubrics.

### **Timing of tasks**

The timing of the tasks also caused problems during the piloting. In some cases the tasks took an inordinate amount of time as the examiner's right to intervene had not been pre-specified. After viewing the candidates' video-taped performances, the team concluded that if the suggested time limit for task completion was over and it was evident that the candidate(s) had failed to achieve the task, the examiner should intervene and round off the conversation by saying something positive such as 'That'll do, thank you'. In order to help the examiner keep to the suggested time limits, it was recommended that a clock should be placed in the room in the future, preferably outside the candidates' field of vision so that it would not distract them.

## **The Speaking assessment scale in action**

### **Lessons learnt while applying the scale**

Markers made a number of useful comments on the use of the rating scales, which can be seen in Chapter 6, Appendix 6.2

#### **General comments on the rating scale**

- On the whole the scale worked well, and many of the problems can probably be attributed to problems with the tasks.
- The layout of the scale should be different: all the four criteria with the bands and band descriptors should be put on one page. In addition, there should be fewer bands.
- Separate scales will probably have to be developed for the two or three different levels (Basic, Intermediate, Advanced). Otherwise it is difficult to make appropriate distinctions.
- Very often the warmer was the part that elicited the most language. Should it perhaps be assessed? If not, then the time allocated to it will have to be controlled more carefully.
- To the lowest level of all the criteria should be added: 'there's not enough language produced to assess' as is done in 'Speech Quality'.

#### **Criterion-related comments**

##### ***Overall communication***

- 'Prompting' is mentioned at levels 0 and 1, then it is forgotten about and is only mentioned again at level 5. What about the in-between levels? 'Successful communication with some prompting' is level 5. This might be more appropriate somewhat lower.
- What happens if the candidate misunderstands the task? This is not addressed in the scale.
- The distinction between levels 6 and 7 could be refined as follows: 'communication is successful, uses some interactive and communicative strategies, contributes effectively throughout the interaction'. For level 6 the wording could be changed as follows: 'communication is usually successful, mostly contributes effectively throughout the interaction, but sometimes may require prompting'.

##### ***Grammar***

- The levels are probably set too high. What candidates can be expected to do in writing should not necessarily be expected in speaking, too.
- 'Complex structures' is introduced at level 6. Is it life-like and natural to use complex structures in speaking?
- It is difficult to differentiate between levels 1 and 2 and also between levels 3 and 4. The wording of level 3 could be changed as follows: 'reasonable control of grammar in a middling range of simple structures, sometimes grammatical mistakes might hinder comprehension'. Level 4 could be reworded as 'generally controlled use of grammar in a good range of simple structures, grammar mistakes do not impede comprehension'.

### **Vocabulary**

- It is hard to differentiate between the levels, especially the lower ones. Nine levels are too many, and as a consequence there is very little difference between descriptors.
- At the lowest level the line in brackets could be erased.
- At level 1 the wording could be changed as follows: 'limited repertoire consisting of only strings of words or phrases, which may often be used inappropriately, difficult to comprehend'.
- Must there be 'regular inaccuracies' at level 5? Does 'regular' mean *regular in time* or *the consistent misuse of certain items*?
- Abstract topics are task-dependent. If the task does not cater for abstract topics or offers no chance to express fine shades of meaning, candidates can only reach level 6.
- The word 'inaccuracy' might have to be defined.

### **Speech quality**

- There is no fluency sub-criterion.
- 'Sounds' and 'intonation' should be more detailed.
- The distinction between 1 or 2, or 3 and 4, 5 and 6 is very subjective, and distinctions need to be made clearer to increase the reliability of marking.

To summarise the most important conclusions: the four-page long marking scheme turned out to be rather user-unfriendly. All the raters felt that it was extremely difficult to use the grading scale in that format. It was suggested that the layout should be changed and perhaps the degree of detail given for each band should be edited. All the four assessment criteria ('overall communication', 'grammar', 'vocabulary', 'speech quality') with all the band descriptors should be put on one page. As seen above, there were problems with each individual criterion, which were mostly due to the inconsistency, ambiguity and incompleteness of the wording of the descriptors. It was often felt impossible to differentiate properly between certain levels as the distinctions were not made salient enough. For example, phrases like 'good control', 'generally good control' or 'reasonable control' can be very difficult to interpret in a reliable way. Furthermore, the vocabulary grading scale was found to be somewhat task-dependent: only if the task generated vocabulary on abstract topics was it possible to put the candidate in Band 7, 8 or 9. All the shortcomings of the grading scales highlighted above show that a thoroughly revised version is needed before the next round of piloting takes place.

### **Rater reliability**

Markers assessed the video-taped performances in pairs in order to allow the estimation of the inter-rater reliability – the agreement among raters. The latter can be measured by calculating the correlation coefficient, using the Spearman Rank Order correlation formula. Table 11.2 below shows the results of the inter-rater reliability calculations.

*Table 11.2: Inter-rater reliability for pairs of raters*

	Pair 1 : 2 (16 cases)	Pair 1 : 7 (13 cases)	Pair 3 : 4 (8 cases)	Pair 3 : 5 (10 cases)	Pair 4 : 5 (11 cases)
Spearman	.69 *	.94 *	.52	.41	.87 *

Notes: significant correlation coefficients are marked with \*  
Level of significance =  $p < .05$

As can be seen in Table 11.2, only three out of the five pairs were found to show significant agreement between the two raters' assessments of candidates' performances.

Individual pairs of raters were also examined in terms of how lenient or strict they were when assessing performance on the tasks. On comparing the mean scores on the individual criteria across all the three tasks in each pair, no significant pattern was found that would have reflected a tendency on the raters' part to assess one particular task more leniently or strictly than the others. In fact, lenience or strictness of marking varied from task to task as well as from rater to rater.

The following tables are intended to show the mean scores on three of the oral assessment criteria (Oc = overall communication; Vc = vocabulary; Gr = grammar) across all the three tasks (Oc1 = overall communication in Task 1, etc.). Speech quality was not rated separately for each task by all the markers, therefore scores on that criterion have been excluded from this analysis. The mean total score for each rater is also indicated (mTotal).

Table 11.3: Pair 1 & 2

	Oc1	Gr1	Vc1	Oc2	Gr2	Vc2	Oc3	Gr3	Vc3	mTotal
rater 1	2.43	2.12	2.62	2.56	2.25	2.31	2.62	2.12	2.25	21.31
rater 2	2.5	2.37	2.43	2.43	2.31	2.37	3.06	3.00	3.18	23.68

As can be seen, rater 1 seems to be a bit stricter than rater 2 (see especially task 3).

Table 11.4: Pair 1 & 7

	Oc1	Gr1	Vc1	Oc2	Gr2	Vc2	Oc3	Gr3	Vc3	mTotal
rater 1	5.61	4.92	5.23	5.46	4.84	5.15	5.38	4.69	5.15	46.46
rater 7	5.38	4.76	4.92	5.30	4.76	5.07	5.46	4.84	5.30	45.84

Rater 1 appears to be more lenient on the first two tasks, but she was somewhat stricter than her partner on task 3.

Table 11.5: Pair 3 & 4

	Oc1	Gr1	Vc1	Oc2	Gr2	Vc2	Oc3	Gr3	Vc3	mTotal
rater 3	3.75	3.87	4.37	4.25	4.37	4.37	4.25	4.12	4.25	37.62
rater 4	4.00	4.00	4.50	3.87	3.87	4.00	3.75	4.00	4.62	37.14

On task 1, rater 3 was stricter than rater 4, but on the other two tasks she was more lenient than the other (except for Vc3).

Table 11.6: Pair 3 & 5

	Oc1	Gr1	Vc1	Oc2	Gr2	Vc2	Oc3	Gr3	Vc3	mTotal
rater 3	3.5	3.7	4.1	3.7	3.5	3.5	3.9	3.8	4.0	33.7
rater 5	3.5	3.7	4.0	3.1	3.4	3.8	4.9	4.1	4.9	35.4

Rater 3 was clearly much stricter than rater 5 on the third task.

Table 11.7: Pair 4 & 5

	Oc1	Gr1	Vc1	Oc2	Gr2	Vc2	Oc3	Gr3	Vc3	mTotal
rater 4	3.91	3.69	4.13	4.15	3.79	3.87	4.2	3.82	3.98	35.58
rater 5	4.05	3.87	4.05	3.94	3.81	4.03	4.55	4.22	4.58	37.21

Rater 5 seems to have been generally more lenient than rater 4, but she was clearly more lenient on task 3.

These findings emphasise the crucial importance of marker training as well as the need for a revision of the marking scales themselves. The different patterns of raters' assessments might have been due to the potential ambiguity or usability of the marking scales, or to the lack of training. Once the scales have been revised, markers should be trained to use the scales so that they will interpret the band descriptors in similar terms.

## **Changes to the speaking assessment scales**

As part of the ongoing developmental process, the Year 12 group has revised the analytical scales planned to be used for the assessment of spoken performances in the new érettségi exam. The new scales of January 2000 (see the scales in Appendix 11.1) are level-specific, i.e. there are separate ones for the Intermediate and Advanced levels. However, they are general scales and task-specific scales will be produced together with the items annually. The revision was based on

- the Common European Framework of reference
- earlier versions of scales assessing speaking produced by the members of Year 10 and Year 12 teams jointly
- assessors' comments on the original scales above after they marked videotaped sample performances based on tasks that had been produced by item writers in the Project
- scales produced by the Year 10 team for the Basic Examination
- scales used in the Cambridge exams by UCLES assessors.

The four main criteria are slightly modified versions of the ones in the original scales. The new headings are the following:

- Task achievement and content
- Grammar and spelling
- Vocabulary
- Organization, cohesion and layout.

These appear to give a clearer picture of which sub-criteria (organized in bullet points, appearing in each band) the assessor has to pay attention to within a criterion.

Following the feedback comments on the original scales and after extensive discussions with testing experts, several other amendments have been made. There are fewer bands in the scales at present since they have been made level-specific. The distinctions between the different bands have been refined with more elaborate wording. The layout has been modified – all four criteria are on one page now, thus making them much more user-friendly.

The new scales were immediately pre-tested: the Project's item writers were asked to comment on them in the light of their earlier experience. They did not receive extensive guidance on how to use them (e.g. how they are supposed to use the blank bands), because we wanted to find out to what extent they could interpret these without instructions as this might provide us with precious information on what else, in what form we should include.

The **general comments** showed that the item writers found these scales much more user-friendly than the earlier versions. The new format and the clearer arrangement made them more transparent and easier to use. Most of the item writers, however, called our attention to the fact that they needed more explicit guidance on how to use the scales, especially the blank bands (2, 4 and 6). Moreover, they also expressed misgivings about giving feedback on scales without actually having assessed sample performances.

**Criterion-related comments related to the latest version of the speaking scales**

- There are still several definitions that are hard to interpret and should be refined. Such examples are ‘appropriate’ and ‘basic but appropriate’ or ‘sufficiently accurate’ and ‘mostly accurate’.
- Two item writers consider that the four main categories should not be of equal weight and they would both attribute less importance to speech quality.
- Certain adjectives used can result in subjective judgements, e.g. appropriate, basic, adequate.
- Communicative ability should be the first criterion mentioned.
- Sub-criteria could be emphasized more by putting them into italics in the tables, e.g.

<b>Organisation, cohesion, layout</b>	
<b>7</b>	In Intermediate level tasks: <ul style="list-style-type: none"> <li>• <i>Organisation</i> effective</li> <li>• <i>Cohesive devices</i>: various</li> <li>• <i>Paragraphing and layout</i>: fully appropriate</li> </ul>

- Some inconsistencies still remained in the scales, e.g. Vocabulary, band 5 lacks reference to range.
- The pass level should be made more striking, e.g. by putting grey background behind.
- Some item writers asked for task-specific criteria, which is in harmony with our understanding of the future scales.

After analyzing and applying the results of the pre-test mentioned above, the Year 12 group is planning to try the next draft of the scales on a small sample in April/ May 2000, using the new Speaking tasks produced recently.

**Oral test performance: test language in focus**

**Introduction**

This final section will summarise a small-scale study that aimed to explore how two sets of oral performance samples, one conducted in the traditional examiner-led mode, the other in the paired mode, differed in terms of the discourse produced.

It was hypothesised that the difference between a native or near-native speaker examiner’s and a candidate’s communicative language ability in the target language as well as their status in the testing context (the examiner-interlocutor being primarily responsible for a successful elicitation of language sample) would have a considerable impact on the discourse of the interaction. Thus, the main research questions were formulated in the following way:

1. Does the discourse produced differ if the interlocutor is an examiner or a peer candidate?

2. If yes, what are these differences in terms of style of interaction, rights and duties of interactants and dominance?

## **Method**

### ***The data***

Out of the original set of 49 video-taped performances from the Szeged pilot exams, 6 were selected and transcribed by the researcher, in which the same interaction task was carried out in both the individual and paired modes. Interactants were expected to organise an afternoon and an evening programme for an English penfriend. The six performance samples were selected in such a way that there would be candidates with both lower and higher levels of language proficiency. Proficiency ratings were given by two independent raters based on the video-recordings, using the grading scale developed for the speaking skill assessment, as described in Chapter 6. The total score candidates could get was 90, out of which the selected samples represent performances by both lower-level proficiency test takers (their scores ranged from 11-33) and higher-level ones (their scores ranged from 60-75). Both examiner-interlocutors in the selected samples had considerable experience in oral testing. Examiner 1 was a native speaker while Examiner 2 was a near-native speaker of English.

### ***Analysis***

It was thought that examining *who controlled* topics would give insights into the style of interaction in the two modes. Therefore, a working definition of *topic* was produced. To achieve this, the focus of each utterance was examined and was rated as a *topic* if it

- introduced a new theme/focus for the exchange in the form of a statement or a question
- introduced a new aspect of a given theme under discussion in response to a question
- was an extension of a given theme by one of the speakers in response to an utterance made earlier by the other speaker.

In *or-questions*, the alternative themes were treated as two different topics. Since some of the above categories were felt to be a bit loose, it was decided that the following should not be considered as a *topic* or *topic change*:

- paraphrasing / repetition of a given theme by either speaker
- responses to yes/no and confirmation-check questions
- reformulations of previous statements as questions
- repair or comprehension check questions
- responses that merely state agreement or disagreement with the other speaker's previous utterance.

After defining how topics were to be identified, a sub-category was established at the data analysis stage: topics introduced in questions were counted separately from topics initiated in statements. The reason for this distinction was the belief that questions were more likely to be an effective means of topic control than statements. When responding to the latter, an interlocutor seems to have more freedom to decide whether to ignore or react to the topic introduced in them. This category system is far from perfect, especially because statements can sometimes function as questions. Nevertheless, for the purposes of this study, it was felt that speakers' topic initiations could give a rough indication of who was in control of topic development in the two oral test modes.

### **Topic control**

Once the topics were identified, they were counted for each speaker and then examined again in adjacent turns in order to see whether they were ratified by the other speaker, i.e. whether there was some kind of response to the topic initiation. The reason why topic initiation and ratification were checked was that in this way it was possible to examine how participants reacted to each.

A marked sample transcript has been included below to demonstrate how the different conversational analyses were performed for the purposes of the study. First of all, let us examine what was considered to be a topic (topic is printed in bold) and whether it was seen to be ratified by the other speaker (topics which were ratified are also italicised).

#### *Sample Transcript*

- 1 C9: hello Laci  
C8: hello er + do you know that er + tomorrow **there is a (.) concert** (.) at (.)  
night o'clock  
C9: hm + // yes  
5 C8: // nine o'clock  
C9: \* er I think too (.) that we could + *vagy* + **we can + take (.)**  
**him too** but + **there will be a football match** (.) you know at nine *vagy*  
eight o'clock and + **I want to + see it** + watch it (.) want to watch it  
C8: + er yes + er + **we (.) should er see** the: er match and + **watching TV**  
10 and after that + we can er see the: concert  
C9: + yes, it would be great + he + **he likes sports and music** + er after (.)  
that + **we can go to (.) eat er at some restaurant** what do you think  
C8: yes it's a: great idea ++ and +++ and **there is a big problem** + er I don't  
know if (.) he likes our favourite band  
15 C9: yes + but er er + we should ask him if he + **we should ask him** what  
kind of music + he likes so + the er + then (.) make the programme if (.)  
we (.) can go to this concert or not + but er **where can we eat at the**  
**evening** + do you know a great place  
C8: yes **there is a: good restaurant** er near to my flat + and there is a lot of  
20 traditional foods, so it's a great place to eat something + //yes  
C9: //I think too + **he can (.) taste the (.)**  
**Hungarian (.) kitchen**  
C8: yes

In the above transcript, there are 13 topics highlighted, out of which 10 were considered to be ratified by the other party (5 by each of them). There is only one question asked by Candidate 9, the other topics were introduced in the form of statements. It can also be seen that somewhat more topics were initiated by Candidate 9 (8 vs. 5) but three topics introduced by him received no response from the other candidate. Overall, it can be concluded from this sample that neither speaker tried to control the conversation exclusively, and the style of discourse can be described as one where both speakers were mostly willing to react to their partner's initiations.

### **Rights and duties of the speakers**

With respect to rights and duties, the samples were examined from the point of view of topic initiations, whether conversation openings and closings were performed and by which party, and whether backchannel signals were used to show that a message was getting through. As can be seen in the sample transcript, there is a conversation opening by Candidate 9 but no explicit closing is done by either speaker. It seems that the artificiality of the testing context as the context of conversational exchange does not

prompt candidates to use conversation openings and closings automatically unless they are prescribed by the assessment task itself. A more important issue to examine is who performs openings and closings in the individual format: the examiner-interlocutor or the candidate? In the paired mode, this duty can always be performed by the candidates themselves. As for backchannel signals, in the sample transcript above, there is only one given in line 4.

### ***Dominance***

The question of who opens and closes the interaction is related to issues of who dominates the performance. Dominance of one speaker over the other can also be reflected in the quantity of talk and the length of turns, as well as interruptions. As for quantity of talk examined in this study, the number of words was counted and then divided by the number of turns in order to get an index of the average length of turns. For example, in the sample transcript, Candidate 9 spoke a bit more than Candidate 8 since the total number of words was 120 and 81 respectively, while the average length of turns was 13.5 words/turn for Candidate 9 and 20 words/turn for Candidate 8.

After demonstrating the methods of conversation analysis through examples, in the next section the results of all the analyses will be summarised and discussed in order to answer the main research questions of the study.

## Results and discussion

In order to provide an overall picture of the topic initiations and ratifications, a summary is shown in Table 11.8.

Table 11.8: Topic initiations and ratifications in the individual and paired modes

		individual mode								paired mode						
		Ts. 1.		Ts. 3		Ts. 5		average		Ts. 2.		Ts. 4.		Ts. 6		average
		E1	C1	E1	C4	E2	C7	E	C	C2	C3	C5	C6	C8	C9	C
			L		H		L			H	H	L	L	L	L	
Topic	Q	14	2	8	7	4	1	8.6	3.3	1	2	6	-	1	-	1.6
Initiations	S	4	6	6	5	5	1	5	4	10	3	-	7	7	5	5.3
Ratified	Q	14	2	8	7	-	1	7.3	3.3	1	2	6	-	1	-	1.6
Topics	S	1	6	5	5	1	1	2.3	4	7	3	-	5	4	5	4
Non-	Q	1	-	-	-	4	-	1.6	-	-	-	-	-	-	-	-
Ratified	S	2	-	1	-	4	-	2.3	-	3	-	-	2	3	-	1.3
Topics																
Total no. of		18	8	14	12	9	2	13.6	7.3	11	5	6	7	8	5	7
TI																
Total no. of		15	8	13	12	1	2	9.6	7.3	8	5	6	5	5	5	5.6
RT																

Note: Ts= Transcript; E= Examiner; C= Candidate; Q= Question; S= Statement; TI= Topic Initiation; RT= Ratified Topics; H= candidate with a higher proficiency level; L= candidate with a lower proficiency level; Averages were calculated in relation to the interactants.

As can be seen, in the individual mode examiner-interlocutors initiated almost twice as many topics as their counterparts (13.6 vs. 7.3 on average). However, there is no difference in the average topic initiations by candidates in the two modes (7.3 vs. 7). Topics initiated by examiners were often expressed directly in a question form, whereas candidates preferred to introduce their own topics in the form of statements irrespective of the mode, although Candidate 4 was exceptionally eager to ask questions. In fact, he sometimes responded to his partner's question with another question, as the extracts below will show:

Extract 1. (lines 1-2)

E1: so we have our (.) our **penfriend visiting** (.) **what shall we do with him**

C4: + a:h + erm + **do you like listening er** (.) **music listening to music**

Extract 2. (lines 13-15)

E1: yeah! that would be a good idea + erm **my favourite football team has got a**  
**match** (.) this evening at eight o'clock + **shall we go and watch that together**

C4: oh well er: + er what ++ **is it going to be a good match?**

The equal number of topic initiations by Candidate 4 and his examiner-interlocutor may be due to the higher proficiency level of the candidate. If we examine the topic initiations in the individual mode by Candidates 1 and 7, who are at a lower level of proficiency, achieving a balance between interlocutors with respect to conversational rights seems to be difficult. This can be explained by the huge proficiency gap that exists between the two interactants and, as a consequence of this, the examiner – interlocutor's efforts to elicit a long enough language sample from the candidate. The performances in the paired mode cannot be characterised by equal distribution of topic initiations either, but perhaps the overall topic initiations are more balanced. This may be attributed to the appropriate matching of candidates, whose proficiency levels did not differ hugely within the pairs. To sum up the above findings, it seems that there is a

difference between the styles of interaction in the individual and the paired modes in terms of who controlled the interaction.

Table 11.9 shows the distribution of the speakers' rights and duties, which are narrowed down to who performed conversation openings and closings and the frequency of using backchannel signals (topic initiations have already been discussed above).

Table 11.9: Speakers' rights and duties

	individual mode								paired mode							
	Ts. 1.		Ts. 3		Ts. 5		average		Ts. 2.		Ts. 4.		Ts. 6		average	
	E1	C1 L	E1	C4 H	E2	C7 L	E	C	C2 H	C3 H	C5 L	C6 L	C8 L	C9 L	C	
Conversation Openings	-	-	√	-	-	-	-	-	√	-	√	-	-	√	-	
Conversation Closings	-	√	-	-	-	-	-	-	-	-	-	-	-	-	-	
Backchannel Signals	6	-	4	1	-	-	3.3	0.3	1	3	1	-	-	1	1	

Note: Ts= Transcript; E= Examiner; C= Candidate; Average refers to the average frequency of backchannel signals; Averages were calculated in relation to the interactants; H= candidate with a higher proficiency level; L= candidate with a lower proficiency level.

Unfortunately, there were too few performance samples in this study to draw conclusions about conversation openings and closings: there were very few openings and even fewer closings (only 1). Backchannel signals were similarly scarce, and were mostly given by the examiner-interlocutor (E: 3.3. vs. C: 0.3 in the individual mode & 1 in the paired mode on average).

The issue of who seemed to dominate the interactions in the two modes has already been discussed in some detail. Table 11.10a & 11.10b are intended to provide new perspectives by showing the amount of talk produced, the length of turns and the number of interruptions.

Table 11.10a: Features of dominance in the individual mode

	individual mode							
	Ts. 2.		Ts. 4.		Ts. 6		average	
	E1	C1 L	E1	C4 H	E2	C7 L	E	C
No. of words	258	110	176	156	121	17	185	94
Av. length of turns	11.7	5.2	10.3	9.1	13.4	2.8	11.8	5.7
Interruptions	-	-	1	-	-	-		

Table 11.10b Features of dominance in the paired mode

	paired mode						
	Ts. 2.		Ts. 4.		Ts. 6		average
	C2 H	C3 H	C5 L	C6 L	C8 L	C9 L	C
No. of words	114	107	69	74	81	120	94.16
Av. length of turns	19	21	8.6	9.2	20	13.5	15.21
Interruptions	1	-	-	-	-	-	

Note: Ts= Transcript; E= Examiner; C= Candidate; Average length of turns=

number of words/ number of turns; Averages relate to the interactants; H= candidate with a higher proficiency level; L= candidate with a lower proficiency level

The average number of words and length of turns in the six performance samples suggest that the distribution of power between the interactants in the individual and pair modes was not always equal: when talking to the lower proficiency level test takers (C1 & C7), the examiner-interlocutors clearly dominated the interaction but in the case of the higher level candidate (C4) both the amount of talk and the average length of turns seemed to be similar for both interactants. As opposed to this uneven quantity of talk produced by different proficiency level testees in the individual mode, candidates performing the very same role play task showed a more even distribution of dominance between themselves irrespective of their level of proficiency: the quantity of talk produced by the candidates was quite similar within all the three pairs. As can be seen in Table 11.10b, the interactions in the paired mode can be characterised by much longer turns and shorter overall discourse lengths than in the individual mode: while candidates produced 5.7 words per turn on average in the latter mode, in the paired mode the average length of turns was 15.21 words.

With respect to the assumed impact of proficiency level on candidates' performance, we can conclude that while there were no big differences between lower and higher level candidates from the point of view of topic initiations (cf. Table 11.8 above), there were considerable differences with respect to the quantity of talk. As can be seen in Table 11.10b, higher level proficiency candidates (C2 & C3) in the paired mode produced less talk than their counterpart (C4) in the individual mode, but at the same time they produced longer turns. A similar pattern characterises lower proficiency level candidates: the quantity of talk seems to be less in the paired mode (compare C2 with C5 & C8, except for C9, who produced slightly more talk than C2), but the turns in the paired mode were again longer. These findings suggest that the pace of the interaction in the individual mode was faster, and interactants spent less time producing a turn whereas in the paired mode candidates were more patient with each other and let their partners make longer contributions within individual turns. This latter feature might be interpreted as a sign of co-operation between the peer-interactants. This assumption, however, does not mean that examiner-interlocutors were not co-operative. The number of interruptions is extremely low in both modes, which may be regarded as a sign of mutual support or co-operation in both modes.

It may be argued that the nature of co-operation was different in the two modes. Examiner-interlocutors may have tried to help the candidates by initiating a greater number of topics and using a greater number of backchannel signals, especially while talking to lower level candidates. As a result of that, they tended to both control and dominate the interaction. In the paired mode, however, candidates may have tried to help each other by allowing their partners to hold the floor for a longer time. This strategy could have had a twofold purpose: on the one hand, talking patiently with one's partner may have generated an encouraging and supportive atmosphere. On the other hand, by producing longer turns candidates were able to gain time, which they could use for preparing their next contributions while their partners were producing a turn. This assumption, however, could only be confirmed by asking the testees themselves as to what was really going on in their minds. Unfortunately, they were not interviewed after the oral pilot exam and the follow-up questionnaires did not focus on this issue either.

## **Conclusion**

This small-scale study was designed to investigate the main features of the interactional style of two sets of oral test performance samples in which the same role play task was carried out by candidates talking to an examiner-interlocutor, on the one hand, and a peer candidate, on the other. The findings show that there is not always a balance between the interactants in terms of conversational rights and duties in the individual mode, especially in the case of lower proficiency candidates. However, we cannot claim that interactions in the paired mode always reflect an even distribution of the same rights and duties.

In answer to the main research questions, discourse differences were found between the individual and paired modes of oral proficiency testing. It seemed to matter considerably whether the candidate's interlocutor was an examiner or a peer since there were more topic initiations and backchannel signals by examiner-interlocutors in the individual mode than by candidates in the paired mode. An imbalance of power between interactants in the individual mode was also reflected through the longer turns produced by the examiner-interlocutors. In the paired mode, although candidates produced shorter stretches of discourse, their turns were longer, compared to those of either the examiners or the single candidates, and more balanced across all the candidates. However, there were no real differences between the two modes in terms of interruptions, conversation openings and closings.

With regard to an assumed impact of proficiency level on candidates' performance in the two modes, a greater degree of dominance by examiner-interlocutors over candidates was observed in the case of lower-level candidates. In the paired mode, however, no significant differences in terms of dominance or topic control were found that seemed to have resulted from the fact that the interactants had either a lower or a higher level of proficiency. This finding suggests that a different style of co-operation may have been in operation in the two testing modes. In the paired mode, candidates were more patient with each other allowing their partners to spend longer on their contributions within individual turns. In the individual mode, however, the pace of interaction was faster, and the examiner-interlocutors tried to do their best to elicit language from the candidates by taking responsibility for topic initiation, which in turn, resulted in them controlling the interaction.

## Appendix 11.1a Assessment scales for the Speaking test – Intermediate level Version 2 Jan 2000

	<b>Grammar and discourse management</b>	<b>Vocabulary</b>	<b>Communicative ability</b>	<b>Speech quality</b>
<b>7</b>	In Intermediate level tasks <ul style="list-style-type: none"> <li>message is conveyed effectively using mostly accurate grammar with only occasional minor errors</li> <li>speech organisation is coherent</li> </ul>	In Intermediate level tasks candidate <ul style="list-style-type: none"> <li>uses appropriate and varied vocabulary</li> </ul>	In Intermediate level tasks candidate <ul style="list-style-type: none"> <li>is able to maintain effective communication</li> <li>contributions are always relevant</li> <li>shows sensitivity to turn-taking</li> <li>does not require assistance in conveying message</li> </ul>	In Intermediate level tasks <ul style="list-style-type: none"> <li>candidate produces the features of spoken English (individual sounds, stress, intonation) well enough to be understood easily</li> </ul>
<b>6</b>				
<b>5</b>	<ul style="list-style-type: none"> <li>message is conveyed adequately using sufficiently accurate grammar with frequent minor inaccuracies</li> <li>speech organisation is mostly coherent</li> </ul>	<ul style="list-style-type: none"> <li>uses appropriate vocabulary</li> </ul>	<ul style="list-style-type: none"> <li>is able to maintain flow of communication</li> <li>contributions are mostly relevant</li> <li>may lack sensitivity to turn-taking</li> <li>does not require major assistance</li> </ul>	<ul style="list-style-type: none"> <li>candidate produces the features of spoken English (individual sounds, stress, intonation) sufficiently well to be understood</li> </ul>
<b>4</b>				
<b>3</b>	<ul style="list-style-type: none"> <li>message is conveyed mostly adequately using a limited range of structures, with occasional major and frequent minor inaccuracies in grammar</li> <li>speech organisation may lack coherence</li> </ul>	<ul style="list-style-type: none"> <li>uses basic but appropriate vocabulary</li> </ul>	<ul style="list-style-type: none"> <li>has difficulties in maintaining flow of communication</li> <li>contributions may be inappropriate</li> <li>often lacks sensitivity to turn-taking</li> <li>may require assistance</li> </ul>	<ul style="list-style-type: none"> <li>candidate produces the features of spoken English (individual sounds, stress, intonation) sufficiently well to be understood but with some strain on the listener</li> </ul>
<b>2</b>				
<b>1</b>	<ul style="list-style-type: none"> <li>occasional breakdowns in communication prevents the message being conveyed, using mostly inaccurate grammar with frequent major errors</li> <li>speech organisation lacks coherence</li> </ul>	<ul style="list-style-type: none"> <li>uses limited and/or inappropriate vocabulary</li> </ul>	<ul style="list-style-type: none"> <li>is not able to maintain flow of communication</li> <li>contributions are inappropriate and/or irrelevant</li> <li>does not attempt turn-taking</li> <li>requires assistance</li> </ul>	<ul style="list-style-type: none"> <li>it is difficult to understand message because candidate produces some features of spoken English (individual sounds, stress, intonation) inaccurately</li> </ul>
<b>0</b>	No assessable output from candidate	No assessable output from candidate/	No assessable output from candidate	No assessable output from candidate

	<b>Nyelvhelyesség és szövegfelépítés</b>	<b>Szókinccs</b>	<b>Kommunikációs készség</b>	<b>Beszédminőség</b>
<b>7</b>	<p>Középszintű feladatokban a vizsgázó</p> <ul style="list-style-type: none"> <li>• a mondanivalóját hatékonyan közvetíti, többségében pontos nyelvtani szerkezeteket használ, esetlegesen előforduló kisebb hibákkal</li> <li>• beszédének felépítése koherens</li> </ul>	<p>Középszintű feladatokban</p> <ul style="list-style-type: none"> <li>• odaillő, változatos szóhasználat</li> </ul>	<p>Középszintű feladatokban a vizsgázó</p> <ul style="list-style-type: none"> <li>• képes a hatékony kommunikáció fenntartására</li> <li>• megnyilvánulásai mindig lényegesek, a tárgyhoz tartozóak</li> <li>• odafigyel a beszélgetés általános szabályaira</li> <li>• nem igényel segítséget mondanivalója közvetítéséhez</li> </ul>	<p>Középszintű feladatokban a vizsgázó kiejtése</p> <ul style="list-style-type: none"> <li>• kiejtése elég jól tükrözi a beszélt nyelv sajátosságait (mind az egyes hangok, mind a hangsúly és intonáció tekintetében) ahhoz, hogy beszéde könnyen érthető legyen</li> </ul>
<b>6</b>				
<b>5</b>	<ul style="list-style-type: none"> <li>• mondanivalóját megfelelően közvetíti, a nyelvtani szerkezeteket kielégítően pontosan használja, gyakori kisebb hibákkal</li> <li>• beszédének felépítése jórészt koherens</li> </ul>	<ul style="list-style-type: none"> <li>• odaillő szóhasználat</li> </ul>	<ul style="list-style-type: none"> <li>• képes a kommunikáció folytonosságának fenntartására</li> <li>• megnyilvánulásai többségükben a tárgyhoz tartozóak</li> <li>• időnként nem figyel oda a beszélgetés általános szabályaira</li> <li>• nem igényel nagyobb segítséget mondanivalójának közvetítéséhez</li> </ul>	<ul style="list-style-type: none"> <li>• elég jól tükrözi a beszélt nyelv sajátosságait (mind az egyes hangok, mind a hangsúly és intonáció tekintetében) ahhoz, hogy beszéde érthető legyen</li> </ul>
<b>4</b>				
<b>3</b>	<ul style="list-style-type: none"> <li>• mondanivalóját nagyrészt megfelelően közvetíti, a nyelvtani szerkezetek szűk skáláját használja, időnkénti nagyobb és gyakori kisebb hibákkal</li> <li>• beszédének felépítése nem mindig koherens</li> </ul>	<ul style="list-style-type: none"> <li>• egyszerű, de odaillő szóhasználat</li> </ul>	<ul style="list-style-type: none"> <li>• csak nehézségek árán képes fenntartani a kommunikációt</li> <li>• megnyilvánulásai esetenként nem megfelelőek</li> <li>• gyakran figyelmen kívül hagyja a beszélgetés általános szabályait</li> <li>• esetenként segítséget igényel mondanivalójának közvetítéséhez</li> </ul>	<ul style="list-style-type: none"> <li>• elég jól tükrözi a beszélt nyelv sajátosságait (mind az egyes hangok, mind a hangsúly és intonáció tekintetében) ahhoz, hogy beszéde érthető legyen, de a megértés némi erőfeszítést igényel a beszélgetőtárstól</li> </ul>
<b>2</b>				
<b>1</b>	<ul style="list-style-type: none"> <li>• a kommunikáció esetenkénti megszakadása megakadályozza a mondanivaló közvetítését, a nyelvtani szerkezeteket nagyrészt helytelenül használja gyakori nagyobb hibákkal</li> </ul>	<ul style="list-style-type: none"> <li>• szűk skálán mozgó és/vagy oda nem illő szóhasználat</li> </ul>	<ul style="list-style-type: none"> <li>• nem képes a kommunikáció fenntartására</li> <li>• megnyilvánulásai nem megfelelőek és/vagy nem a tárgyhoz tartozóak</li> <li>• meg sem próbálkozik a beszélgetési általános szabályinak betartásával</li> </ul>	<ul style="list-style-type: none"> <li>• annyira helytelen (mind az egyes hangok, mind a hangsúly és intonáció tekintetében), hogy mondanivalóját nagyon nehéz megérteni</li> </ul>

	• <i>beszédének felépítése nem koherens</i>		• <i>segítség igényel mondanivalójának közvetítéséhez</i>	
<b>0</b>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>

## Appendix 11.1b

### Assessment scales for the Speaking test – Advanced level – Version 2 Jan 2000

	<b>Grammar and discourse management</b>	<b>Vocabulary</b>	<b>Communicative ability</b>	<b>Speech quality</b>
<b>7</b>	In Advanced level tasks <ul style="list-style-type: none"> <li>message is conveyed effectively using a wide range of linguistic structures, although errors might occur in complex ones</li> <li>speech organisation is consistently coherent</li> </ul>	In Advanced level tasks candidate <ul style="list-style-type: none"> <li>precise and varied vocabulary</li> </ul>	In Advanced level tasks <ul style="list-style-type: none"> <li>candidate contributes fully and effectively to the communication</li> <li>contributions are always relevant and appropriate</li> <li>candidate participates with ease and maintains the norms of turn-taking</li> </ul>	In Advanced level tasks <ul style="list-style-type: none"> <li>candidate handles the features of spoken English (individual sounds, stress, intonation) effectively</li> </ul>
<b>6</b>				
<b>5</b>	<ul style="list-style-type: none"> <li>message is conveyed adequately using an adequate range of linguistic structures, although minor errors occur</li> <li>speech organisation is adequately coherent</li> </ul>	<ul style="list-style-type: none"> <li>varied vocabulary</li> </ul>	<ul style="list-style-type: none"> <li>candidate contributes effectively to the communication</li> <li>contributions are mostly relevant and appropriate</li> <li>candidate is sensitive to turn-taking</li> </ul>	<ul style="list-style-type: none"> <li>candidate produces the features of spoken English (individual sounds, stress, intonation) well enough to be understood easily</li> </ul>
<b>4</b>				
<b>3</b>	<ul style="list-style-type: none"> <li>message is conveyed mostly adequately using an adequate range of linguistic structures, although some major and minor errors occur</li> <li>speech organisation is mostly coherent with occasional inconsistencies</li> </ul>	<ul style="list-style-type: none"> <li>appropriate vocabulary</li> </ul>	<ul style="list-style-type: none"> <li>candidate contributes sufficiently to the communication</li> <li>contributions are mostly relevant but may be inappropriate</li> <li>candidate may lack sensitivity to turn-taking</li> </ul>	<ul style="list-style-type: none"> <li>candidate produces the features of spoken English (individual sounds, stress, intonation) sufficiently well to be understood</li> </ul>
<b>2</b>				
<b>1</b>	<ul style="list-style-type: none"> <li>uses inadequate linguistic structures, errors may obscure the message</li> <li>speech organisation lacks coherence</li> </ul>	<ul style="list-style-type: none"> <li>limited and/or inappropriate vocabulary</li> </ul>	<ul style="list-style-type: none"> <li>candidate contributes little to the communication</li> <li>contributions are irrelevant and/or inappropriate</li> <li>candidate lacks sensitivity to turn-taking</li> </ul>	<ul style="list-style-type: none"> <li>it is difficult to understand the message because candidate produces some features of spoken English (individual sounds, stress, intonation) poorly</li> </ul>
<b>0</b>	No assessable output from candidate	No assessable output from candidate	No assessable output from candidate	No assessable output from candidate

	<b>Nyelvhelyesség és szövegfelépítés</b>	<b>Szókinccs</b>	<b>Kommunikációs készség</b>	<b>Beszédminőség</b>
<b>7</b>	<p><i>Emelt szintű feladatokban a vizsgázó</i></p> <ul style="list-style-type: none"> <li>• <i>mondanivalóját hatékonyan közvetíti, széles skálán mozgó nyelvi szerkezeteket használ, a bonyolultabb szerkezetekben esetlegesen el•forduló kisebb hibákkal</i></li> <li>• <i>beszédének felépítése következetesen koherens</i></li> </ul>	<p><i>Emelt szintű feladatokban</i></p> <ul style="list-style-type: none"> <li>• <i>pontos és változatos szóhasználat</i></li> </ul>	<p><i>Emelt szintű feladatokban a vizsgázó</i></p> <ul style="list-style-type: none"> <li>• <i>teljes mértékben és hatékonyan járul hozzá a kommunikációhoz</i></li> <li>• <i>megnyilvánulásai mindig tárgyhoz tartozóak és megfelelőek</i></li> <li>• <i>könnyedén, nehézségek nélkül vesz részt a kommunikációban, betartja a beszélgetés általános szabályait</i></li> </ul>	<p><i>Emelt szintű feladatokban a vizsgázó</i></p> <ul style="list-style-type: none"> <li>• <i>a beszélt angol nyelv sajátosságaival (mind az egyes hangok, mind a hangsúly és intonáció tekintetében) hatékonyan bánik</i></li> </ul>
<b>6</b>				
<b>5</b>	<ul style="list-style-type: none"> <li>• <i>mondanivalóját megfelelően közvetíti, a nyelvi szerkezetek megfelelően széles skáláját használja, esetlegesen előforduló kisebb hibákkal</i></li> <li>• <i>beszédének felépítése összességében koherens</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>változatos szóhasználat</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>hatékonyan járul hozzá a kommunikációhoz</i></li> <li>• <i>megnyilvánulásai többségükben a tárgyhoz tartozóak</i></li> <li>• <i>odafigyel a beszélgetés általános szabályaira</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>kiejtése elég jól tükrözi a beszélt nyelv sajátosságait (mind az egyes hangok, mind a hangsúly és intonáció tekintetében) ahhoz, hogy beszéde könnyen érthető legyen</i></li> </ul>
<b>4</b>				
<b>3</b>	<ul style="list-style-type: none"> <li>• <i>mondanivalóját nagyrészt megfelelően közvetíti, a nyelvi szerkezetek megfelelően széles skáláját használja, néhány nagyobb és kisebb hibával</i></li> <li>• <i>beszédének felépítése jórészt koherens</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>odaillő szóhasználat</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>elfogadható mértékben járul hozzá a kommunikációhoz</i></li> <li>• <i>megnyilvánulásai esetenként nem megfelelőek</i></li> <li>• <i>nem mindig figyel oda a beszélgetés általános szabályaira</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>kiejtése elég jól tükrözi a beszélt nyelv sajátosságait (mind az egyes hangok, mind a hangsúly és intonáció tekintetében) ahhoz, hogy beszéde könnyen érthető legyen</i></li> </ul>
<b>2</b>				
<b>1</b>	<ul style="list-style-type: none"> <li>• <i>a nem megfelelő nyelvi szerkezetek használata és így az előforduló nyelvi hibák akadályozzák a mondanivaló eredményes közvetítését</i></li> <li>• <i>beszédének felépítése nem koherens</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>szűk skálán mozgó és /vagy oda nem illő szóhasználat</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>csak kis mértékben járul hozzá az eredményes kommunikációhoz</i></li> <li>• <i>megnyilvánulásai nem megfelelőek és/vagy nem a tárgyhoz tartozóak</i></li> <li>• <i>figyelmén kívül hagyja a beszélgetés általános szabályait</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>kiejtése annyira helytelen (mind az egyes hangok, mind a hangsúly és intonáció tekintetében), hogy mondanivalóját nagyon nehéz megérteni</i></li> </ul>
<b>0</b>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>	<i>Nincs értékelhető nyelvi megnyilvánulás</i>