

INTO EUROPE
The Speaking Handbook

INTO EUROPE

Series editor: J. Charles Alderson

Other volumes in this series:

Reading and Use of English

The Writing Handbook

Listening

Into Europe *The Speaking Handbook*

Ildikó Csépes
Györgyi Együd

CONTENTS

<i>Series Editor's Preface</i>	9
<i>Acknowledgements</i>	11
PART ONE General Introduction	
Introduction	17
Chapter 1: To the teacher	19
Chapter 2: What may influence test takers' oral performance?	27
PART TWO Designing oral examination tasks	
Introduction	35
Chapter 3: The Interview	37
Chapter 4: The Individual Long Turn	51
Chapter 5: Discussion	85
Chapter 6: Role play	103
PART THREE Examiner training	
Introduction	135
Chapter 7: Training Interlocutors	137
Chapter 8: Training Assessors	161
Recommendations for Good Practice	173
APPENDICES	
Appendix 1: Self-Assessment Statements for Speaking	180
Appendix 2: A2/B1 Level Speaking Assessment Scale	184
Appendix 3: B2 Level Speaking Assessment Scale	186
Appendix 4: Benchmarks and Justifications for DVD Sample 8.1	188
Appendix 5: Justifications for DVD Sample 8.2	189
Appendix 6: Benchmarks for DVD Sample 8.2	190
Appendix 7: Examples of Candidate Language for Interpreting the Speaking Assessment Scale	191
Appendix 8: Justifications for DVD Sample 8.3	192
Appendix 9: Benchmarks for DVD Sample 8.3	193
Appendix 10: Benchmarks and Justifications for DVD Sample 8.4	194
Appendix 11: Benchmarks for DVD Sample 8.5	195
Appendix 12: Justifications for DVD Sample 8.5	196
Appendix 13: List of reference books and recommended readings	197
Appendix 14: Contents of DVD	198

To János, Ágika and Zsófi

SERIES EDITOR'S PREFACE

The book is the second in the **Into Europe** series. The series in general is aimed at both teachers and students who plan to take an examination in English, be it a school-leaving examination, some other type of national or regional examination, or an international examination. Hopefully that examination will be a recognised examination which is based on international standards of quality, and which relates to common European levels – those of the Council of Europe.

However, unlike the first book in the series (*Reading and Use of English*) this book is especially aimed at **teachers** who are preparing their students for English examinations, or who may themselves have to design and conduct oral examinations in English. Assessing a learner's ability to speak a foreign language is a complicated and difficult task. Not only must the teacher know what tasks to set students when testing their speaking ability – what the features of good tasks are, what mistakes to avoid when designing oral tasks – but the teacher must also know how to assess the students' performance as fairly as possible. It is often said that testing speaking is a subjective matter and in a sense this is true and inevitable. But it does not have to be unreliable or unprofessional, and teachers can learn how to improve their ability to design tasks as well as their ability to judge performances more reliably. This book will help all teachers who feel the need to do this.

The authors of this book have long experience of teaching and assessing English. Moreover, as part of a British Council Project they have for the past six years and more been actively involved in designing speaking tasks, in piloting those tasks, and in devising appropriate procedures for the assessment of students' performances. They are the authors of a series of courses aimed at making teachers more aware of what is involved in assessing speaking, and they have developed, piloted and delivered highly successful in-service training courses to help teachers become more professional interlocutors and assessors. In Part Three of this book, those courses are described in more detail.

The British Council-funded Project was conducted under an agreement with the Hungarian Ministry of Education, through its agency OKI (the National Institute of Education). The task of the Project was to produce test specifications, guidelines for item writers and test tasks for the reform of the Hungarian School-leaving English Examination. The test tasks produced (Reading, Writing, Listening, Use of English and Speaking) were tested on large samples of students similar to those who would take school-leaving examinations in the future. The Project also trained raters of students' spoken performance, and developed in-service training courses for teachers of English, to help them become aware of

the demands of modern European examinations, and how best to prepare their students for such examinations.

It is in order to support teachers that the British Council has decided to publish the speaking tasks that were developed, as well as videos of students performing on those tasks. Building on the authors' experience, and incorporating their expertise and advice, this Handbook for Speaking is thus an invaluable resource for preparing for modern English oral examinations.

ACKNOWLEDGEMENTS

We would like to express our heartfelt thanks to Charles Alderson, who has been the consultant for the Hungarian Examinations Teacher Support Project since its inception and without whose inspiration, unfailing encouragement and editorial support this book would never have been completed. We are convinced that his uncompromising professionalism shown in test development and examination reform in Hungary has set a great example for us and all the other Project members to follow. We also wish to express our gratitude to Edit Nagy, the Project Manager, who was the originator of the British Council's support to Examination Reform in Hungary. Without Edit the Project would never have started and would never have achieved what it has. We thank her for her dedication to examination reform and her endeavours in ensuring high professional standards in the project work. This book would never have been conceived without Charles and Edit - Thank you!

Below we list all those individuals whom we wish to thank for writing items, attending training courses, taking part in Editing Committee meetings, designing teacher-support materials and courses, benchmarking and standard setting, and participating in all the other diverse tasks in examination construction. We owe you all a great debt of thanks.

We wish to thank all the Hungarian secondary school students who appeared on the DVD for agreeing to be videoed during their oral examinations as well as their English teachers and the school headmasters for their assistance and permission to set up the pilot speaking examinations. The recordings were made in the following schools in Hungary: Babits Mihály Gimnázium (Budapest); Berzsenyi Dániel Gimnázium (Budapest); Deák Ferenc Kéttannyelvű Gimnázium (Szeged); Gábor Dénes Műszaki Szakközépiskola (Szeged); Gárdonyi Géza Ciszterci Gimnázium (Eger); JATE Ságvári Endre Gyakorló Gimnázium (Szeged); KLTE Gyakorló Gimnázium (Debrecen); Krúdy Gyula Kereskedelmi és Vendéglátóipari Szakközépiskola és Szakiskola (Szeged); Szilágyi Erzsébet Gimnázium (Eger); Táncsics Mihály Gimnázium (Budapest); Teleki Blanka Gimnázium (Székesfehérvár).

We are very happy to acknowledge the support of KÁOKSZI, currently responsible for examination reform and its implementation, its former Director, Sarolta Igaz and most especially of Krisztina Szollás, who promoted the cause of quality examination reform when we seemed to have more opponents than allies.

We are also extremely grateful to the British Council for its unfaltering support over the years, especially the support of Directors Paul Dick, John Richards and

Jim McGrath, and their able Assistant Directors Ian Marvin, Peter Brown, Nigel Bellingham and Paul Clementson. We have counted on and benefited enormously from your support in good times and in bad.

We also acknowledge gratefully the support of our consultants, listed below, without whose expertise, experience and encouragement, we would not have got as far as we have.

Without the enthusiastic participation of countless secondary school teachers and their principals and students, we would not have been able to pilot and improve the test tasks: to you, we owe a great deal.

We wish to thank Dezső Gregus of the DZ Studio, the video and DVD production manager, whose dedication and flexibility have always been highly appreciated. Thanks to his professionalism, over the years since 1998 the Project has accumulated a large pool of high-quality video-recorded speaking examinations, which provided the basis for the DVD compilation.

And finally to our editors Béla Antal and Gábor Hingyi, and our publishers, thank you for your input, support and encouragement. We are privileged to have had the support of the Teleki Foundation, its manager Béla Barabás, and his assistant Viktória Csóra. We hope you are happy with the results.

And to you, the reader, thank you for using this book and we hope you enjoy and benefit from the results.

PEOPLE AND INSTITUTIONS

British Council Project Manager

Nagy Edit

KÁOKSZI Project Manager

Szollás Krisztina

Editing and layout

Antal Béla and Hingyi Gábor (??)

Video and DVD production

Gregus Dezső (DZ Studio, Szeged)

Item writers

Ábrahám Károlyné, Böhm József, Bukta Katalin, Csolákné Bognár Judit, Fehérváryné Horváth Katalin, Gál Ildikó, Grezsu Katalin, Gróf Szilvia, Hardiené Molnár Erika, Hegyközi Zsuzsanna, Lomniczi Ágnes, Magyar Miklósné, Margittay Livia, Nyirő Zsuzsanna, Sándor Ernőné, Schultheisz Olga, Sulyok Andrea, Szabó Kinga, Weltler Csilla

Team members developing the assessor and interlocutor training materials

Gál Ildikó, Schultheisz Olga, Sulyok Andrea, Szabó Kinga, Bukta Katalin, Böhm József, Hardiené Molnár Erika

Other item writers and teachers who took part in item production, pilot examinations and benchmarking, or made invaluable comments

Barta Éva, Berke Ildikó, Czeplédi Csaba, Vidáné Czövek Cecilia, Cser Roxane, Cseresznyés Mária, Dirner Ágota, Dóczi Brigitta, Horváth József, Kiss Istvánné, Kissné Gulyás Judit, Luszti Ágnes, Martsa Sándorné, Nábrádi Katalin, Némethné Hock Ildikó, Nikolov Marianne, Sándor Éva, Pércsich Richárd, Philip Glover, Szabó Gábor, Szalai Judit, Szollás Krisztina, Tóth Ildikó, Weisz György

Colleagues providing technical assistance for the Project

Himmer Éva, Révész Rita

Editing Committee

Cseresznyés Mária, Dávid Gergely, Fekete Hajnal, Gróf Szilvia, Kissné Gulyás Judit, Nikolov Marianne, Nyirő Zsuzsanna, Philip Glover, Szollás Krisztina

OKI English Team leaders

Vándor Judit, 1996–1999

Öveges Enikő, 1999–2000

Project consultants

Richard West (University of Manchester)

Jane Andrews (University of Manchester)

John McGovern (Lancaster University)

Dianne Wall (Lancaster University)

Jayanti Banerjee (Lancaster University)

Caroline Clapham (Lancaster University)

Nick Saville (Cambridge ESOL)

Nick Kenny (Cambridge ESOL)

Lucrecia Luque (Cambridge ESOL)

Annette Capel (Cambridge ESOL)

Hugh Gordon (The Scottish Qualifications Authority)

John Francis (The Associated Examining Board)

Vita Kalnberzina (Latvia)

Ülle Türk (Estonia)

Zita Mazuoliene (Lithuania)

Stase Skapiene (Lithuania)

SCHOOLS TAKING PART IN THE PILOTING OF TASKS FOR THE HUNGARIAN EXAMINATIONS REFORM PROJECT

Ady Endre Gimnázium, Debrecen; Apáczai Csere János Gimnázium és Szakközépiskola, Pécs; Babits Mihály Gimnázium, Budapest; Batthyányi Lajos Gimnázium, Nagykanizsa; Bencés Gimnázium, Pannonhalma; Berze Nagy János Gimnázium, Gyöngyös; Berzsenyi Dániel Gimnázium, Budapest; Bethlen Gábor Reformá-

tus Gimnázium, Hódmezővásárhely; Bocskai István Gimnázium és Közgazdasági Szakközépiskola, Szerencs; Boronkay György Műszaki Szakiskola és Gimnázium, Vác; Bolyai János Gimnázium, Salgótarján; Ciszterci Rend Nagy Lajos Gimnáziuma, Pécs; Deák Ferenc Kéttannyelvű Gimnázium, Szeged; Debreceni Egyetem Kossuth Lajos Gyakorló Gimnáziuma, Debrecen; Dobó István Gimnázium, Eger; Dobó Katalin Gimnázium, Esztergom; ELTE Radnóti Miklós Gyakorló Gimnáziuma, Budapest; ELTE Trefort Ágoston Gyakorló Gimnáziuma, Budapest; Eötvös József Gimnázium, Tata; Fazekas Mihály Fővárosi Gyakorló Gimnázium, Budapest; Gábor Áron Gimnázium, Karcag; Gábor Dénes Gimnázium, Műszaki Szakközépiskola és Kollégium, Szeged; Gárdonyi Géza Gimnázium, Eger; Herman Ottó Gimnázium, Miskolc; Hunfalvy János Gyakorló Kéttannyelvű Közgazdasági és Külkereskedelmi Szakközépiskola, Budapest; I. István Kereskedelmi és Közgazdasági Szakközépiskola, Székesfehérvár; Janus Pannonius Gimnázium, Pécs; JATE Ságvári Endre Gyakorló Gimnázium, Szeged; Karinthy Frigyes Kéttannyelvű Gimnázium, Budapest; Kazinczy Ferenc Gimnázium, Győr; Kereskedelmi és Vendéglátóipari Szakközépiskola, Eger; Kölcsey Ferenc Gimnázium, Zalaegerszeg; Kossuth Lajos Gimnázium, Budapest; Krúdy Gyula Gimnázium, Győr; Krúdy Gyula Gimnázium, Nyíregyháza; Krúdy Gyula Kereskedelmi, Vendéglátóipari Szakközépiskola és Szakiskola, Szeged; Lauder Javne Zsidó Községi Iskola, Budapest; Lengyel Gyula Kereskedelmi Szakközépiskola, Budapest; Leőwey Klára Gimnázium, Pécs; Madách Imre Gimnázium, Vác; Mecsekaljai Oktatási és Sportközpont, Pécs; Mikszáth Kálmán Gimnázium, Pásztó; Móricz Zsigmond Gimnázium, Budapest; Németh László Gimnázium, Budapest; Neumann János Közgazdasági Szakközépiskola és Gimnázium, Eger; Neumann János Informatikai Szakközépiskola, Budapest; Óbudai Gimnázium, Budapest; Pásztorvölgyi Gimnázium, Eger; Pesti Barnabás Élelmiszeripari Szakközépiskola, Szakmunkásképző és Gimnázium, Budapest; Petrik Lajos Vegyipari Szakközépiskola, Budapest; Pécsi Művészeti Szakközépiskola, Pécs; Premontrei Szent Norbert Gimnázium, Gödöllő; PTE Babits Mihály Gyakorló Gimnázium, Pécs; Radnóti Miklós Kísérleti Gimnázium, Szeged; Révai Miklós Gimnázium, Győr; Sancta Maria Leánygimnázium, Eger; Sipkay Barna Kereskedelmi és Vendéglátóipari Szakközépiskola, Nyíregyháza; Sport és Angoltagozatos Gimnázium, Budapest; Szent István Gimnázium, Budapest; Szilády Áron Gimnázium, Kiskunhalas; Szilágyi Erzsébet Gimnázium, Eger; Talentum Gimnázium, Tata; Táncsics Mihály Gimnázium, Kaposvár; Táncsics Mihály Közgazdasági Szakközépiskola, Salgótarján; Teleki Blanka Gimnázium, Székesfehérvár; Terézvárosi Kereskedelmi Szakközépiskola, Budapest; Toldy Ferenc Gimnázium, Nyíregyháza; Városmajori Gimnázium, Budapest; Vásárhelyi Pál Kereskedelmi Szakközépiskola, Budapest; Veres Péter Gimnázium, Budapest; Vörösmarty Mihály Gimnázium, Érd; 408.sz. Szakmunkásképző, Zalaszentgrót; Wigner Jenő Műszaki Szakközépiskola, Eger

Györgyi would like to express her heartfelt thanks and gratefulness to her husband, János for his endless love, support, encouragement and patience throughout the years which have led to the publication of this book. Without him

she would never have been able to contribute to the examination reform in Hungary and co-author this publication.

She is also extremely grateful to Richard West of the University of Manchester, who gave her professional guidance and support during her studies at the University of Manchester.

Györgyi also would like to thank her former colleague and friend, Len Rix of the Manchester Grammar School, and her former English Assistants, Peter Neal (1993), Jonathan Prag (1994), Stewart Templar (1994-1995), Robert Neal (1995), Tom Walker (1996), Elliot Shaw (1996-1997) and Andrew Duncan Logan (1997) for their enthusiastic interest in ELT issues in Hungary, which first inspired her to initiate school-based research projects in assessing speaking performances, and to get involved in the examination reform project.

PART ONE

GENERAL INTRODUCTION

Introduction

The book is in three main parts. In Part One, we discuss general issues related to the assessment of speaking ability in line with modern European standards. In Chapter 1, we focus on the main features of modern English speaking examinations: skills to be assessed, task types, levels of achievements according to common European standards and quality control issues such as standardisation, benchmarking and training of examiners. In Chapter 2, we review the main variables that may influence test takers' oral performance in order to raise test developers' awareness of their positive or negative impact. Chapter 2 also discusses the individual and the paired mode of oral performance assessment.

Since test developers have full control over the design of examination tasks, in Part Two we discuss features of good and bad speaking tasks by providing examples of different task types developed within the Hungarian Examinations Reform Teacher Support Project. We present guidelines for item writers who wish to design interview questions (Chapter 3), picture-based individual long turn tasks (Chapter 4), discussion activities for the individual and paired mode (Chapter 5) and role plays for the individual and the paired mode (Chapter 6).

Part Three deals with how interlocutors and assessors can be trained in order to standardise speaking examinations. Chapters 7 and 8 describe the interlocutor and assessor training model developed by the Hungarian Examinations Reform Teacher Support Project. In Chapter 7, sample training activities such as simulation/role play tasks are presented in order to highlight how future interlocutors can gain the necessary confidence in their role. The demands of the interlocutor's job are further highlighted through DVD performances that display both standard and non-standard interlocutor behaviour. Similarly to the training of interlocutors, in Chapter 8 sample activities and guidelines are presented through which future assessors can be provided with hands-on experience in assessing speaking performances both in language classes and in examination situations.

Uniquely, this book illustrates different options in the assessment of speaking as it is accompanied by an invaluable resource of oral performance samples on DVD, which features Hungarian learners of English at a wide range of proficiency levels. The DVD includes carefully selected performances in order to demonstrate

- a number of speaking tasks in action (both good and bad tasks)
- how picture-based elicitation tasks work with and without question prompts for the interlocutor;

- whether it makes a difference if the interlocutor's contributions (e.g. questions and comments) are scripted/prescribed or not;
- how candidates' performance differs on individual and paired discussion tasks;
- standard and non-standard interlocutor behaviour;
- sample benchmarked performances.

Finally, we provide recommendations for good practice by discussing how the principles described for assessing speaking can be applied in classroom assessment contexts and how ongoing quality assurance can be provided in order to adhere to modern European standards. The washback effect of modern European speaking exams is also considered, as teachers need to understand that high quality exams should have a positive impact on the quality of English language teaching. The hope is that learners will practise tasks that require them to use English in life-like situations as part of their exam preparation. And if they learn to cope with such tasks, they will be guaranteed to succeed in using English with real people outside the language classroom, as well as in the speaking examination itself.

Chapter 1

To the Teacher

This handbook is intended to help teachers who have to administer or design oral tests to develop and conduct modern English oral examinations. We discuss what is involved in the assessment of speaking ability: what it is we want to measure, what is likely to influence candidates' performances, what task types and examination modes can be used to elicit language for assessment purposes, and how to prepare examiners to conduct exams and assess candidates' performances reliably.

The most important feature of modern English examinations is that they should present candidates with tasks that resemble as closely as possible what people do with the language in real life. Modern speaking examinations, therefore, focus upon assessing a learner's ability to use the language in lifelike situations, in which they have to perform a variety of language functions. The candidate's performance has to be spontaneous since in real life we rarely have the chance to prepare for what we want to say. In real life, when two people are engaged in a conversation, they take turns to initiate and to respond, they may decide to change the topic of the conversation suddenly or they may interrupt each other in order to take a turn. As a result, conversations may take very different directions depending on the speakers' intentions. In short, the speaker's contributions may be fairly unpredictable in real-life conversations. In contrast, in many traditional exam settings, it is customary to allow candidates to make notes for up to 15 minutes on the topic on which they will be examined, before they are actually required to speak. Modern English speaking examinations, on the other hand, try to create circumstances for candidates in which they can convey messages spontaneously. To achieve this, examination tasks engage test takers in language performance in such a way that their contributions are not rehearsed or prepared in advance. This can be ensured, for example, by using tasks that are related to carefully designed contexts, which may be provided by a role-play situation or a set of pictures, for example, and by training examiners to follow specific guidelines for behaviour such as intervening when it is obvious that the candidate is reciting a rehearsed text. The task should discourage candidates from reciting memorised texts as much as possible because a rehearsed performance cannot provide sufficient evidence that the learner is able to participate fluently in real communicative events.

While it is obvious that candidates should be able to use the language spontaneously in lifelike situations, it is equally important that their language ability needs to be assessed in different speech contexts. A proficient speaker is capable of taking part in various different activities, which can be grouped into two main categories: productive activities (one-way information flow activities) and interactive activities (two-way information flow activities). In the former group of activities the language user produces an oral text that is received by an audience of one or more listeners. For instance, the language user may be asked to speak spontaneously while addressing a specific audience in order to give

instructions or information. However, in other cases s/he may be asked to speak from notes, or from visual aids. In interactive activities, the language user acts alternately as speaker and then as listener with one or more partners in order to realise a specific communicative goal.

In order to ensure the validity of any speaking examination, the speaking activities should be carefully selected, taking into account the language needs of the target population and the purpose of the examination. Speaking activities must always be relevant for the candidates who take the given exam. For example, if the speaking examination is intended for young adults who have no specific purpose for the use of English, it is highly unlikely that any of the following activities would be valid language use activities: giving speeches at public meetings, giving sales presentations or negotiating a business transaction. All these language use activities are unusual as it is only a specific group of language users (business people) who may be required to perform them in real life. In modern speaking exams that aim to assess candidates' overall speaking ability in English for no specific purpose, the following one-way information flow tasks are frequently employed: describing experiences, events, activities, habits, plans, people; comparing and contrasting pictures; sequencing activities, events, pictures; giving instructions or directions. Interactive or two-way information flow activities, on the other hand, include transactions to obtain goods and services, casual conversation, informal or formal discussion, interview, etc.

Modern European speaking examinations should provide candidates with appropriate opportunities to demonstrate that they can communicate in the target language in order to convey messages and realise their communicative goals, and in order to make themselves understood and to understand others. Naturally, learners at different levels of proficiency will perform different speaking activities with more or less accuracy and fluency. Learners at low levels of proficiency cannot be expected to perform the same range of language activities as learners at higher levels. In order to ensure that language proficiency is understood in similar terms and achievements can be compared in the European context, the Council of Europe has devised a common framework for teaching and assessment, which is called 'The Common European Framework of Reference for Languages', or CEF for short (Council of Europe, 2001). When assessing learners' oral abilities, examination tasks should be designed in such a way that they are closely related to the oral production and interactive activities that represent specific levels of language proficiency within the Council of Europe Framework.

The CEF scale has six major levels, which start with "beginner" or "false beginner" and go up to "highly advanced". The levels are labelled with letters and numbers since what is considered "false beginner" or "highly advanced" varies greatly in different contexts. In the Framework, the lowest level is marked as A1 and the highest level is labelled as C2. Each level should be taken to include the levels below it on the scale. The descriptors report typical or likely behaviours of learners at any given level by stating what the learner *can do* rather than what s/he cannot do.

For overall spoken production, the CEF levels are specified in the following way:

Table 1 CEF Overall Oral Production

	OVERALL ORAL PRODUCTION
C2	Can produce clear, smoothly flowing well-structured speech with an effective logical structure which helps the recipient to notice and remember significant points.
C1	Can give clear, detailed descriptions and presentations on complex subjects, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.
B2	Can give clear, systematically developed descriptions and presentations, with appropriate highlighting of significant points, and relevant supporting detail.
	Can give clear, detailed descriptions and presentations on a wide range of subjects related to his/her field of interest, expanding and supporting ideas with subsidiary points and relevant examples.
B1	Can reasonably fluently sustain a straightforward description of one of a variety of subjects within his/her field of interest, presenting it as a linear sequence of points.
A2	Can give a simple description or presentation of people, living or working conditions, daily routines, likes/dislikes, etc. as a short series of simple phrases and sentences linked into a list.
A1	Can produce simple mainly isolated phrases about people and places.

(Council of Europe, 2001: page 58)

Note that the B2 level in the middle of the scale has a subdivision. The descriptor that follows immediately after B1 is the criterion level. The descriptor placed above it defines a proficiency level that is significantly higher than the criterion level but does not achieve the next main level, which is C1.

Spoken interaction is also described at the six main levels of the Framework. In addition to the scale for overall spoken interaction, which has a subdivision for A2, B1 and B2 (see below), there are sub-scales available for describing learners' performance in the following areas (the page numbers in parentheses refer to the English version published by Cambridge University press in 2001):

- understanding a native speaker interlocutor (p. 75);
- conversation (p. 76);
- informal discussion with friends (p. 77);
- formal discussion and meetings (p. 78);
- goal-oriented co-operation such as repairing a car, discussing a document, organising an event, etc. (p. 79);
- transactions to obtain goods and services (p. 80);
- information exchange (p. 81);
- interviewing and being interviewed (p. 82).

These scales can provide detailed guidelines for test developers because the level descriptors for the different language use contexts clearly state what a learner is expected to be able to do.

Table 2 CEF Overall Spoken Interaction

	OVERALL SPOKEN INTERACTION
C2	<i>Has a good command of idiomatic expressions and colloquialisms with awareness of connotative levels of meaning. Can convey finer shades of meaning precisely by using, with reasonable accuracy, a wide range of modification devices. Can backtrack and restructure around a difficulty so smoothly the interlocutor is hardly aware of it.</i>
C1	<i>Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language.</i>
B2	<i>Can use the language fluently, accurately and effectively on a wide range of general, academic, vocational or leisure topics, marking clearly the relationships between ideas. Can communicate spontaneously with good grammatical control without much sign of having to restrict what he/she wants to say, adopting a level of formality appropriate to the circumstances.</i> <i>Can interact with a degree of fluency and spontaneity that makes regular interaction, and sustained relationships with native speakers quite possible without imposing strain on either party. Can highlight the personal significance of events and experiences, account for and sustain views clearly by providing relevant explanations and arguments.</i>
B1	<i>Can communicate with some confidence on familiar routine and non-routine matters related to his/her interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thoughts on more abstract, cultural topics such as films, books, music etc.</i> <i>Can exploit a wide range of simple language to deal with most situations likely to arise whilst travelling. Can enter unprepared into conversation on familiar topics, express personal opinions and exchange information on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).</i>
A2	<i>Can interact with reasonable ease in structured situations and short conversations, provided the other person helps if necessary. Can manage simple, routine exchanges without undue effort; can ask and answer questions and exchange ideas and information on familiar topics in predictable everyday situations.</i> <i>Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters to do with work and free time. Can handle very short social exchanges but is rarely able to understand enough to keep conversation going of his/her own accord.</i>
A1	<i>Can interact in a simple way but communication is totally dependent on repetition at a slower rate of speech, rephrasing and repair. Can ask and answer simple questions, initiate and respond to simple statements in areas of immediate need or on very familiar topics.</i>

(Council of Europe, 2001: page 74)

The level descriptors of CEF are also available for learners to assess their own abilities. Sample self-assessment statements for speaking, covering the full range of performance levels from A1 to C2, together with a separate scale for overall performance, are included in Appendix 1.

Modern European examinations are linked to the CEF levels as different levels of achievement can only be interpreted across Europe if the intended level for an examination is clearly defined. International examinations such as the Cambridge ESOL main suite exams (KET, PET, FCE, CAE, CPE) are aimed at a specific CEF level each. The Preliminary English Test (PET) is aimed at B1 while the Certificate of Advanced English (CAE) is at C1. However, it is not only international examinations that should define their intended levels in relation to the Framework but all high-stakes modern language examinations should do so. The planned Hungarian school-leaving examination is, according to current documents, said to be at A2 and B1 for the so-called Intermediate level, and at B2 for the so-called Advanced level. However, modern European examinations are usually aimed at only one CEF level at a time because learners' language abilities can be measured more reliably in this way. The reliability of an examination is strengthened if candidates have to perform on tasks that are indeed able to elicit the desired features of performance that characterise the given CEF level.

It is important to note, however, that a speaking examination can only be claimed to measure language ability at a particular CEF level if there is sufficient evidence that the examination has gone through a rigorous programme of quality control. First of all, examination tasks need to be moderated by an editing committee (a group of experienced item writers), then piloted with real students and finally revised in the light of the results of the pilot exams. It usually takes several rounds of careful editing, piloting and revision before a task is given its final shape. In this book all the tasks that are claimed to work well have been piloted and revised where necessary. Thus, we are confident that the tasks recommended as good examples are of high quality, and they are fully in line with the highest possible (European) standards of item production.

However, even after extensive piloting and revision we cannot claim that a task does indeed measure a specific level and therefore, a procedure called "standard setting" has to be employed to establish what level the task is at. During this procedure experts are asked to examine in detail the relationship between the task and the Common European Framework and to decide which (minimum) level a learner must be at if he or she is able to complete the task successfully. Standard setting is normally accompanied by another procedure called "benchmarking", during which testing experts view videos in order to examine how candidates perform and reach a consensus about the level of particular candidate performances.

It is important that trainers for modern European speaking examinations should be thoroughly familiar with benchmarking procedures and be capable of conducting benchmarking workshops. Devising and conducting benchmarking sessions is a skilled task which requires considerable experience and on-the-job training. The Hungarian Examinations Reform Teacher Support Project of the

British Council devised, piloted and implemented a set of benchmarking procedures between 2001 and 2003. The benchmarking experience of the Hungarian Examinations Reform Teacher Support Project and the materials developed by them could serve as good examples for facilitating local standard setting processes.

Modern European speaking exams not only pay attention to what they measure but how they measure it. Modern European speaking examinations are administered with at least two examiners (an interlocutor and an assessor) present in the examination room. It is essential to standardize the way the interlocutor (the examiner who is responsible for conducting the exam) elicits language from the candidate. It is equally important to standardise the way the assessor (a silent observer at the exam) judges or marks the test taker's performance. These two roles cannot be managed simultaneously as the interlocutor has to make sure that the candidate produces a sufficient amount of language for assessment purposes and that the candidate responds to the given task. Judging the quality of the performance requires an observer, an independent judge with a completely different perspective on the exam. The assessor is silent and seemingly passive throughout the exam and therefore s/he can make more reliable judgements about the quality of the candidate's performance. Since the assessor's judgements have to be made in accordance with clearly specified criteria for assessment, the assessor must consult the assessment scale during the exam in order to arrive at specific ratings for different aspects of the performance. And when performance is assessed task by task in an exam, it becomes even more demanding to give scores since different tasks may require the use of different assessment scales. Clearly, in such speaking examinations it is impossible to perform the interlocutor and assessor's jobs simultaneously. If live marking is not possible, i.e. there is only one examiner administering the speaking exam, candidates' performances must be video- or audio-recorded so that they can be assessed later in examination centres.

Standardisation is one of the most important aspects of quality control in modern European language assessment. It is only by establishing and employing standard, commonly accepted procedures for conducting exams and assessing performances that oral test scores can be compared across different examination bodies and the examination result can be reliable and valid. It is only by rigorous procedures for standardisation that we can ensure that a specific score means the same – it represents the same achievement – no matter where the candidate takes the test.

In order to ensure a high degree of reliability of assessment and score comparability, modern speaking examinations should be standardised, i.e. employ common standards in all possible aspects of the elicitation and assessment procedures. Without effective training of interlocutors and assessors, an examination score cannot be a valid indicator of a learner's oral proficiency. For example, an interlocutor's non-standard performance may negatively influence the candidate's performance, or the assessor may interpret candidates' performances in his/her own terms instead of evaluating them in relation to clearly defined level descriptors, (which are usually illustrated in assessor training courses through

benchmarked performances). Benchmarked performances provide the means through which the reliability of subjective marking is enhanced and ensured. Without them it is impossible to establish the standard to which assessors should be marking. Thus, benchmarking is an essential component of any quality control system for oral language examinations.

In order to ensure the reliability and validity of modern European speaking examinations, the following is recommended as a means of standardising the exam:

- providing an Interlocutor Frame, which includes all the instructions and suggested question prompts the interlocutor will need in order to conduct the exam;
- providing guidelines for interlocutor behaviour;
- training interlocutors to be able to apply the Interlocutor Frame and internalize the guidelines;
- employing analytic assessment scales that have been piloted and revised;
- training assessors to apply the scales and interpret the descriptors in the intended way;
- monitoring interlocutors and assessors.

This book will highlight each step listed above. The language elicitation and assessment procedures, examiner training materials and speaking assessment scales that are used as examples have been carefully designed, piloted and revised several times by the Hungarian Examinations Reform Teacher Support Project. This book is intended primarily to share the insights and materials that the Project has accumulated in the area of assessing speaking skills.

Chapter 2:

What May Influence Candidates' Oral Performance?

It is difficult to assess a candidate's oral performance reliably. A learner may perform differently on different tasks, or with different interlocutors, or if they take an exam individually or in pairs. The assessment of oral language ability is clearly a complex matter as there are a number of variables that may positively or negatively influence test takers' performance. In order to elicit the candidates' "best performance", test designers must consider carefully the potential impact of all the possible factors that may shape candidates' performances and their assessment. This chapter is intended to raise teachers' awareness of the potential sources of variation in test takers' oral performance.

When developing speaking exams, test designers should provide the candidates with circumstances in which their performance will be primarily influenced by their language ability and not by irrelevant aspects of the testing situation. In addition to language ability, candidates' performance may also be affected by their personal characteristics, their knowledge of the world, their emotions and anxiety level and their ability to plan and implement their plans (their "strategic competence"). Let us consider the potential impact of each of these variables.

Personal characteristics include the test taker's age, sex, nationality, native language, level and type of general education, type and amount of preparation or prior experience with a given test. When designing examination tasks, test developers need to consider carefully how the individual characteristics of the test taker for whom the test is intended will interact with the given task. For example, if the test is aimed at a population of both sexes, none of the tasks should favour either sex or any particular age group. A role play task in which the candidate has to ask for a perm at the hairdresser's may embarrass male candidates, especially if they have little or no hair. It can be equally embarrassing for teenagers to assume unfamiliar roles in a role play where they have to pretend to be the manager of a multi-national company who is discussing the schedule of a business trip with his secretary (played by the interlocutor), or pretending to be the parent of a 16-year old child discussing whether the child may go on a hiking trip abroad (where the child's role is taken by the interlocutor). It is very important to take candidates' life experiences and real life needs for the language into account when choosing topics and roles.

If tasks require specific *knowledge of the world*, candidates' performances may also be negatively affected. For example, if the topic of discussion is genetically modified food, candidates might not feel at ease to contribute sufficiently to the interaction as the topic is heavily dependent on background knowledge. Unless candidates know something about this topic and so have the necessary background information, the conversation is likely to come to a halt fairly quickly. We must also bear in mind that knowledge about the target language culture is part of candidates' general knowledge of the world, and so it should be treated separately from language ability. When the purpose of the test is to measure how fluently and

accurately candidates manage to convey messages in English, it is quite unimportant whether candidates know the famous sights of London, or the names of British political parties. A test taker may be a perfectly fluent language user without knowing these facts. We do not wish to suggest, however, that teaching about the target language culture is a waste of time. Learners' motivation for learning can be greatly enhanced by making them familiar with different traditions, beliefs, attitudes, and social conventions of the target language culture (also referred to as 'sociocultural knowledge'). In language testing, however, candidates' responses to a task should never depend on knowledge of facts and figures, or how much they know about the topic of a given task.

Another important aspect of candidates' individual characteristics is their *emotional response* to the test task (known as their 'affective schemata'). Affective schemata provide the basis on which language users assess a language use task and its setting in relation to past emotional experiences. Some authors suggest that controversial topics may stimulate some test takers to perform at a high level because they may have strong feelings about a particular viewpoint. This is very important when it comes to choosing the topic of discussion activities. Different topics will evoke different associations and emotional responses from candidates. However, there are certain topics – such as divorce, death, natural disasters, diseases, personal income – that are likely to affect some of the candidates in a negative way. For example, if the candidate's parents have just divorced, or one of his/her close relatives is seriously ill. When designing test tasks, we should consider very carefully every topic in terms of how it will affect test takers. In general, potentially distressing topics should be avoided altogether in order to establish the most favourable circumstances for all the candidates and to help them perform to the best of their abilities.

The mediating role between candidates' individual characteristics and the test task and setting is played by *strategic competence*, which is a general ability enabling an individual to make the most effective use of their skills in carrying out a task. This ability is made up of the following strategies:

- planning (e.g. judging what can be presupposed or taken as given);
- execution (e.g. taking and releasing the floor, co-operating);
- evaluation (e.g. monitoring the interaction when the speaker uses facial expressions or body language);
- communication repair (e.g. asking for and giving clarification).

In spoken interaction language users make use of these strategies in order to successfully participate in the conversation. As a result, they create meaning(s) together with their partner(s).

The language assessment procedure, however, is influenced not only by the candidates' individual characteristics, their strategic competence and the characteristics of the task that are responsible for the performance. It is just as important to consider the other participants in the interpretation of performance: the interlocutor whose job is to facilitate the elicitation of the performance and the

assessor who marks the performance in the light of his/her interpretation of the pre-specified assessment criteria.

The following diagram shows a visual representation of the possible interactions between the variables that may influence the interpretation of candidates' oral test performance.

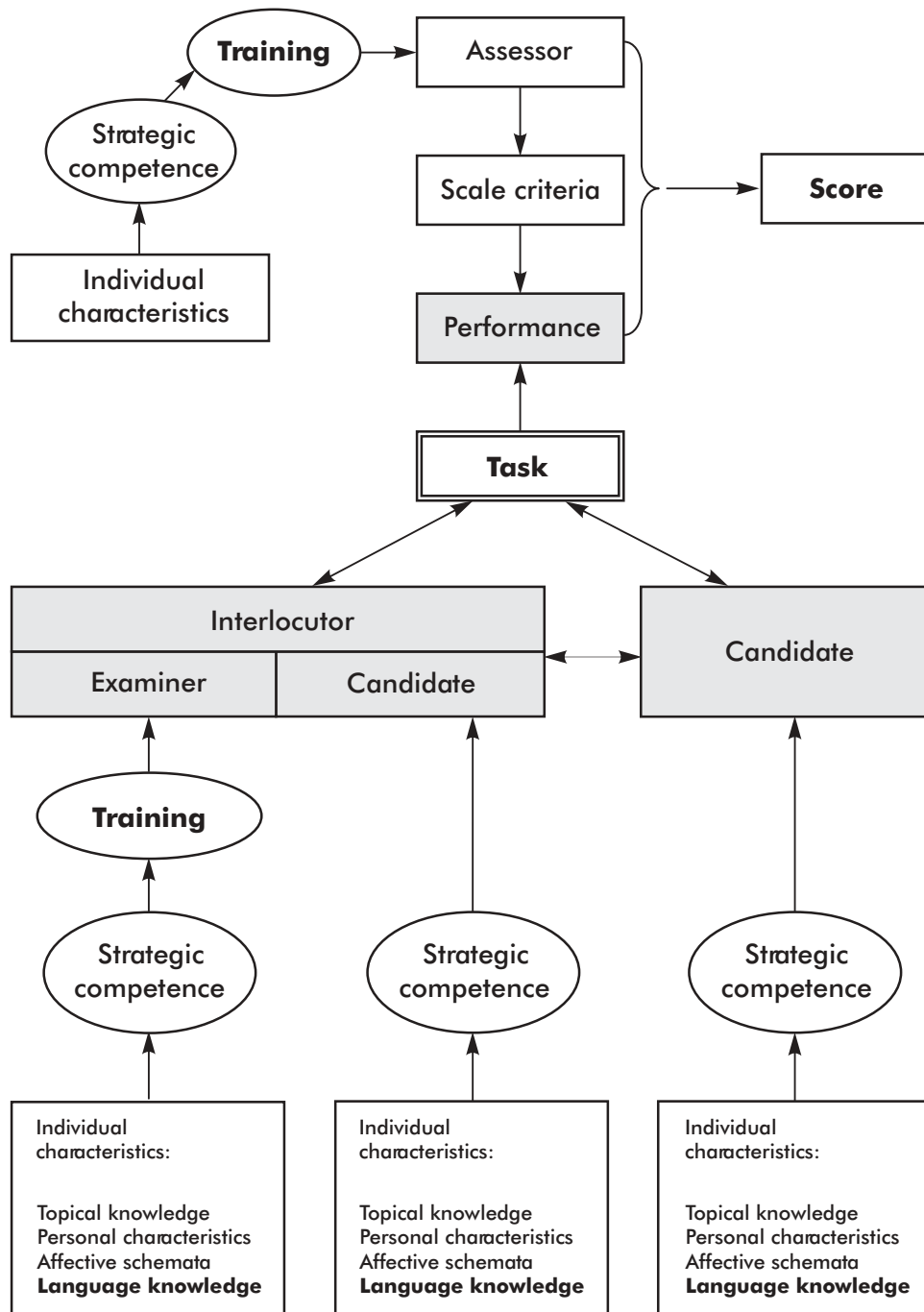


Figure 1. An extended model¹ of oral performance testing

¹ Based on Bachman & Palmer (1996) and McNamara (1996)

In the centre of the diagram above, there is a triangle, the three angles of which represent the key contributors in this speech event: the candidate, the interlocutor (who may be an examiner or a candidate) and the task. The double-headed arrows indicate that there is a two-way interaction between these key elements, the overall outcome of which is test performance. Performance is evaluated by a trained assessor, who uses scale criteria in order to arrive at a score, which is then taken as a measure of the candidate's language knowledge. As can be seen clearly in the diagram, there is no direct link between the score and the candidate's language knowledge. Therefore, it is important to stress that the score is only an indirect index of language knowledge and only partly a direct index of actual performance since specific assessor behaviour and assessment scales may systematically affect the scores candidates get.

The diagram above shows that the interlocutor may also be another candidate, which is the case in paired examinations. In paired-task performance, there may actually be three participants involved in the interaction since the examiner-interlocutor is still present, although his/her role is normally very limited. In the case of the group oral exam, where three or more candidates participate in a task, the number of participants engaged in the interaction is even bigger, which also means that the potential interaction between the individual variables of the participants is likely to be even more complex.

The evidence for candidates' language abilities can be positively influenced in one-to-one and paired/group oral exams in two ways: examiner training and task design. Examiner behaviour must be monitored and standardized through training in order to eliminate unwanted variation in test takers' performance. For example, in paired tasks the examiner can intervene and minimize unwanted variation in test takers' performance that may result from pairing up partners who fail to match each other for different reasons (e.g. level of proficiency in L2, personality or lack of familiarity with each other). The design of test tasks in oral proficiency testing is also very important. In fact, test developers have full control over this variable alone, whereas the other variables are likely to be much more difficult to control or monitor. The central role of tasks is underscored in Figure 1 by placing 'Task' in the middle of all the possible interactions. As the characteristics of test tasks are always likely to affect test performance, it is important to understand and control them. In other words, their effects cannot be eliminated. That explains why in Part Two we discuss different speaking tasks in detail.

Next, we will consider how the format of the test may influence performance. Since the one-to-one, live oral proficiency interview (OPI) has been regarded as a yardstick in oral proficiency testing, we will first discuss the potential impact of the examiner-interlocutor on test takers' performance in the individual mode, and then we will examine more closely the paired exam format.

The face-to-face oral proficiency interview

In the 1980s the validity of the oral proficiency interview (OPI) began to be questioned as it became more widespread as a testing method. Research has shown

that the oral interview test is sensitive to a change in the speech style ("interview" style vs. "reporting" speech style) and a change of the topic of discussion. Some authors point out that a serious drawback of the interview is related to the differences between the interviewer and the candidate in terms of their status and role in the interaction (the candidate speaks to a superior and may be unwilling to take the initiative). Others have criticized the interview format for not being able to replicate all the features of real life communication such as motivation, purpose and role appropriacy. A comparison of an oral proficiency interview and an informal conversation reveals significant differences between these two modes of interaction. The basic characteristics of conversation include unplannedness, and the potentially equal distribution of rights and duties in talk (e.g. initiating and closing topics). It is argued that in OPIs emphasis is put on successful elicitation of language and not on successful conversation. As a result, the interaction of the interview is controlled by the interviewer to a large extent. Thus, many authors question the validity of the OPI as an appropriate means of measuring conversational ability.

The three main differences between OPIs and ordinary conversation are as follows:

- The topical and turn-taking systems of OPIs are different from those of ordinary conversation.
- The goals of the participants are also different, and in OPIs there are specific constraints on participants' contributions (e.g. examiners have a pre-defined agenda).
- Participants in OPIs sometimes have a different understanding of what is going on (e.g. the candidate may misinterpret the examiner's intentions), which, if this happens, can be clarified, discussed or resolved in ordinary conversations.

As has been pointed out earlier, the examiner-interlocutor is responsible for making sure the candidate produces a sufficient amount of language for assessment purposes. However, research into oral test discourse has shown that there may be variation in the linguistic behaviour of the examiner since interlocutors may modify their speech when talking to candidates. For example, the interlocutor may rephrase questions to check comprehension, or repeat questions more slowly. Also, they may give cues to candidates on the topic of a question before it is asked, or give evaluative responses to a candidate answer. While most of these strategies seem to be quite useful as they can help candidates to remain in the conversation, there are some which are not so desirable in language assessment. For instance, when the interlocutor uses statements as question prompts (e.g. *So you can't drive*), s/he simply deprives candidates of demonstrating their language ability as only yes-no confirmations are required. Moreover, such strategies may not be consistently employed by all examiners. Such inconsistency in examiner behaviour raises two important issues. On the one hand, candidates will have unequal opportunities if the examiner fails to give the

same kind of support to each. On the other hand, the validity of scores may be negatively affected as the scores should be based on candidates' abilities rather than depend on examiner behaviour or support.

The examiner-interlocutor's individual characteristics are also likely to influence the candidate's performance. Although examiner behaviour can be controlled and monitored through training to a large extent, candidate performance may still be influenced by factors that are difficult to control. Research has shown that the examiner's L1 background or acquaintanceship with the candidate can influence candidate performance.

In a study that involved two interlocutors with different L1 backgrounds (an American and a Japanese interviewer), the interlocutor's culture had an important influence on the interviewing strategies used. After receiving similar training, the two interviewers conducted six interviews in their own mother tongue. (The interviews were intended to measure English as a second language and Japanese as a second language.) Although candidates received similar scores when interviewed by two different interlocutors, a study of the discourse of the interviews revealed substantial differences in terms of the use of display questions (i.e. questions to which the answer is known), overarticulation and lexical simplification. The Japanese interviewer employed significantly more of these features than the American examiner. These findings suggest that the interviewers took different paths to arrive at equivalent assessments. The Japanese style emphasized authority over the interviewee through attention to form, while the American style reflected attention to content.

The interlocutor's acquaintanceship with candidates may also influence the communication strategies used by the participants. One study explored how a familiar and an unfamiliar interlocutor conducted oral proficiency interviews with the same set of candidates. An analysis of the test discourse showed that candidates talking to the familiar examiner more often resorted to so-called 'help-requesting strategies' than when they were faced with an unfamiliar examiner-interlocutor. The unfamiliar interlocutor, on the other hand, was found to employ a balanced mixture of strategies by responding to requests for help as well as using strategies that focused on eliciting and clarifying messages. Thus, there seems to be some evidence that candidates' familiarity or lack of familiarity with the interlocutor may generate different types of oral discourse.

We argue that interlocutor variability related to interlocutor support or familiarity with the candidate could be minimized if the elicitation procedure were more controlled or structured. This can be achieved by standardizing the elicitation procedure with the help of an Interlocutor Frame, which clearly specifies the allowable contributions of the interlocutor (see Chapters 3 and 7 for more detail).

The paired exam

The validity of the paired oral exam has been researched considerably less in comparison with the interview although peer-to-peer interactions avoid the unequal power relations that characterise examiner-examinee interactions. Nowadays the format is widely used by international exams such as the main suite

exams of Cambridge ESOL, which introduced the paired format in the early 1990s because it seemed to have several advantages over the one-to-one interview format. For example, interaction in the paired format was found to be more varied than in the individual mode since the candidate's partner was not only the examiner but another candidate as well.

Testing candidates in pairs and small groups is motivated by five main reasons:

- dissatisfaction with the oral interview as the only test format to assess oral proficiency,
- search for new tasks that can elicit different patterns of interaction from those elicited by the interview format,
- desire to generate positive washback on teaching by encouraging more interaction between learners,
- to mirror good language teaching practice,
- time saving and cost reduction, as it seems to be less expensive to test candidates in groups.

However, several potential problems have been discussed in relation to the paired format. Pairing up candidates may entail potential problems of mismatch between them with respect to their proficiency levels and/or personality. If the personalities are markedly different, this may affect both the performance and the assessment of the candidates. When a candidate has to work with an incomprehensible or uncomprehending partner, this may negatively influence the candidate's performance. Moreover, in the belief that they are helping their partners, more proficient candidates might not perform at their best. Some have argued that it is impossible to make a valid assessment of the same candidate's abilities when s/he is clearly disadvantaged by a mismatching partner. Substituting the assessor's real impressions of the candidate's performance with hypotheses concerning how s/he would have performed with a different partner has to be ruled out for obvious reasons. Therefore, it is vital for language testers to understand the impact of mismatch between candidates' proficiency levels and/or personality on test performance in order to eliminate harmful effects or unwanted variation.

In contrast to such negative views of the paired format, some have expressed positive views with regard to the beneficial impact and positive features of the peer-to-peer examination. For example, intermediate-level university students in Italy were found to show noticeable willingness to communicate and collaborate with each other when they took a classroom oral test in pairs. The role-play tasks used in the exam managed to elicit a large sample of language, showing a high level of student involvement. Students felt that they had control, which in turn gave them greater confidence. Paired orals were more likely to make students feel at ease and use language in a more natural and purposeful way than in the oral interview, where they would always have to address the teacher as a superior.

Experience of pilot oral examinations conducted within the Hungarian Examinations Reform Teacher Support Project shows that the paired format supports good teaching and is greatly appreciated by students. Students'

questionnaire responses have revealed a reduction in levels of stress, which was probably due to the fact that they knew each other and were interested in similar things to talk about. As a result, students were able to co-operate satisfactorily with their partner.

What seems to be common to the views concerning the paired exam we have reported so far is that all of them fail to provide adequate research evidence that could support the claims and counter claims. Sadly, there has been little empirical research focusing on the effects of the elicitation mode on the scores in paired orals. Investigations conducted in Japan, however, explored whether test takers' scores were influenced by the degree of acquaintanceship between the candidates. The results of the study showed a significant effect: close acquaintanceship positively affected candidates' ratings as they received higher scores when interacting with friends. In a Turkish context, acquaintanceship was also found to be a significant factor, but its impact worked in the opposite direction. Turkish candidates performed better (i.e. got higher scores) when they had to interact with a stranger. Thus, it seems that acquaintanceship is likely to be a culture-specific variable, which means that its effect may vary depending on the cultural background of the test takers. As acquaintanceship appears to be an important variable that can influence test performance, it is recommended that candidates should be allowed to choose their own peer partners.

Another important variable that is believed to influence candidates' performance in the paired exam – candidates' level of proficiency – has been examined in the context of the new Hungarian school-leaving exam in English (proposed by the Hungarian Examinations Reform Teacher Support Project). An empirical study investigated how test takers' scores were influenced when they had to engage in a pair-task performance with partners at different levels of proficiency. Contrary to beliefs about the negative impact of mismatching partners with respect to proficiency, the study showed that there was no statistically significant difference between candidates' performance ratings across different candidate pairings, formed on the basis of differing degrees of mismatch or overlap between candidates' proficiency levels. Peer partners' differing levels of proficiency did not affect candidates' oral test scores. This finding, therefore, provides evidence in favour of the validity of the paired speaking exam format.

In this chapter, we have considered some variables that may influence candidates' performance in oral language assessment. It is very important for test developers to bear in mind that test takers' performance may be either enhanced or negatively influenced by the various facets of the testing situation. Not all language testers are researchers who can explore the relationship between different variables, but at least they should be aware of how different tasks and variation in interlocutor behaviour may influence candidates' performance. In order to eliminate or minimize unwanted variation in performance and to help candidates to give their "best performance" at the exam, language testers should be familiar with what makes good test tasks and how to train oral examiners. Therefore, in Parts Two and Three we will discuss the most important issues and principles related to task design and examiner training.

PART TWO

DESIGNING ORAL EXAMINATION TASKS

Introduction

The design of speaking tasks will crucially determine the extent to which valid and reliable assessments can be made of candidates' oral language ability. Examination tasks are the most important means that test developers have to influence what kind of language performance will be elicited from test takers. Therefore, exam developers should make sure that their tasks are appropriate for the purpose of the exam (general vs. specific purpose; proficiency vs. achievement testing), the intended group of test takers (in terms of their age, educational background, needs for the language, background knowledge, etc.) and the level of proficiency that the given exam aims to measure. Without considering carefully the effects and demands of tasks, which can be judged best through piloting, candidates' performance cannot be taken as an appropriate basis for making valid judgments about the specific abilities and skills the exam is targeted at.

Similarly to language teaching, there is no best method in language testing either, as different task types will measure different aspects of language knowledge and may present different demands. As has already been suggested, a variety of elicitation techniques or tasks should be used in order to sample candidates' proficiency in a number of different contexts. Certain task types are mainly used in the individual mode, others are recommended for use in the paired mode.

Thanks to the Hungarian Examinations Reform Teacher Support Project, a variety of task types in different examination formats (individual vs. paired) have been tried out. In this part of the book, we will discuss the lessons learnt from the pilot oral exams. As we have experimented with four main task types, you will find guidelines for designing the following:

- interview
- picture-based individual long turn
- discussion
- role-play

In order to highlight different options in task design, we will discuss why certain tasks are bad and what features of the design are likely to make a speaking task work well. In many cases you can check or make further judgments about the appropriacy of the tasks presented by viewing sample performances on the accompanying DVD. We would like to emphasize that some of the tasks we present are NOT to be used in class because they are examples of poor design. We believe that by discussing the bad features of specific tasks, the reader can gain

better insights into various pitfalls that are to be avoided. It is also important to note that some of the accompanying performances on the DVD exemplify how students perform on tasks that are problematic. These performances are intended to highlight problems with the TASK and not with the STUDENT. A performance may be poor because of the task and students might have been able to perform better if they had had better tasks.

Chapter 3:

The Interview

Traditionally, oral examinations have been based on an interview led by an examiner, whose job is to put questions to the candidate in order to elicit language performance for assessment purposes. The interview is often considered to be a prototypical exam format although its scope is limited because the examiner is in control of the conversation, initiating and concluding topics and so the flow of information is one-way. Such an imbalance in conversational rights and duties does not reflect the way we use language in everyday communication since language users often initiate, redirect and conclude topics, and they often want to get information, not simply give it. In order to make up for the shortcomings of the interview format, modern European examinations use different task types that can measure different aspects of language ability by requiring the candidate to perform in a variety of language use contexts.

Nevertheless, the interview format is one useful test method, provided that it is used in conjunction with other methods, such as the ones described and illustrated in this Handbook. In an interview, the candidate is often asked about a number of different topics that s/he can relate to. This test format is probably suitable for eliciting personal information about the candidate and getting him/her to express opinions on certain issues. These are obviously valid purposes for language elicitation as language users should be able to introduce themselves, to talk about their hobbies, plans, jobs/school, families, and they should also be able to say what they think about relevant issues. Of course, the degree of detail expected in their responses will vary according to the proficiency level that the candidate is being examined for. However, the interview also performs a social function as it helps to establish a relationship between the examiner and the candidate.

The interview as an elicitation procedure can be free (unstructured) or guided (structured). The danger of conducting oral interviews without any guidance is that interlocutors may differ markedly in the way they ask questions. Research has shown that in unstructured interviews with the same candidate being interviewed by two different interlocutors, the two interlocutors used very different elicitation techniques. The interlocutors differed most in

- the way they structured topics:
One of the interlocutors developed the topics systematically, while the other interlocutor was unsystematic. The first interlocutor extended topics by drawing on information provided by the candidate while the second interlocutor failed to do so. In addition, the first interviewer closed each topic clearly while the second interviewer tended to close topics only indirectly.
- the questioning techniques they used:
The first interlocutor systematically employed closed questions (*yes/no* and *or-questions*), which were followed by questions that explicitly elicited an extended re-

sponse from the candidate. In contrast, the second interlocutor used closed questions, which were followed only by echoing what the candidate had just said.

- the feedback they gave:

The first interlocutor gave feedback indicating understanding and interest while the second one gave little positive feedback or indication of interest.

As a result of the differences between the two interlocutors' style, the candidate's performance was very different in the two exams. When four independent raters marked the candidate's two performances, they drew different conclusions about the candidate's ability: in one of the interviews the candidate was characterised as being willing and responsive while in the other she was labelled as unforthcoming and uncooperative. Therefore, it is not surprising that the raters also gave different scores.

Variation in interviewer style has serious implications for the validity of the exam. If performances vary according to the interviewer's style, which performance reflects the candidate's true abilities? Which score should we take as a valid indicator for the candidate's oral proficiency? It is very difficult to answer such questions. Therefore, language testers must design examination procedures in such a way that ensures equal treatment for all candidates. Language testers should try to design speaking examinations and interlocutor training procedures which help minimise unwanted variation in interviewer style even at the expense of sacrificing the so-called 'naturalness' of the conversation for the sake of test fairness. For some people it may seem that if the examiner's behaviour is strictly controlled and his/her contributions are scripted, the naturalness of the interaction between the interlocutor and the candidate is lost. However, an oral interview can never replicate a real-life encounter as it represents a special activity type in itself. An interview is not a conversation. It is very rarely the case that in a test the participants would really like to get to know each other and exchange information or opinion about a specific topic. In reality, the interlocutor's job is to elicit a rateable language sample from the candidate, whose performance will be assessed according to some criteria.

Standardizing interlocutor behaviour

In modern European speaking examinations, interview procedures are guided in order to ensure that interlocutors adopt standard procedures when conducting the examination. The extent to which a speaking examination is guided may vary from exam to exam, especially because much depends on the type of speaking activity used. In role plays, for example, the interlocutor's contributions cannot be fully scripted as it is impossible to foresee all the possible reactions by candidates to the task prompts.

In order to standardize the examination procedure as much as possible, the interlocutor's contributions should be as guided as possible. For this purpose, the modern European speaking examination model presented in this Handbook includes an *Interlocutor Frame*. This document prescribes what should happen in the

examination and when it should happen. The interlocutor's contributions (questions and instructions) are also specified as it may make a big difference to a candidate's performance if the interlocutor paraphrases the questions. In the test development phase all the interview questions were piloted and therefore the wording of each question matches the intended proficiency level. Piloting exam questions is essential as the interlocutor may not be able to use the most appropriate words when paraphrasing on the spur of the moment, and in the end s/he might confuse rather than help the candidate. This is exactly what seems to happen in DVD Samples 3.1 and 3.2. The DVDs are intended to demonstrate how the lack of an Interlocutor Frame may influence the Interlocutor's behaviour, which in turn may negatively influence the candidate's performance. The two samples demonstrate that without an Interlocutor Frame, the same examiner may conduct exams in a very different way. Both DVD samples show a single candidate's performance in an interview.

Table 3 contains guiding questions to help the viewer to evaluate the performances in DVD Sample 3.1 and 3.2. Before viewing the performances you should read through the questions. While watching the DVD, you can take notes in the boxes provided. You are advised not to go on reading until you have filled in Table 3 below.

Table 3 Guiding Questions for DVD Sample 3.1 and 3.2

	Sample 3. Candidate: Zsolt	Sample 3.2 Candidate: Zoltán
1. Does the Interlocutor use paraphrasing ?		
2. Does the Interlocutor use display questions ?		
3. Does the Interlocutor mostly use global questions or ask for specific information concerning the topic?		
4. Who talks more: the Interlocutor or the Candidate?		
5. Do the Interlocutor's questions always help the Candidate to perform? If no , why?		
6. Does the Interlocutor make unnecessary comments ?		
7. How is the Interlocutor's behaviour different in Sample 3.1 and Sample 3.2 ?		

The DVDs show that without an Interlocutor Frame the examiner often paraphrases his own questions and sometimes the paraphrased questions do not help the candidate at all (e.g. *What can you do? What's good in Sándorfalva? What do you do at home in the evenings?*). The paraphrasing strategy does not seem to work in DVD Sample 3.1, where the interlocutor's efforts to help the weak candidate (Zsolt) at all costs result in the interlocutor dominating the interaction. In DVD Sample 3.2, because of the interlocutor's personal interest in the topic of football, the questions become highly specific and therefore the candidate (Zoltán) has only limited scope for producing an extended response (e.g. *Which league does your football team play in?; How many matches did the team win?; Who is number one?*), and some of the questions focus on the knowledge of specific vocabulary items (e.g. *Where do you play?; What position do you play?*). In order to keep the conversation going, in DVD Sample 3.2 the interlocutor makes unnecessary comments (e.g. *We're going to talk a bit; So you have to work hard*) and uses display questions as well (e.g. *It's a village, isn't it?*). The latter should be avoided as this question type elicits information already known to the questioner, and its use seems to further strengthen the artificial or unnatural features of the interaction. The differences in the interlocutor's style in these two samples make it difficult to compare the two candidates' performances reliably.

In interviews the interlocutor should follow specific guidelines. Standard behaviour with respect to interviewer contributions may include recommendations such as:

- Use global questions for elicitation.
- Use *wh*-questions instead of *yes/no* questions whenever possible.
- Never ask more than one question at a time.
- Do not talk more than necessary: refrain from making unnecessary comments.
- Do not interrupt or finish what the candidate wants to say.
- Do not ask questions that require background knowledge.
- Avoid ambiguous and embarrassing questions.
- Use genuine questions and avoid display questions.
- Maintain eye contact with the candidate when talking to him/her.

Designing interview questions

In interviews, candidates are often asked to introduce themselves and say something about their family, job/school, hobbies, or something that relates to their everyday life. The range of questions appropriate for this elicitation format is fairly limited and thus it can be quite easy for the candidate to contribute since many of the questions are predictable. In other words, candidates can rehearse their responses to questions which relate to their personal background. As long as the interlocutor varies the questions and does not allow the candidate to recite a monologue, the interview format can provide a good opportunity for candidates to talk about themselves and interviews are often a good and natural way to begin an oral exam. The questions themselves, however, should be designed with care as

not all questions are likely to trigger the desired reaction or to elicit the desired language from the candidate.

Consider the following questions that aim to elicit personal information about the candidate at A2 or B1:

Table 4 Sample Interview Questions Intended for A2 & B1

Sample Set 1	Sample Set 2
1.1 Where are you from? Tell me about your town/village.	2.1 Where do you live?
1.2 How do you spend your free time?	2.2 Do you have friends?
1.3 What do you like most about television?	2.3 Do you like watching TV?
1.4 Tell me about your family.	2.4 Do you have any brothers or sisters?
1.5 What kind of films do you like? Why?	2.5 What is your favourite film?

If we compare the two lists in Table 4, we find that the questions in Set 2 are much less appropriate than in Set 1. In Set 2, the candidate is given much less scope for contributions as the questions are either *yes/no questions* or they ask for specific information. The question ‘Where do you live?’ can be fully answered by mentioning the name of the town/village where the candidate lives. The candidate may not necessarily realise that s/he could also describe the place in some detail. If the candidate uses good test-taking strategies, s/he may provide an extended response to all the questions in Set 2 although none of the question prompts require him/her explicitly to do so. In Set 1, however the candidate is explicitly asked to mention more than just a place name, the title of a film or to give a short *yes* or *no* answer. Questions 1.1–1.4 are global questions, they invite the candidate to speak about a specific topic but at the same time they do not restrict his/her response as much as questions 2.1–2.5. When the candidate is asked what s/he likes most about television, the range of possible answers is virtually unlimited. Similarly, candidates may respond to question 1.2 in a number of different ways: some will just list activities they like doing in their spare time while others may also mention who they spend their free time with, what their favourite place is like, etc. Although there are two questions asking for specific information in Set 1, it must be pointed out that in both cases there are follow-up questions that are intended to extend the scope of answer for the candidate. In question 1.5, for example, the candidate is asked to justify his/her choice. This is a very useful interviewing strategy in general.

At higher proficiency levels, candidates must show that they are capable of talking not only about themselves but about more general issues as well. In Table 5, good and bad examples of interview questions are shown. Because of the requirements of the higher proficiency level, the topic focus is somewhat different

from that of Set 1 and 2. Our assumption is that the wording of the questions matches the intended level (B2). This needs to be ensured by appropriate piloting.

Table 5 Sample Interview Questions Intended for B2

Sample Set 3	Sample Set 4
3.1 Tell me about your plans for the future.	4.1 Do you like going to the cinema?
3.2 How do you think American culture influences our life?	4.2 Why do you think many countries see the American influence as a threat to their national identity?
3.3 How have the technological inventions of the last century changed our life?	4.3 Which scientific inventions of the 20 th century do you think have changed people's lives the most?
3.4 Which is your favourite historical period? Why?	4.4 If you had been born 500 years earlier, how different do you think your life would be?
3.5 What may influence the way people are dressed?	4.5 Are you fashion conscious? What articles of clothing are fashionable these days?

The first pair of questions (3.1 & 4.1) shows that even at a higher proficiency level the candidate may be asked personal questions, especially at the beginning of the examination as a kind of warmer in order to put him/her at ease. However, the personal questions should still be phrased in accordance with the guidelines mentioned above. Question 4.1 fails to comply with the guidelines and so it is inappropriate: it is a closed question and requires only a minimal response. Nevertheless, candidates at this level seem to be more willing to take the initiative and they may give an extended response even to a closed question.

In accordance with the requirements of the B2 level, the candidate will primarily be asked about his/her opinion concerning general issues. Questions 3.2–3.5 are examples of the kind of questions that may be put to the candidate. As shown in Table 5, although the same topics are explored in Set 4 as in Set 3, still the focus of the questions seems to be inappropriate. Question 4.2 is strongly biased: it is based on the assumption that the candidate agrees with the statement although it may not be true. When answering questions 4.3 and 4.4, the candidate may not possess the necessary background knowledge or may not be able to activate it. If the candidate is asked to justify something s/he does not agree with, or has little or no information about, his/her response is likely to be fairly short. The problem with question 4.5 is twofold: on the one hand, candidates (especially teenage boys) may not have the necessary background knowledge. On the other hand, the question simply requires the candidate to list some fashionable items of clothing. This task seems to measure one's vocabulary knowledge rather than one's oral communication skills.

To sum up, interview questions should be phrased in accordance with the requirements of the given proficiency level and in such a way that they provide the

candidate with the maximum scope for response. Ideally, the interview is built around global questions, relating to neutral themes, which exclude, for example, current political issues, disasters or sensitive matters such as income or health problems.

Sample interview questions

As suggested above, the interlocutor should follow a standard procedure for conducting the exam. In the following sets of interview questions, this principle is followed as the interlocutor's sheet includes not only the questions s/he may ask but almost all the contributions (introduction and other comments) the interlocutor may make during this examination phase. On the following pages there are eight sets of sample interview questions: Sets 1–4 are intended for A2 & B1 while Sets 5–8 are intended for B2. Note that these interviews are only the first part of an oral examination that consists of a number of other, different tasks.

THE GUIDED INTERVIEW

SET 1

(Intended Levels: A2 & B1)

2-3 minutes, 3-4 questions

Hello. My name's ... *(the Interlocutor introduces him/herself)*

What's your name? *(to the candidate)*

I'd like to ask you some questions, ... *(the Interlocutor uses the candidate's name)*

2. What kinds of shops are there near your home?
3. Where do you go when you do the shopping?
4. Is learning languages important for you? Why/Why not?
5. What do you like most about your school?
6. Can you describe the home you would like to have in the future?

After 2-3 minutes

Thank you.

DVD Sample 3.3 shows a sample candidate performance on these interview questions.

THE GUIDED INTERVIEW
SET 2

(Intended Levels: A2 & B1)
2-3 minutes, 3-4 questions

Hello. My name's ... *(the Interlocutor introduces him/herself)*

What's your name? *(to the candidate)*

I'd like to ask you some questions, ... *(the Interlocutor uses the candidate's name)*

1. What do you like most about your school?
2. Can you tell me about your favourite subjects at school?
3. What do you hope to do in the next few years?
4. Can you tell me about the family you would like to have in the future?
5. How do you usually spend your holidays?
6. Do you like shopping? Why/Why not?

After 2-3 minutes

Thank you.

THE GUIDED INTERVIEW
SET 3

(Intended Levels: A2 & B1)
2-3 minutes, 3-4 questions

Hello. My name's ... *(the Interlocutor introduces him/herself)*

What's your name? *(to the candidate)*

I'd like to ask you some questions, ... *(the Interlocutor uses the candidate's name)*

1. Can you tell me something about your family?
2. Why did you choose to study in this school?
3. What do you enjoy most about learning English?
4. Where are you from? Can you tell me about your town / village?
5. Do you go out in your free time? *(If yes)* Where do you go?
6. What do you and your friends like wearing when you go out?

After 2-3 minutes

Thank you.

**THE GUIDED INTERVIEW
SET 4**

(Intended Levels: A2 & B1)
2-3 minutes, 3-4 questions

Hello. My name's ... *(the Interlocutor introduces him/herself)*

What's your name? *(to the candidate)*

I'd like to ask you some questions, ... *(the Interlocutor uses the candidate's name)*

1. What do you and your friends like wearing at school?
2. Can you tell me about the job you would like to have in the future?
3. Where are you from? Is it easy to get to different places in this town/village?
4. Have you got any hobbies? *(If yes)* How did you become interested in it / them?
5. Do you go out in your free time? *(If yes)* Where do you go?
6. What do you enjoy most about learning English?

After 2-3 minutes

Thank you.

THE GUIDED INTERVIEW

SET 5

(Intended Level: B2)

5-6 minutes

Hello. My name's ... *(the Interlocutor introduces him/herself)*

My colleague is just going to listen. *(the Interlocutor refers to the silent Assessor)*

What's your name? *(to the candidate)*

First, I'd like to ask you some questions. *(the interlocutor says the candidate's first name)*

What are your plans after finishing this school?

I see, thank you.

Let's talk about something else.

Recently it has become easier to move to and work in a foreign country.

- Would you like to move to and work in a foreign country? Why/Why not?
- What difficulties might you have to face in a foreign country?
- If you moved to a foreign country, in what ways do you think your life would change?

OK, that's enough, thank you.

Let me ask you now about teenagers.

Parents sometimes find it difficult to get along with their teenage children.

- How much of the housework should teenagers do at home?
- Should parents always allow their children to do what they like?
- Should teenagers get a lot of pocket money? Why/Why not?
- Why do teenagers often find it easier to get along with their grandparents?

After 5-6 minutes

That will do, thank you.

DVD Sample 3.4 shows a sample candidate performance on these interview questions.

The Guided Interview

SET 6

(Intended Level: B2)
5-6 minutes

Hello. My name's ... *(the Interlocutor introduces him/herself)*

My colleague is just going to listen. *(the Interlocutor refers to the silent Assessor)*

What's your name? *(to the candidate)*

First, I'd like to ask you some questions. *(the interlocutor says the candidate's first name)*

What do you like doing in your spare time?

I see, thank you.

Let's talk about something else.

It seems that television has changed people's lives a lot.

- Do you think it has an important part in many people's life? Why/Why not?
- Could you easily do without it? Why/ Why not?
- In what ways can television be harmful or dangerous?
- Is there enough variety of TV programmes nowadays?

OK, that's enough, thank you.

Let me ask you now about summer jobs.

It seems that many students look for a summer job or a part-time job.

- What are the chances of finding a good job nowadays?
- What is the ideal part-time job / summer job for a teenager?
- Do you agree with those students who want to spend most of their summer working? Why/Why not? / Would you spend most of your summer holiday working?

After 5-6 minutes

That will do, thank you.

THE GUIDED INTERVIEW
SET 7

(Intended Level: B2)
5-6 minutes

Hello. My name's ... *(the Interlocutor introduces him/herself)*

My colleague is just going to listen. *(the Interlocutor refers to the silent Assessor)*

What's your name? *(to the candidate)*

First, I'd like to ask you some questions. *(the interlocutor says the candidate's first name)*

What do you like about your school?

I see, thank you.

Let's talk about something else.

Nowadays many people learn foreign languages, even after school.

- Why do you think it is important to learn foreign languages?
- Which are the most popular foreign languages? Why?
- In what kinds of jobs is it essential to know foreign languages?
- What are the advantages of speaking more than one foreign language?
- What are the best ways of learning languages?

OK, that's enough, thank you.

Let me ask you now about advertisements.

It seems that people buy everything that is well advertised.

- Have you ever bought anything because of an advertisement you saw/read/heard? What was it?
- Which products are advertised too often? Why?
- What do you find the most effective way of advertising?

After 5-6 minutes

That will do, thank you.

THE GUIDED INTERVIEW SET 8

(Intended Level: B2)

5-6 minutes

Hello. My name's ... *(the Interlocutor introduces him/herself)*

My colleague is just going to listen. *(the Interlocutor refers to the silent Assessor)*

What's your name? *(to the candidate)*

First, I'd like to ask you some questions. *(the interlocutor says the candidate's first name)*

Which is your favourite season of the year? Why?

I see, thank you.

Let's talk about something else.

It seems that shopping habits have changed over the last 10 years.

- When and where do people typically do their shopping nowadays?
- What are the advantages and disadvantages of shopping in big chain stores?
- Do you think that young people and the older generation have similar or different shopping habits? Why?
- How have shopping centres or plazas changed people's lifestyle?

OK, that's enough, thank you.

Let me ask you now about collecting things.

It seems that nowadays hobbies like collecting stamps, coins or other small things are less popular with young people.

- Do you agree? Why? / Why not?
- Which hobby seems to be the most popular with young people today? Why?
- What is the relationship between people's lifestyles and their hobbies?
- How do you imagine the hobby of the future generation? Will it be the same or different?

After 5-6 minutes

That will do, thank you.

In this chapter, we have examined how oral interviews could be guided in order to enhance the reliability and fairness of oral examinations. We have argued that the use of an *Interlocutor Frame*, which describes the procedure for conducting the exam and includes prescribed and/or potential contributions by the interlocutor during the exam, can ensure that the elicitation procedure is comparable across different candidates. As a result, variability in examiner behaviour can be minimised. We have also discussed some design considerations for writing interview questions by comparing different question types. The eight sets of interview questions provided at the end of the chapter are intended as good examples for this elicitation method. In the following chapter, we will explore another popular oral testing technique: how to use pictures in order to elicit an individual long turn from candidates.

Chapter 4:

The Individual Long Turn

Another common elicitation technique in oral examinations aims to provide candidates with an opportunity to demonstrate their ability to produce long turns by describing or comparing and contrasting visual prompts. Pictures are the most widely used prompts for eliciting language from candidates. The reason for this can be found in the advantages of this technique:

- Pictures can be obtained quite easily, they are part of our everyday life. Newspapers, magazines, books, leaflets, postcards, photographs, drawings and pictures from the internet provide inexhaustible and immensely varied sources for test designers and item writers.
- Well-chosen pictures can offer economic and effective ways to elicit a long turn from the candidates without providing them with any language input in the target language to copy.
- While the topic of the test is determined by the picture, candidates have the freedom to show their mastery of the target language. Pictures provide excellent opportunities for personal reactions and interpretations.
- Deficiencies in reading comprehension cannot prevent candidates from doing well at such speaking tasks: they produce language about what they see.

However, this apparently easy technique is beset with pitfalls. Picture selection is one of the hardest tasks for item writers. It is very easy to choose a “nice” picture which will not elicit the required quality and amount of language because it is not suitable for testing purposes. The most common problems with pictures are the following:

- The picture is not challenging enough, and does not contain enough stimuli to elicit language from the candidate.
- The picture is culture-dependent, requiring special background knowledge from the candidate.
- The topic of the picture is distressing, offensive, violent or taboo, which may affect the candidate’s performance.
- Surreal, abstract and symbolic pictures can prevent candidates, especially at lower levels, from performing instead of facilitating their language output.
- Bizarre, unrealistic situations in the pictures are unlikely to elicit appropriate, life-like language output from candidates.
- Using too many pictures to compare and contrast for one task makes the candidate’s task very difficult, often impossible to carry out. Instead of producing more and more varied language, the candidate might be incapable of coping with the quantity of information.
- Using only one picture without the opportunity to compare and contrast might lead to a simplistic physical picture description instead of exploring the

given topic from different angles. This elicits very limited language output both in terms of quantity and variation of vocabulary.

- If the pictures used for the “compare and contrast” task-type are too similar, they are not suitable for eliciting sufficiently varied language: a limited range of vocabulary and structures will be produced by the candidate.
- Pictures without action are prone to elicit only a limited, descriptive language output.
- Over-crowded pictures that contain too many details can puzzle the candidates and prevent them from focusing on a single main issue. Even well-trained candidates can end up trying to describe every single minor detail instead of concentrating on the main task.

The success of an individual long turn task depends not only on the success of the selection of picture prompts. Examiner behaviour is just as important. Detailed guidelines for interlocutors can help them to conduct this part of the speaking examination in a standardised manner, eliciting the best possible language output from the candidate.

Modern European examinations apply *Interlocutor Frames* to ensure standardised administration of speaking tests. The interlocutor’s contributions (questions and instructions) are carefully guided and described in as much detail as possible in the Frame. The interlocutor is expected to ask questions ONLY from a set of pre-specified options indicated on the interlocutor’s sheet. The interlocutor is only allowed to make comments that are in compliance with the guidelines for conducting the examination.

In the independent long turn tasks demonstrated in this chapter the interlocutor gives the candidate a task sheet with two or three picture prompts. The candidate has to compare and contrast them, and give some personal reaction to the pictures. The Interlocutor’s copy of the task sheet contains question prompts. The interlocutor has to select and ask follow-up questions even if the candidate manages to talk about the pictures at length. It is not necessary for the interlocutor to use the question prompts in the order listed – the number of questions used is likely to depend on the quantity and quality of candidates’ responses. Naturally, those candidates whose contribution is too short will have to be asked more question prompts in order to elicit enough language for assessment. If a candidate is obviously weak (e.g. s/he is struggling painfully with the task, keeps repeating what has been said before) and it seems quite clear that s/he cannot complete the task, it is better to say *Thank you* and proceed to the next part of the examination. On the other hand, if a candidate’s individual turn is too long, the interlocutor may have to politely but firmly interrupt him/her by saying *Thank you. That will do*. Even in such cases, the interlocutor should ask the candidate one or two of the question prompts.

**SHORT VERSION OF THE INTERLOCUTOR FRAME
FOR THE INDIVIDUAL LONG TURN TASKS
WITH PICTURE PROMPTS**

Now I'm going to give you some pictures. I'd like you to talk about them.

Please draw a number, ...

It is number ...

Here are your pictures.

Read out the task specific instructions, wait 5-10 seconds.

Shall I repeat what you have to do?

Wait max. 30 seconds

Let's start now.

After 5-6 minutes or when the candidate has finished

That's enough. Thanks. / Thank you.

In the following section sample speaking tasks with explanations are provided to demonstrate some common problems with the selection of picture prompts and task design.

Sample speaking tasks to demonstrate some common problems with the selection of picture prompts and task design

In this section we demonstrate some typical mistakes in selecting picture prompts. Task design problems are also exemplified, offering practical options for remedy.

SAMPLE 1: GOING OUT

CANDIDATE'S SHEET



SAMPLE 1: GOING OUT INTERLOCUTOR'S SHEET

These pictures show different ways of going out. Look at them in order to compare and contrast them. Start when you are ready.

Picture 1

- | | |
|------------------------------|-----------------------------------|
| 1. Where are the two ladies? | 4. What are they wearing? |
| 2. What are they doing? | 5. What can you see on the table? |
| 3. What do they look like? | 6. What are they drinking? |



Picture 2

- | | |
|----------------------------------|---|
| 1. What are these people doing? | 4. What can you see on the table? |
| 2. What does the room look like? | 5. Are these people enjoying themselves? Why? |
| 3. How do you like this place? | |

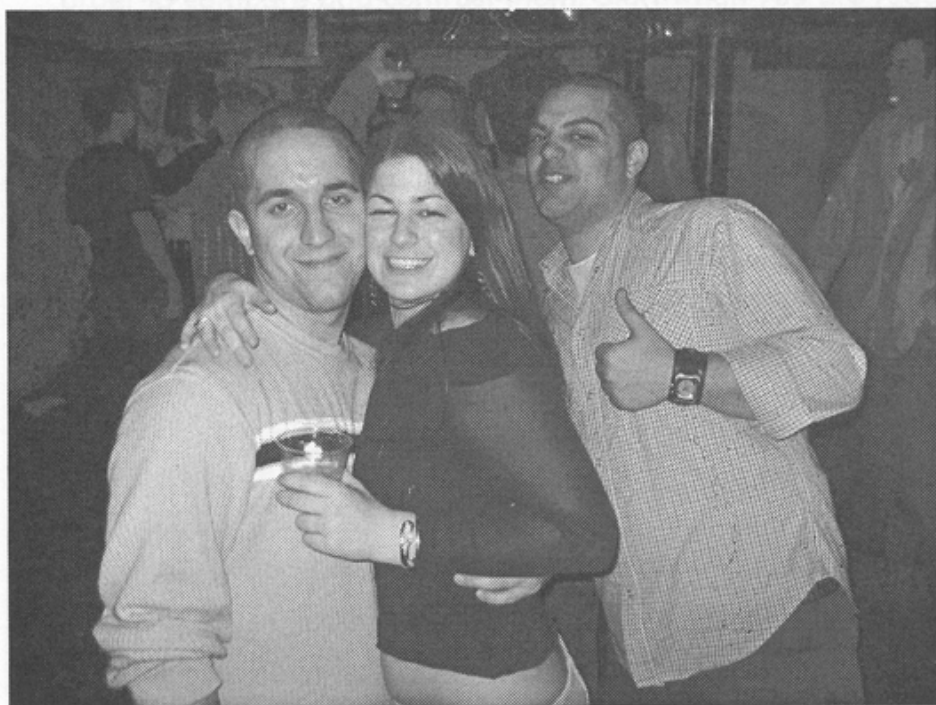


This is a “compare and contrast”-type task. Saying in the rubric that the pictures show two different ways of going out is quite misleading because they show fairly similar situations and there is not enough contrast between them. Pictures with much more significant contrast should be chosen for this task-type.

The interlocutor has two different sets of questions, one for each picture. This encourages dealing with the pictures individually, one by one, instead of comparing and contrasting them. It is recommended that each question should relate to both pictures, giving the candidate maximum scope for contributing to the task.

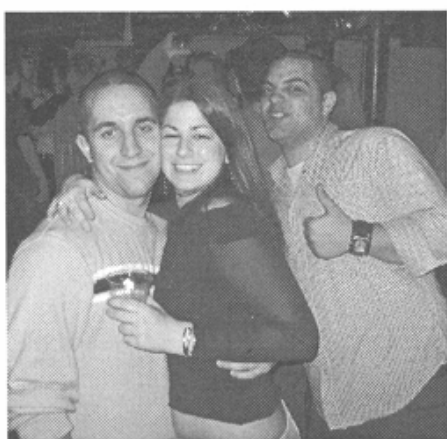
SAMPLE 2: CELEBRATIONS

CANDIDATE'S SHEET



SAMPLE 2: CELEBRATIONS INTERLOCUTOR'S SHEET

These pictures show people at parties. Look at them in order to compare and contrast them. Start when you are ready.



1. Which picture is similar to parties in your family?
2. Which picture is similar to parties you go to with your friends?
3. When do you celebrate together with your family?
4. When do you celebrate together with your friends?
5. How are parties with your family members different from parties with your friends?
6. Which party in the pictures would you like to be invited to? Why?
7. How do you prepare for a party as a host?
8. How do you prepare for a party as a guest?
9. When did you last go to a party? Tell me about it.

This is another “compare and contrast”-type task. There are some differences between the two pictures, but the situations are again fairly similar and there is not enough contrast between them. Pictures with much more significant contrast should have been chosen for this task.

The interlocutor has one set of questions which refers to both pictures. Unlike the previous example, here the nature of the questions encourages contrasting and comparing the pictures to a certain extent. However, questions 3, 4, 7, 8 & 9 do not relate directly to the pictures.

SAMPLE 3: OUTDOORS

CANDIDATE'S SHEET



SAMPLE 3: OUTDOORS INTERLOCUTOR'S SHEET



Questions:

1. What is common in these pictures?
2. What can be the relationship between the people in the pictures?
3. Do you think these four people belong to the same family? Why?
4. Which activity would you choose for your family?
5. Where would you go near your home to do these activities?
6. Which is the best season for these activities?
7. Where do you think you could find an ideal place for these activities?

The task does not have any rubrics, which is unacceptable, since the candidate does not really know what s/he is expected to do. It is wrong to leave it to the interlocutor to set the tasks in a non-standardised manner, since this would allow for unequal, unfair treatment.

The two pictures show very similar situations, both show a parent spending free time with their child. Only the activities and perhaps the places are different. This does not offer enough space for the candidate to elaborate on the differences. Talking about similarities, however, elicits the same type of vocabulary throughout the test. Some of the questions seem to be unnatural or even silly, for example:

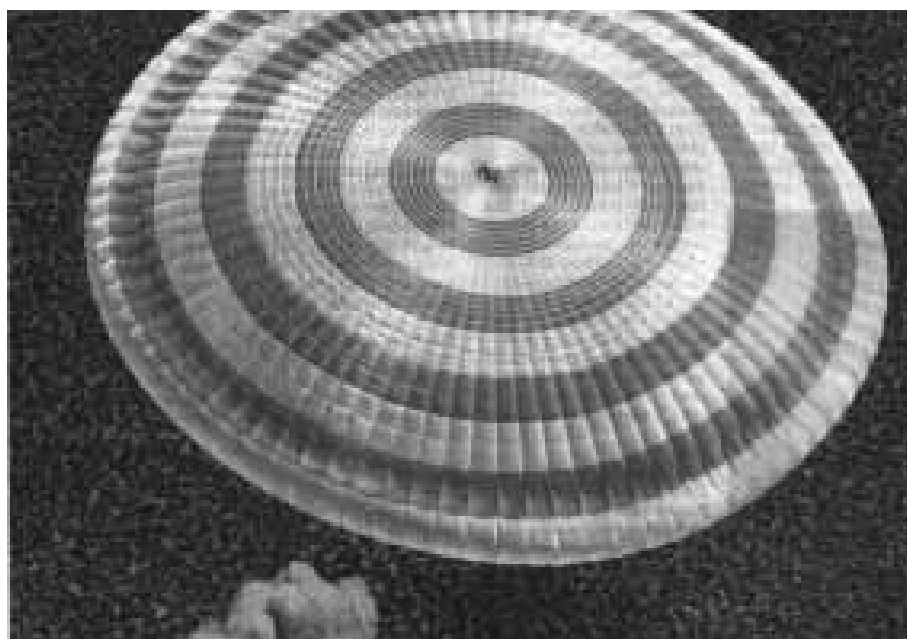
– *Do you think these four people belong to the same family?*

(They are in different pictures, in different places, why should they?)

– *What can be the relationship between the people in the pictures?*

(It is highly likely that they are mother and son and father and son.)

**SAMPLE 4A: PARACHUTIST
CANDIDATE'S SHEET**



**SAMPLE 4B: DOCTOR'S SURGERY
CANDIDATE'S SHEET**



SAMPLE 4A: PARACHUTIST INTERLOCUTOR'S SHEET

Show the candidate the two pictures and ask them to select one.

- First, ask the candidate to describe the action in the picture.
- Second, ask the candidate to suggest who may have taken the picture and why.

This task requires the candidate to select one picture, describe it and make deductions.

The first picture is far from most candidates' everyday, real-life experiences. The parachutist is doing a rather unusual activity. It is very difficult to describe it (for a weak candidate performance see DVD Sample 4.1 where the candidate asks for help and the interlocutor tries to assist with unscripted questions, which the candidate seems not to understand), because

- a) it requires special vocabulary items which even good candidates do not have (see DVD Sample 4.2 where even a really good candidate suffers due to lack of special vocabulary);
- b) there is not a lot happening in the picture, apart from the parachutist falling with the help of his parachute, so there is nothing else to describe or talk about. DVD Sample 4.2 shows the difficulty even a good candidate has with interpreting the picture. As a result, the interlocutor seems to feel obliged to ask (unscripted) questions which he hopes will help the candidate, but which do not offer much support.

The second picture, a photo of a doctor's surgery must have been taken decades ago. Everything is old-fashioned, out-of-date, and so different from what candidates can experience in real life that it is quite difficult for them to relate to. Many special vocabulary items would be needed to describe the surgery and the action in detail. For sample candidate performances see DVD Sample 4.3 where the candidate seems to grope for words, and the interlocutor offers suggestions (e.g. 'weighing machine'). The candidate finds it difficult to say anything about the picture. In DVD Sample 4.4, after giving a brief literal description of the picture a good candidate moves away from simply describing the picture in order to say anything at all.

The second instruction for the interlocutor does not offer a real opportunity for widening the scope of the long turn. There is no reference made to the candidate's own life experience, thus the language output will be impersonal and artificial.

SAMPLE 5: MOBILE PHONES

CANDIDATE'S SHEET



SAMPLE 5: MOBILE PHONES INTERLOCUTOR'S SHEET

Look at these two pictures. Please describe and compare them.

4. April 7. 13:25:14
S



The task is to describe and compare the two visuals. Both pictures show people using their mobile phones in unusual, bizarre, and somewhat dangerous situations, which are highly unrealistic. Although the pictures are interesting in themselves, apart from the mobile phones there is no common theme linking them. Candidates might be puzzled and confused by them, especially because there are no prescribed questions for the interlocutor to elicit language from candidates. Both the candidate and the interlocutor are pretty much left alone with this difficult task. These unreal and stage-managed pictures cannot elicit enough language even from good candidates, and the less able ones will simply freeze. For a sample candidate performance see DVD Sample 4.5 where the candidate has difficulty interpreting the pictures.

SAMPLE 6: DANGEROUS SPORTS

CANDIDATE'S SHEET



SAMPLE 6: DANGEROUS SPORTS

INTERLOCUTOR'S SHEET

The following pictures show people doing some unusual and dangerous sport. Choose two or three pictures and describe what the people are doing and say what is dangerous about these activities. You do not need to name all activities shown.



The topic is appropriate and the number of pictures show considerable variety, but the problem is that out of the five pictures four involve people in the air in one way or another. This means that the instruction for the candidate to say what is dangerous about them is likely to elicit more or less the same type of language. Using dangerous sports that are as different from each other as possible would elicit more varied language output from the candidates. The interlocutor's questions, however, relate to all five pictures, are quite varied and suitable for personalising the topic. With different pictures the task might work better.

For a sample candidate performance see DVD Sample 4.6. The lack of numbers on the pictures does not help the candidate refer to them, but the pictures do not in any case elicit much language from the candidate, which is why the interlocutor tries to find topics that might encourage the candidate to say more.

SAMPLE 7: DANGERS OF SMOKING
CANDIDATE'S SHEET



SAMPLE 7: DANGERS OF SMOKING INTERLOCUTOR'S SHEET

Describe the following advertisement. Say what the advertisement tries to tell people and how effectively it does it.

The subject matter is of a sensitive nature, which is not appropriate in an examination situation. Item writers should avoid the topic of death because there may be a candidate who has recently lost a family member or knows someone who has cancer because of heavy smoking, for example. Because of the judgemental nature of some of the interlocutor's questions, even those candidates who have smokers in their families or smoke themselves might feel uncomfortable while discussing this topic. The task could be used for a classroom discussion, but certainly not in a stressful examination situation.

For a sample candidate performance see DVD Sample 4.7. The interlocutor seems to feel driven to ask further (unscripted) questions which do not get much out of the candidate.



Sample speaking tasks to demonstrate good practice in the selection of picture prompts and task design

In this section we demonstrate good practice. The picture prompts in the following tasks proved to be suitable for comparison and contrasting. They can be easily related to each other, yet there are significant differences between them. Each task stimulates extended use of vocabulary and structures within the same topic area. The presence of people in the pictures in everyday, life-like situations offers candidates good opportunities to go beyond a superficial physical description, and widen the scope of language output. The interlocutor's questions are also suitable for eliciting good language, because some of them invite candidates to relate the situations to their own life, while others require generalisation and abstraction.

The DVD samples (4.8 to 4.13) show how well-chosen pictures can elicit a range of language from candidates without interlocutors having to force them to speak, in ways that might vary from candidate to candidate and therefore favour some candidates over others. What is important when viewing these videos is not to look for mistakes that candidates might make, but rather to focus on how well the tasks – pictures and questions – enable the candidates to perform to the best of their ability.

SAMPLE 8: ROOMS **CANDIDATE'S SHEET**



SAMPLE 8: ROOMS

INTERLOCUTOR'S SHEET

These pictures show two rooms. Look at them in order to compare and contrast them. Start when you are ready.

- What kind of people do you think live in these rooms?
- Do the rooms look like the rooms in your home? What's similar and what's different?
- If you lived in these rooms, what would you change in them?
- Are these rooms typical of Hungarian homes? Why do you think so?



For a sample candidate performance see DVD Sample 4.8.

SAMPLE 9: DOING & WATCHING SPORTS

CANDIDATE'S SHEET



SAMPLE 9: DOING & WATCHING SPORTS

INTERLOCUTOR'S SHEET

These pictures show young people doing and watching sports. Look at them in order to compare and contrast them. Start when you are ready.

- What is good about these activities?
- What is bad about these activities?
- Why do you think these young people enjoy doing these activities?
- Which do you prefer: doing or watching sport? Why?



For a sample candidate performance see DVD Sample 4.9.

SAMPLE 10: HOLIDAYS

CANDIDATE'S SHEET



SAMPLE 10: HOLIDAYS

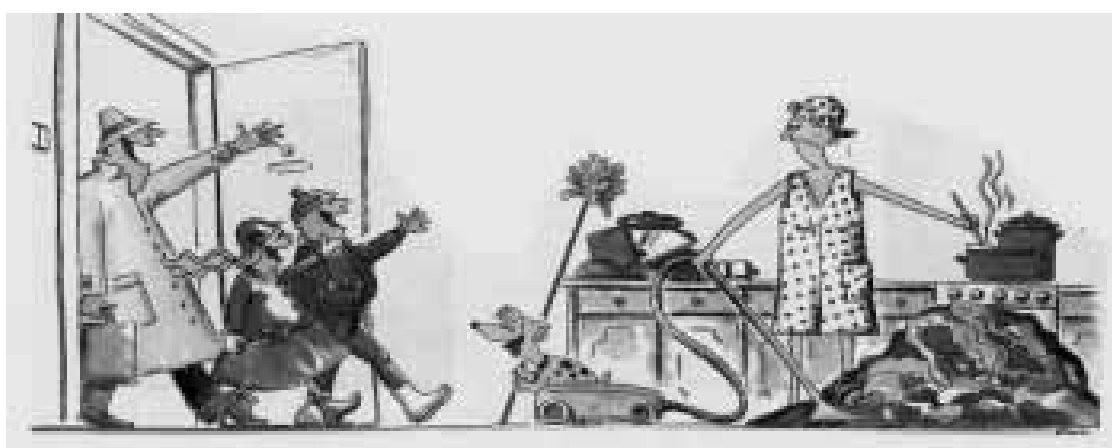
INTERLOCUTOR'S SHEET

These pictures show people spending their holidays. Look at them in order to compare and contrast them. Start when you are ready.

- Which of these places would you choose for a holiday? Why?
- Which of these places wouldn't you choose for a holiday? Why?
- Have you ever been to a place similar to one of these? Tell us about it.
- Would you like to visit either of these places in winter? Why/why not?
- Why do you think people decide to spend their holidays in these places?



SAMPLE 11: DOING HOUSEWORK **CANDIDATE'S SHEET**



SAMPLE 11: DOING HOUSEWORK

INTERLOCUTOR'S SHEET

These pictures show people doing housework. Look at them in order to compare and contrast them. Start when you are ready.

- Which picture is more similar to the family life you would like to have in the future?
- Which picture do you like? Why?
- What do you think will happen next in each of these situations?
- Which picture is more typical of a Hungarian family? Why do you think so?



For a sample candidate performance see DVD Sample 4.10.

SAMPLE 12: LOVE
CANDIDATE'S SHEET



SAMPLE 12: LOVE

INTERLOCUTOR'S SHEET

These pictures show people in love. Look at them in order to compare and contrast them. Start when you're ready.

- Which photo reflects better your idea of love? Why?
- Which one would you prefer to see on the cover of a teenage magazine? Why?
- How well do you think these pictures express the idea of love?
- What may people like about these photos?
- Which photo reflects better your idea of love? Why?



For a sample candidate performance see DVD Sample 4.11.

SAMPLE 13: FAMILIES FROM DIFFERENT AGES CANDIDATE'S SHEET



SAMPLE 13: FAMILIES FROM DIFFERENT AGES

INTERLOCUTOR'S SHEET

These pictures show families from different ages. Look at them in order to compare and contrast them. Start when you're ready.

- Do you think that the role of the mother is different in the two families? Why?
- How do you think these families spend their evenings and weekends?
- What changes in family life do the two pictures reflect?
- How different do you think the lifestyles of these two families are?
- What difficulties might these families have in their daily lives?



For a sample candidate performance see DVD Sample 4.12.

SAMPLE 14: TV PROGRAMMES

CANDIDATE'S SHEET



SAMPLE 14: TV PROGRAMMES

INTERLOCUTOR'S SHEET

These pictures show different TV programmes. Look at them in order to compare and contrast them. Start when you're ready.

- Would this be a good selection of TV programmes for your family? Why / Why not?
- What population do you think these programmes are intended for?
- Which programme do you think Hungarians would prefer to watch? Why?
- Which of these programmes (if any) would you watch in a foreign language? Why?
- What can make these programmes popular with viewers?



For a sample candidate performance see DVD Sample 4.13.

SAMPLE 15: PARENTS WITH CHILDREN CANDIDATE'S SHEET



SAMPLE 15: PARENTS WITH CHILDREN

INTERLOCUTOR'S SHEET

These pictures show parents and children. Look at them in order to compare and contrast them. Start when you're ready.

- What do you think the parent and the child are talking about in these pictures?
- What do you think about these parents?
- Was there a scene in your childhood when a picture similar to these could have been taken?
- Which of the two pictures do you think is about typical Hungarian parents? Why?
- Why do you think these children enjoy being with their parents?



Summary

The samples presented in this chapter demonstrate the mistakes item writers should avoid when selecting picture prompts and designing the individual long turn tasks. Examples of good practice were also shown. The following table summarises the main points, adding some practical tips as guidelines for teachers and item writers for the selection and use of pictures.

Table 6 Guidelines for designing individual long turn tasks

	DOs	DON'Ts
Picture prompts	<ul style="list-style-type: none"> ✓ Use black and white pictures. ✓ Use pictures that are clear and photocopiable (maximum A/4 size). ✓ Vary the source of pictures (photographs, drawings, cartoons, etc.). ✓ Select pictures which show general, everyday, life-like situations which candidates can easily recognise and relate to. ✓ Select pictures which show people in action. ✓ Always use 2-3 pictures to provide the candidate an opportunity for comparing and contrasting. ✓ Select pictures which can be easily related to each other, but with significant differences to elicit a wide range of vocabulary and structures. 	<ul style="list-style-type: none"> ➤ Do not use sensitive, distressing, offensive, violent or taboo topics. ➤ Do not select bizarre, surreal, abstract or symbolic pictures, especially for lower level tests. ➤ Do not select over-crowded pictures. ➤ Do not select pictures which do not contain enough stimuli for language output. ➤ Do not use only one picture. ➤ Do not use too many pictures. ➤ Do not use pictures which are very similar. ➤ If possible, avoid using pictures which do not contain people and actions.
Interlocutor's questions	<ul style="list-style-type: none"> ✓ About 4-6 question prompts should be provided for any set of pictures. ✓ The questions should be independent of one another, i.e. the Interlocutor can choose to ask them in any order. ✓ The questions should always be general enough to relate to all the pictures in the task. ✓ The questions should be worded and focused in such a way that candidates really can produce long turns in response to them. ✓ Each task should have some questions which relate the pictures to the candidate's own experiences and/or opinions. ✓ Each task should be accompanied by follow-up questions which require some degree of generalisation and abstraction from the candidate, depending on the level of the test. 	<ul style="list-style-type: none"> ➤ Do not ask questions which do not relate to the pictures at all. ➤ Do not ask questions which relate to only one of the pictures. ➤ The questions should not elicit a physical description of either of the two pictures. ➤ Do not ask artificial, impersonal questions. ➤ Avoid questions which aim to elicit candidates' opinion about a topic in general as this can easily lead to candidates reciting memorised texts.

Chapter 5:

Discussion Tasks

Discussion activities provide a good opportunity to measure candidates' oral interactional skills since they allow for a two-way information flow. In a discussion, there is usually some kind of opinion gap between the participants because they may hold different views or consider different issues important or unimportant. Thus, candidates are encouraged to negotiate, argue for and against specific ideas or propositions. The most important difference between discussion activities and interviews is that the former focus on exchanging opinions: both the interlocutor and the candidate have to express an opinion, which they need to compare and justify in order to come to some kind of agreement in relation to a specific task. Discussion activities may involve the participants in completing tasks such as comparing, contrasting, listing, rank ordering, redesigning, planning, categorizing, problem solving or selecting. Candidates are generally encouraged to express their own opinion, which distinguishes the discussion technique from a role-play activity. In role-plays, the candidate often takes somebody else's role (e.g. a tourist) in order to reach a particular communicative goal (e.g. to enquire about local sights abroad).

However, discussion activities with an examiner-interlocutor also have a lot in common with interviews as the interlocutor's scope for participating in the conversation is equally limited. The interlocutor cannot contribute to the same extent; s/he cannot hold the floor because it is not the interlocutor whose language proficiency the assessor has to judge. The difference in terms of the social roles and status of the interlocutor and the candidate – so typical of the interview – is unlikely to change or disappear in discussion tasks. Therefore, the candidate may not feel comfortable to take the initiative, to contradict or express disagreement with a partner, who is usually superior in terms of age, language proficiency and has authority due to his/her social role. Because of these limitations of the one-to-one or individual format, peer discussion activities in which participants have similar backgrounds (e.g. age, level of proficiency, interests, life experience) are believed to provide a better means for displaying test takers' interactional abilities.

Observation of pilot exams organized by the Hungarian Examinations Teacher Support Project has revealed that in paired discussion activities peer candidates have more freedom to take the initiative, to challenge or disagree with the partner's points of views than in the individual mode, where the candidate's partner is the examiner. Based on the experience of piloting different discussion activities, both in the individual and the paired mode, we recommend that discussion activities should be designed in the paired format as it allows more scope for candidates to take an active part in a discussion. When talking to a peer partner, candidates often claim to be more relaxed and feel more secure than with an examiner. This is especially true for candidates who are familiar with each other and so they know what to expect from their partner. In the pilot exams, the interests, life experience, beliefs and attitudes of the candidates in pairs often coincided – they were roughly

the same age and studied in the same school – and so they found it easier to relate to a specific topic for discussion in order to exchange their views. In such examination contexts, it seems that candidates can better display their ability

- to express beliefs, opinions, agreement or disagreement;
- to give and seek personal views and opinions in discussing topics;
- to compare and contrast alternatives;
- to comment on the views of others;
- to provide explanation and/or arguments in support of their views.

Since discussion activities in the paired format seem to work better than in the individual mode, we will first focus on how to design paired discussion activities. However, since in some exams candidates may also be asked to perform discussion activities with an examiner-interlocutor, at the end of this chapter we will also suggest guidelines for designing discussion activities in the individual mode.

Paired discussion activities

It is very important that paired discussion tasks relate to life-like situations and engage both candidates by giving them equal opportunities to contribute to the discussion. Successful contribution can be facilitated by giving candidates word or picture prompts, which represent the ideas they may talk about. In this way, candidates can create expectations and activate prior knowledge and/or experience. It is also very important that candidates should be asked to complete some kind of product-oriented task while they are talking. For instance, the task may require them to list some advantages and disadvantages, which they need to agree on with respect to their relative importance. Alternatively, candidates can be asked to select and agree on some important/unimportant or useful/useless aspects of a given topic. Without a specific task focus, however, a discussion may easily become rambling and candidates will end up producing parallel monologues. This is likely to happen, for example, in the case of tasks such as *Paired Discussion Activity 1* below, in which the candidates are not asked to carry out any specific product-focused task (other than ‘discussing’ the problems), and they are not provided with any word or picture prompts either.

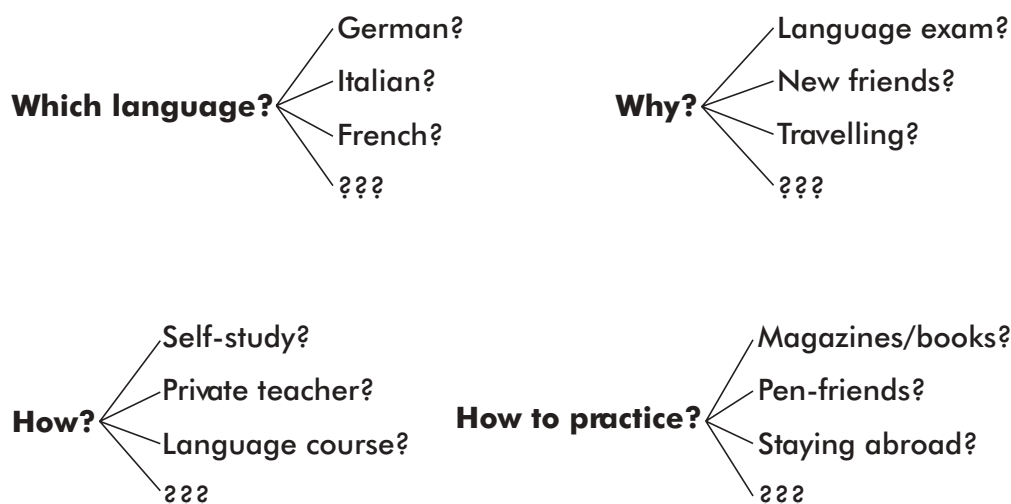
PAIRED DISCUSSION ACTIVITY 1.

Mobile phones have long been a controversial issue. Discuss the problems with your partner.

While it is not advisable to leave a task completely unstructured (such as the one above), we would like to point out that making a discussion task far too structured is equally mistaken. Consider *Paired Discussion Activity 2* below.

PAIRED DISCUSSION ACTIVITY 2.

You and your friend are planning to learn a new foreign language. Discuss the following points on your cards.



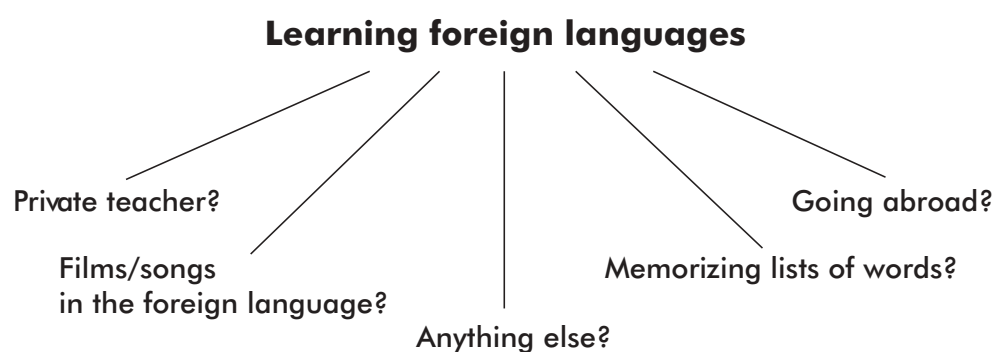
In *Paired Discussion Activity 2*, the task instruction says that the candidates should discuss all the points on their task sheet. There are several word prompts, and both candidates have the same list to consider. These prompts, however, may be problematic for the following reasons. On the one hand, task completion may take a long time, as it could be time-consuming to discuss all the suggested points in detail. On the other hand, there are three question marks linked to each main issue of learning a foreign language and candidates may think that they have to consider a new aspect or idea for each main issue. Because of all that, task completion may be slowed down unnecessarily and could lead to a shift of focus for the candidates: instead of producing a meaningful exchange of opinions, they may simply try to cover all the points listed on the task sheet. The language produced is also likely to be fairly repetitive as candidates may end up repeating the same structures several times. This task could be improved by changing the instruction and the prompts for each candidate. As *Paired Discussion Activity 3* explores the same topic as *Paired Discussion Activity 2*, it is interesting to compare the two tasks.

PAIRED DISCUSSION ACTIVITY 3.

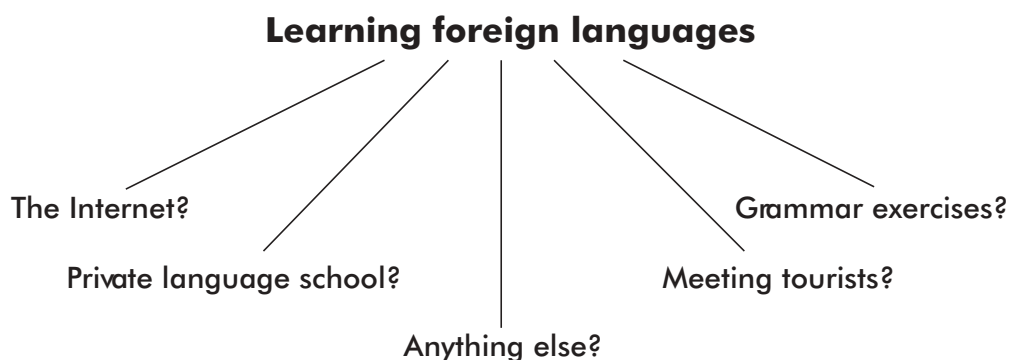
Your school magazine has asked students for advice on learning a foreign language. Think of what has worked for you and discuss with your partner the good and bad points of different ways of learning a foreign language. On your sheet, there are some ideas but you can suggest other things, too. Finally, agree on the three most useful ways.

Start when you are ready.

Prompts for Candidate A:



Prompts for Candidate B:



On comparing the two tasks, *Paired Discussion Activity 3* seems to allow candidates more freedom to choose what they want to talk about than *Paired Discussion Activity 2*. The instruction says that candidates can use the ideas on the sheet but can suggest other things, too. Another major difference between the two tasks is that in *Paired Discussion Activity 3* the two candidates have different prompts to consider. Because of that, it will be more important and perhaps interesting for the two candidates to listen to each other. Finally, the latter task also has a clear task focus

as candidates have to agree on the three most useful ways of learning a foreign language. These small but significant modifications in the design of the task make *Paired Discussion Activity 3* more appropriate for testing candidates' interactional skills in the paired format.

It is very important to choose an appropriate task focus for paired discussion activities. If we examine *Paired Discussion Activity 4*, we find that the two candidates have slightly different task prompts (the name of the profession is the main difference), but still they are not really encouraged to interact with each other. The task fails to require them to bridge any kind of opinion gap. They do not need to compare and contrast their views in order to come to an agreement, for instance. They can simply come up with a monologue describing the profession on their sheet by taking into account the points listed. There is no appropriate focus for this task: the candidates are only required to say whether they would like to choose either of the professions or not. They do not even have to give reasons for their preferences.

PAIRED DISCUSSION ACTIVITY 4.

On your cards, you have the name of two different jobs. Your task is to compare these jobs by discussing the points in the list and then decide whether you would like to choose either of them or not.

Prompts for Candidate A

TEACHER

- What does s/he do?
- Working hours
- Weekends/holiday
- Salary
- Qualifications
- Personality
- Connection with people/colleagues
- Any disadvantages?

Prompts for Candidate B

TAXI DRIVER

- What does s/he do?
- Working hours
- Weekends/holiday
- Salary
- Qualifications
- Personality
- Connection with people/colleagues
- Any disadvantages?

The next sample speaking task (*Paired Discussion Activity 5.A*) is the original version of a paired discussion activity submitted by an item writer for the Hungarian Examinations Reform Teacher Support Project. Based on the guidelines for task design we have presented so far, it should be clear that the task needs modification in order to make it work well in the paired mode. Consider how you could improve this activity. There are some points to help you think through what aspects of the task might need to be changed.

PAIRED DISCUSSION ACTIVITY 5.A

The local government is planning to close down the old local garbage dump and build a new regional one at the edge of the town where you live. The people are divided in their opinion, therefore a referendum must be held. You are participating in a forum some days before the referendum. Evaluate the situation and come to an agreement with your partner. Discuss each argument on your sheets, but you can use other ones, too. Start when you are ready.

Prompts for Candidate A & B

- Costs covered by the European Union
- Close to the river of the town
- Modern technology
- Smell, noise ...
- New jobs for 20
- The town's future as a thermal spa ...?
- Why store the waste of others?
- The present landfill is all but safe
- Toxic wastes?
- Too few new jobs
- Garbage will be selected
- Profit: other settlements will pay for storage
-

Consider the following:

- Is the task appropriately contextualised?
- Is the instruction clearly worded, easy to follow?
- Is there an appropriate task focus for the discussion activity?
- Are the prompts suitable both in terms of quality and quantity?

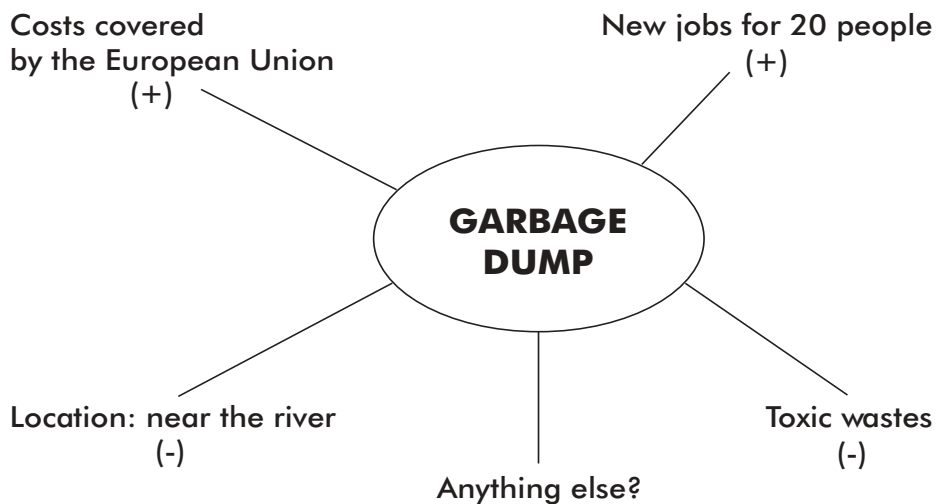
Since *Paired Discussion Activity 5.A* also went through the regular procedure for high quality item production (item moderation, editing, piloting and revision), you can check how the format of the task changed and what the final version, used in

mock exams, looks like (see *Paired Discussion Activity 5.B* below). You can also judge the appropriacy of the modifications for this task by viewing real students performing it on DVD Sample 5.1.

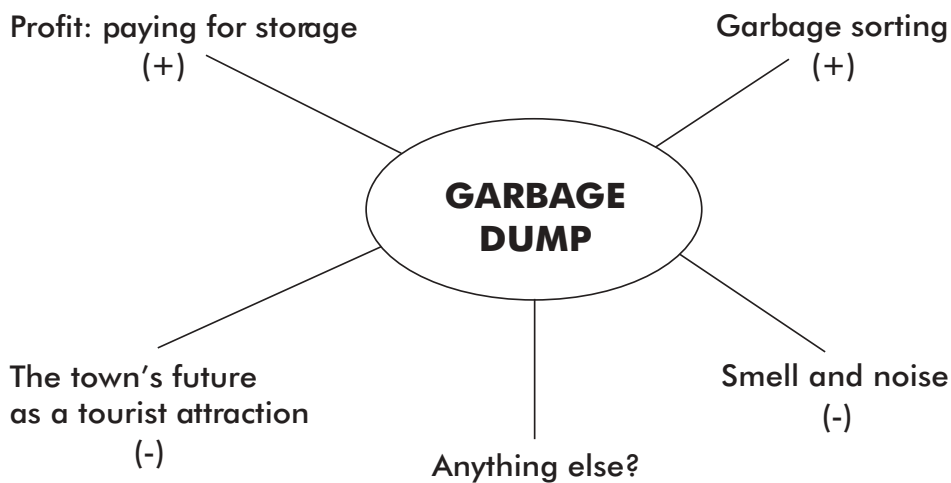
PAIRED DISCUSSION ACTIVITY 5.B (FINAL REVISED VERSION)

The local government is planning to build a garbage dump where the rubbish of your region will be placed. Discuss the advantages and disadvantages of building this regional garbage dump and try to come to an agreement. On your sheets, there are some ideas but you can suggest other ones, too.
Start when you are ready.

Prompts for Candidate A



Prompts for Candidate B



Based on the lessons learnt from piloting a large number of paired activities, we recommend the following set of guidelines for designing paired discussion activities. The guidelines were originally written to help secondary school and university teachers who had been trained in language testing to write test items for the Hungarian Examinations Reform Project. We believe that they are also useful for English teachers in general, even if they have not received specific training in language testing, to help them write better speaking tasks for classroom assessment purposes.

GUIDELINES FOR WRITING PAIRED TASKS

Item writers should bear in mind that

- the reading input given on the candidate's task sheet should be minimal as the exam is not intended to measure candidates' reading ability (use word prompts instead of full sentences, maximum 30 words);
- the task instructions should be read out by the Interlocutor and so they cannot be given on the Task Sheet for the candidates (they should appear on the Interlocutor's Sheet only);
- candidates' potential contributions to the interaction should be balanced, which could be achieved by giving an equal number of visual or word prompts for both candidates;
- candidates should have comparable tasks, i.e. both of them should be required to do the same thing (to list, compare, contrast, select, justify, modify, extend, reduce options, etc.) in order to facilitate a balanced, realistic and smoothly-running exchange between them;
- the interaction between the candidates should be task-based as it seems to give them a meaningful purpose to engage in a conversation (e.g. select the three most/least important aspects of something, rank order items in a given list);
- the tasks have to be guided but not fully controlled, i.e. candidates should have a chance to add something of their own to the exchange (use the word prompt *Anything else?* to indicate that candidates can add their own ideas);
- controversial topics seem to be more likely to generate a discussion than neutral topics;
- distressing, offensive, violent or taboo topics should be avoided;
- the topic should be something on which it is reasonable to expect candidates to have an opinion;
- in order to make the task realistic for candidates, they should define the context of the conversation very carefully, but at the same time they should not force candidates to agree with a point of view that they may not accept in real life. Candidates should be given a chance to voice their own opinion rather than argue for something that they cannot identify with.

In paired discussion activities, the examiner-interlocutor also has a specific role, although his/her presence is much less conspicuous than in the individual mode. Still, it is very important to specify the examiner-interlocutor's potential contributions very carefully as s/he is the only participant who can take action if there is a breakdown in communication for some reason, or when one candidate tries to dominate the conversation.

There may be many reasons why an examiner has to intervene. For instance, candidates may misunderstand the task and so they need to be told again what to do. In other cases, they may forget that they have to reach agreement in relation to the given task (e.g. they need to agree on the three most useful aspects of a given issue) and thus the examiner has to remind them to do so. Some candidates may finish the task too soon and so the interlocutor should indicate to them that they have to continue the discussion because they have not produced an adequate language sample for assessment yet.

DVD Sample 5.2 shows a paired task performance in which the examiner-interlocutor was not required to observe specific guidelines for intervention. The task required the two candidates to discuss what they would like to do together at the weekend. They were given word prompts to focus their discussion (e.g. where to go, what to do, who to invite). Watch the DVD to find out how the candidates' performance was affected by the lack of guidelines for interlocutor behaviour. The paired task was intended to take 5-6 minutes.

Consider the following observation points before you watch the performance:

- How long is the candidates' performance?
- Do candidates produce enough language for a reliable assessment of their abilities?
- Does the interlocutor's contribution facilitate candidates' performance successfully?

As was pointed out in Chapter 1, the reliability of speaking examinations can be enhanced by standardizing the elicitation procedure. This means that interlocutor intervention also has to be clearly specified so that interlocutors can behave in a standard manner in case there is some kind of problem with the candidates' performance. We recommend the following set of guidelines for interlocutor intervention in paired speaking tasks.

GUIDELINES FOR INTERVENING IN PAIRED- TASK PERFORMANCES

It is the interlocutor's duty to intervene if

- there is a communication breakdown;
- there is imbalance between the two candidates' contributions;
- the amount of language produced by the candidates is insufficient.

In such cases the interlocutor should

- repeat all or part of the rubric;
- invite candidates to talk about one specific aspect of the task;
- invite the candidate whose contributions seem to be unsatisfactory (i.e. too short or incomplete) to talk about one specific aspect of the task or to elaborate on something s/he said.

The Hungarian Examinations Reform Project has tried out a number of paired discussion activities with real students in mock exam situations. We recommend the following tasks as appropriate speaking activities to assess candidates performing in pairs. You can also judge how these tasks work as each of them can be seen 'in action' on the DVD. The DVD samples also demonstrate how the above guidelines for interlocutor intervention can be put in practice.

PAIRED DISCUSSION ACTIVITY 6.

(For a sample candidate performance see DVD Sample 5.3)

In the summer, you would like to work together for a month. On your sheets, there are some possible summer jobs but you can suggest other ones. Talk to each other about the good and bad points of the jobs and then try to agree which are the three jobs that you would prefer to do. Start when you are ready.

Prompts for Candidate A & B



PAIRED DISCUSSION ACTIVITY 7.

(For a sample candidate performance see DVD sample 5.4)

Imagine that in your school the students would like to go to free afternoon classes. Talk to each other about what courses the school should organize. On your sheets, there are some ideas but you can suggest other things, too. Talk about how useful these courses would be for students at your school and try to agree about the three most useful ones.

Start when you are ready.

Prompts for Candidate A

Driving

Photography

Dress-making

Word-processing

Anything else?

Prompts for Candidate B

Typing

Aerobics

Using the Internet

Anything else?

Self-defence

PAIRED DISCUSSION ACTIVITY 8.

(For a sample candidate performance see DVD Sample 5.5)

Imagine that the two of you are editing a teenage magazine. In the next issue, there will be articles about love. You have received these four pictures for the cover of the next issue. Discuss which photo should be the cover picture.



PAIRED DISCUSSION ACTIVITY 9.

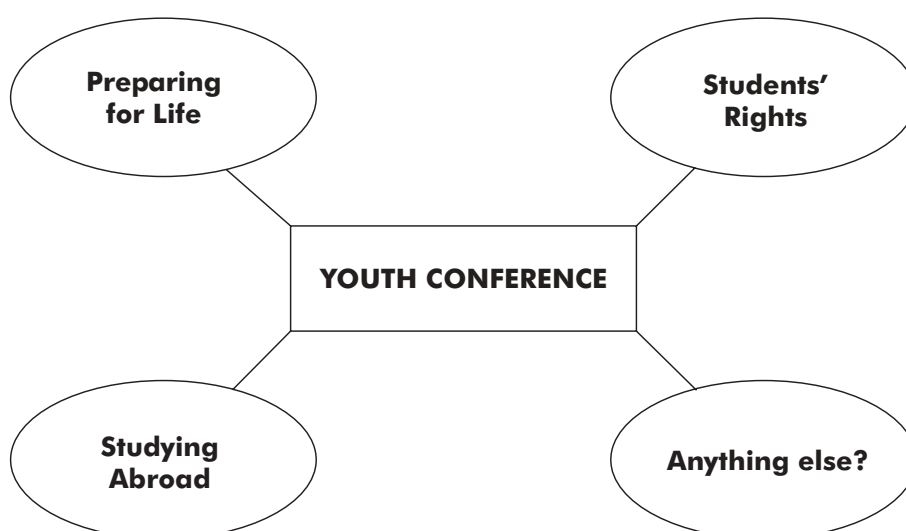
(For a sample candidate performance see DVD Sample 5.6)

You have been invited to an International Youth Conference where you will represent Hungarian secondary school-leavers. On your sheets, there are topics that the organizers would be interested in. Discuss why these issues are important for school-leavers in Hungary, and decide on the two most important issues. Start when you are ready.

Prompts for Candidate A



Prompts for Candidate B



Discussion tasks in the individual mode

Discussion activities with an examiner-interlocutor do not seem to be capable of replicating real-life interaction in the same way as paired discussion activities. The reason for this is that the examiner's scope for contribution to the exchange must be limited, especially in terms of what s/he can add to the discussion. We have already pointed out that the interlocutor cannot hold the floor for a long time as that will automatically reduce the time for the candidate to perform. The interlocutor also has to contribute to the interaction in such a way that encourages the candidate to justify and argue for his/her views.

In order to facilitate a meaningful exchange between the interlocutor and the candidate, we recommend that the interlocutor should assume a specific role while the candidate may act as him/herself, expressing his/her true opinion. The interlocutor can play familiar roles, such as a neighbour or a foreign visitor, as well as less familiar roles, for instance a representative for an organization or a counsellor. It is essential that the discussion activity is designed in such a way that there is a clear opinion gap between the examiner and the candidate. This can be easily ensured by assigning roles to the participants that require them to represent different points of views or to argue for different things. Since the candidate and the examiner cannot be expected to come to an agreement automatically or continue the discussion for long without showing willingness to accept the partner's point of view or reasoning, it should also be clearly specified in the examiner's role description what the outcome of the exchange should be. For example, the instructions for the interlocutor may suggest that s/he should agree fully or partially with some of the candidate's arguments.

While the interlocutor's contributions should be kept to the minimum, s/he should always try to challenge the candidate as much as possible in order to elicit performance in L2 from the candidate to a maximum degree. The prompts for the interlocutor may instruct him/her to

- ask for further justifications from the candidate;
- argue against what the candidate has presented/suggested;
- or to present opposite points of views.

Candidates rarely have a chance to start the conversation in exam situations. Therefore, if possible, they should be instructed to do so in discussion activities in the individual mode. Because the interlocutor is responsible for eliciting language from the candidate, naturally it will be his/her duty to decide when to close the conversation.

As in the case of paired discussion activities, item writers should also bear in mind that the reading input given on the candidate's task sheet should be minimal, and so word prompts should be used instead of full sentences (maximum 30 words). While it is clear that discussion tasks in the individual mode have to be guided, they should not be fully controlled: candidates should have a chance to add something of their own to the exchange. The phrase "*Anything else?*" will indicate to the candidates that they can add their own ideas.

The next two discussion tasks (for the individual mode) are intended as examples to show how the item writing guidelines described above can be applied. The tasks have been piloted in mock exams and there is a sample candidate performance available on DVD for the second task.

DISCUSSION TASK 1 **(FOR THE INDIVIDUAL MODE)**

To be read out for the candidate:

I am a foreign journalist. I am interested in Hungarian young people's film watching habits. Give me your opinion.

On your sheet, there are some ideas but you can suggest other ones too.

You must start the conversation when you are ready.

Prompts for the candidate:

- Cinema or video?
- Type of film
- With whom?
- Language of the film
- Anything else?

For the examiner only:

Ask for the candidate's opinion about

- showing violence on screen
- film adaptations of books

Challenge the candidate in a polite manner so that s/he should be forced to defend his/her arguments.

You must finish the conversation.

DISCUSSION TASK 2

(FOR THE INDIVIDUAL MODE)

(For a sample candidate performance see DVD Sample 5.7)

To be read out for the candidate:

I am your foreign neighbour. We live in a small town in Hungary. The local government is planning to build an international airport in our area. Let's discuss the advantages and disadvantages of building the airport.

**On your sheet, there are some ideas but you can suggest other ones too.
You must start the conversation when you are ready.**

Prompts for the candidate:

- Tourism
- Employment
- Environment
- Anything else?

For the examiner only:

Challenge the candidate in a polite manner so that s/he should be forced to defend his/her arguments.

You worry about an increase

- in the crime rate
- the negative effects on the environment

You are happy that local people will have better transport facilities.

You must finish the conversation.

In this chapter, we have presented good and bad examples for the discussion technique, and have suggested guidelines for designing tasks for both the paired and the individual mode. In the next chapter, we will discuss another popular elicitation method, the role-play, which is also aimed at measuring candidates' interactional ability. The two techniques have a lot in common but we will also consider the differences between the two.

Chapter 6: Role play Tasks

The role-play technique is widely used for assessing oral interaction skills, but the term itself has quite a variety of interpretations and applications. At one end of the scale it refers to highly controlled or semi-guided dialogues, while at the other end it includes free, improvised activities with no guidance whatsoever. The answer to the question why role-plays are suitable for measuring oral interactional skills can be found in the nature of the roles themselves. The roles featuring in examination tasks usually simulate the ones we take in our everyday lives. Most roles are reciprocal (student – teacher, child – parent, customer – shopkeeper, employer – employee, etc.). The reciprocity of roles provides great opportunities for life-like interactions in role-play tasks.

The role-play technique and the discussion tasks covered in *Chapter 5* overlap in some of their features. Both task types involve the participants in completing tasks such as comparing, contrasting, listing, rank ordering, redesigning, planning, categorising, problem solving or selecting, expressing preference. During the test participants may use language functions like asking for and giving information; expressing emotions as well as likes and dislikes; asking for and giving opinions, etc. A number of social formulae may also be used such as opening and closing a conversation; congratulating; saying sorry, etc.

However, we can observe significant differences as well. While in the discussion tasks candidates are generally encouraged to express their own opinion, in role-plays the candidate often takes somebody else's role in order to reach a particular communicative goal through interaction. In well-designed role-play tasks the difference between the social roles and status of the examiner-interlocutor and the candidate, mentioned in *Chapter 3* in connection with the interview and in *Chapter 5* in connection with the discussion tasks, can be reduced. Thus, the flow of interaction in role-plays may run more smoothly and resemble natural conversations more than in the above mentioned two techniques. However, superiority in terms of age, language proficiency and authority (due to the examiner role, which is difficult to ignore even in a role-play situation) will still prevail in the majority of cases, and potentially limit the candidate's language output for psychological reasons.

In this chapter we will first focus on how to design paired role-play tasks. Since candidates are often asked to perform role-play tasks with an examiner-interlocutor, in the second part of this chapter we will demonstrate and analyse role-play tasks in the individual mode as well.

Paired role-play tasks

Similarly to paired discussion tasks, it is essential that paired role-play tasks relate to and simulate real-life situations and provide both candidates with equal opportunities to contribute to the discussion. Successful contribution can be facilitated by the following task design strategies:

STRATEGIES FOR TASK DESIGN IN ROLE-PLAY TASKS

- ✓ Give candidates **familiar** roles.
- ✓ Give them a standardised, clear and concise role description.
- ✓ Provide word or picture prompts which represent the ideas they have talk about according to their roles.
- ✓ Instruct them about their expected contributions to the role-play situation.

Role-play tasks must have a clear focus, otherwise they will not work at all. Each task should have a final goal towards which candidates work throughout the role-play task. This means that candidates should be asked to complete some kind of outcome-oriented task while they are talking. However, assessment should always concentrate on the process of the interaction instead of the result. Information and opinion gap activities provide the best opportunities for life-like situations. *Role-play Task 1* is one of them.

ROLE-PLAY TASK 1: CAMPING HOLIDAY

Interlocutor's Sheet

Camping holiday

You would like to go camping together. Discuss the details of your holiday.

On your task sheets there are some ideas. You must discuss these but you can also discuss others.

Start when you ready.

Prompts for Candidate A

You would like to go to the mountains.

You have a sleeping bag but no rucksack and tent.

Discuss:

- food
- what to do
- when to go
- how to go
- anything else?

Prompts for Candidate B

You would like to go to a lake.

You have a rucksack, but no tent and sleeping bag.

Discuss:

- food
- what to do
- when to go
- how to go
- anything else?

Role-card for Candidate A

Camping holiday

You would like to go to the mountains.

You have a sleeping bag but no rucksack and tent.

Discuss:

- food
- what to do
- when to go
- how to go
- anything else?

Role-card for Candidate B

Camping holiday

You would like to go to a lake.

You have a rucksack, but no tent and sleeping bag.

Discuss:

- food
- what to do
- when to go
- how to go
- anything else?

The situation is clearly set, and the candidates take a very familiar role, they remain themselves. The task uses word prompts, which are short and clear. The role-play is based on an opinion gap, the candidates have different ideas about where to go camping.

ROLE-PLAY TASK 2: PEN-FRIEND

(For a sample candidate performance see DVD Sample 6.1)

Role-card for Candidate A

You and your friend (the other candidate) would like to take your pen-friend 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 pen-friend 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.

Role-card for Candidate B

You and your friend (the other candidate) would like to take your pen-friend 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 pen-friend 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.

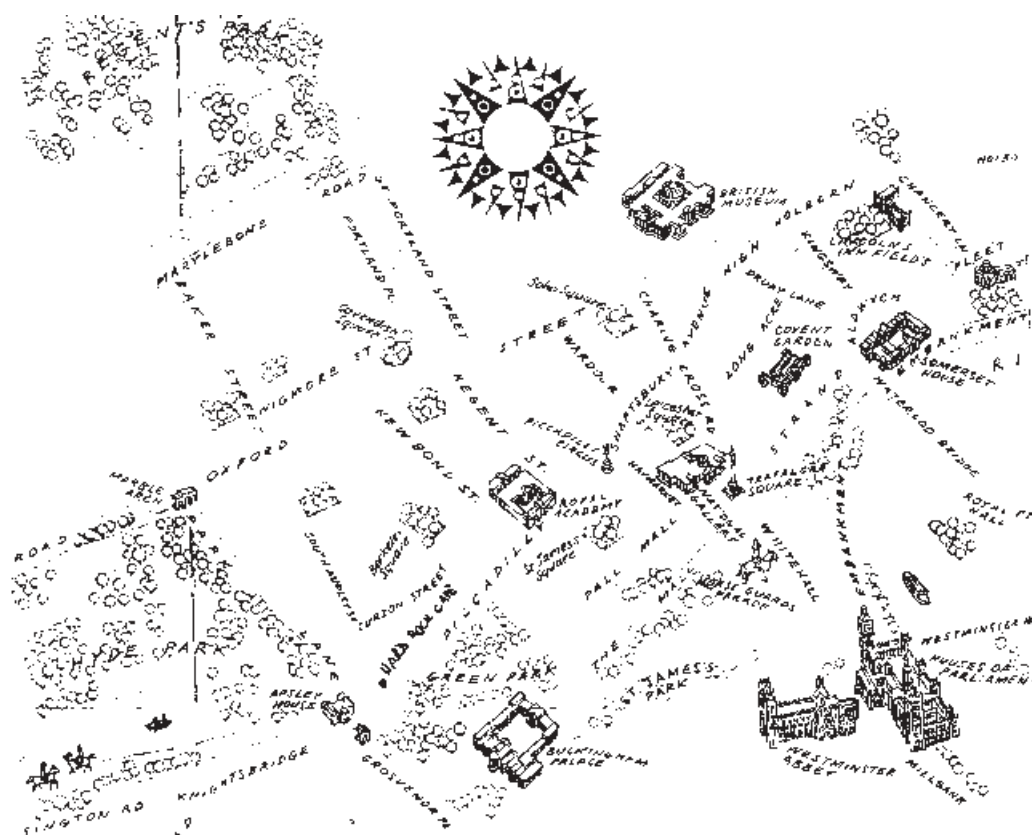
Role-play Task 2 is an example of over-guiding a role-play task. The situation is life-like, candidates do not have to take unfamiliar roles, there are information and opinion gaps in the task (difference between candidates' likes and dislikes and plans) to elicit enough language output. The sample video shows two good candidates who could have done much better if there had been more freedom in the task, if they had been allowed to contribute to the discussion with their own ideas as well. Because of the lack of an Interlocutor Frame, the Interlocutor does little and the candidates spend time reading the instructions. In good tasks this is not the case, since the Interlocutor reads out the standardised instructions.

ROLE-PLAY TASK 3: MAP READING

(For a sample candidate performances see DVD Samples 6.2 & 6.3)

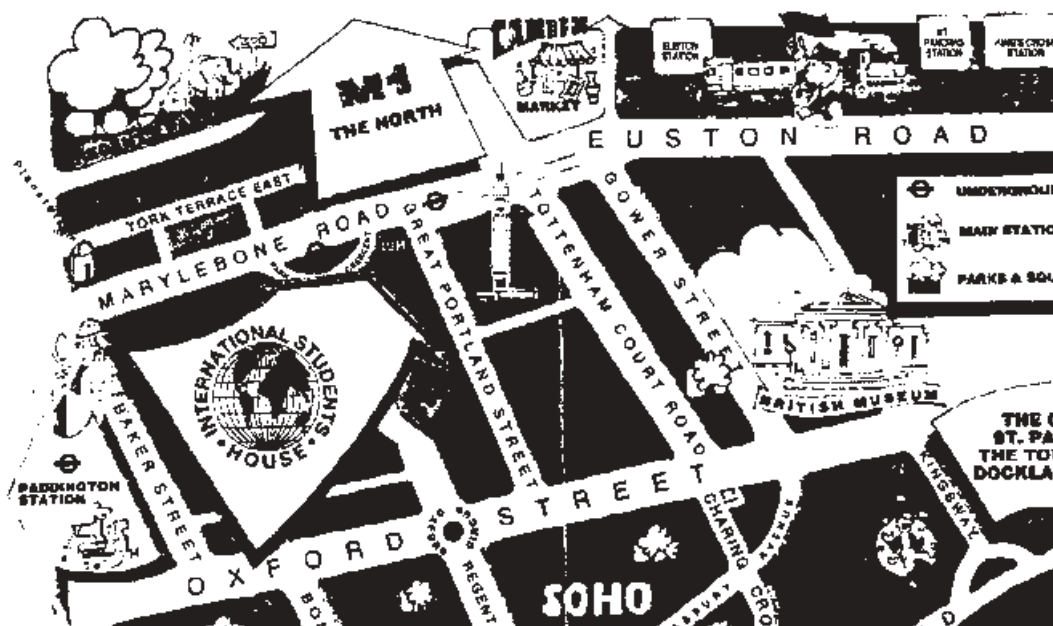
Role-card for Candidate A

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.



Role-card 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.



Role-play Task 3 risks assessing a cognitive skill (the ability to read maps) instead of focusing on language. This cognitive skill goes far beyond language skills, and not everyone possesses it to the same extent. The situation itself is realistic, but because of task design problems it fails as a role-play task. The flow of interaction is one-way because Candidate B has all the necessary information, while Candidate A is in the dark. Thus the roles are unbalanced: Candidate A is reduced to asking very limited questions, while Candidate B gives basic replies. The prompts, which are maps of London, make the situation difficult, since they are not identical and not very clear. There is no Interlocutor Frame and so the interlocutor feels obliged to explain aspects of the task to the students when they show that they find the map difficult to read and are not sure what to do. Even once they have understood the task, as Sample 6.2 shows, Candidate B says much more than Candidate A. Weak candidates really suffer during the task, and – as you can see from DVD Sample 6.2 – their language output is minimal. Even in the case of good candidates (see DVD Sample 6.3), the interaction and language output are one-sided. This kind of task would be suitable for a listening test perhaps, though the quality and the complexity of the map prompt might still be problematic.

ROLE-PLAY TASK 4: A HUNGARIAN SOUP SPECIALITY

(For a sample candidate performances see DVD Samples 6.4 & 6.5)

Instructions for Candidate A (in Hungarian):

Önt felhívja ismerőse, akit tegnap látott vendégül, és megköszöni a vacsorát Önnek. Ugyanakkor megkéri, mondja el a receptjét annak, amit főzött, mert mindenkinek nagyon ízlett.

Translation of instructions:

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:

Verbal prompt for Candidate A:

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 felforralni. Ekkor a tejszín(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 feforrósítjuk. Ha kissé kihűlt, turmixoljuk, átszűrjük és teljesen lehűtjük.

A maradék szőlőszemeket beletéve tálaljuk.

Instructions for Candidate B (in Hungarian):

Ön felhívja barátját, és megköszöni neki az előző napi vendéglátást. Kérdezze ki, hogy készül az a leves, amit tegnap felszolgált, mert mindenkinek nagyon ízlett.

Translation of instructions

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.

The task was intended to measure a student's ability to mediate from one language to another, where it was expected that Candidate A would summarise how to make the soup, based on the recipe. Originally, the task was accompanied by a visual prompt as well (a picture of the Hungarian soup speciality called *Tokaji Szőlőleves*), and the task was tried out with some candidates who were given the instructions in English while some other candidates got the instructions in Hungarian. It was found that the wording of the rubric in L2 as opposed to L1 did not make a big difference as the test takers seemed to understand what they were required to do. However, the task prompt for Candidate A, the recipe in Hungarian, caused serious problems for some test takers. Unfortunately, as can be seen in DVD Sample 6.4, Candidate A simply tried to translate the recipe word for word (possibly because of the difficult vocabulary in the recipe). But the major problem with this task was that it was not suitable for the paired mode as the roles were not equally balanced and so one candidate ended up dominating the interaction (see DVD Sample 6.5 as well).

ROLE-PLAY TASK 5: EXCHANGING HOLIDAY TRIPS FOR SPAIN

(For a sample candidate performance see DVD Sample 6.6)

Instructions 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.).

Your hotel is the following:

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.

Instructions 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, disco, 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 k(nálata: 4 tenispálya, minigolf, asztalitenisz, biliárd, röplabda, futball, fitness-szoba.

CLUB ESPANA TIPP: A sziget egyik legattraktívabb szállodája.

Task 5 had similar shortcomings to Task 4 as the input was lengthy and most candidates wanted to give all the details of the hotel on their task sheet. This meant that the interaction was far from being dynamic. Less able candidates simply tried to translate the whole advertisement and did not attempt to discuss the advantages and/or disadvantages of the two hotels not to mention giving advice. Although more proficient candidates seemed to cope with the task better, their interaction was still not so dynamic as a result of the amount of information each candidate was expected to process and summarise for their partners. A common problem for all candidates was that the advertisements included foreign names such as Lido Llobet, Figueretas, Club Goleta and Plaza d'en Bossa, which turned out to be difficult for candidates to pronounce. DVD Sample 6.6 shows the problems that two somewhat weaker candidates had in coping with the task.

The following two samples (Role-play Task 6 and Role-play Task 7) will demonstrate how badly designed speaking tasks can be improved by careful revision.

ROLE-PLAY TASK 6: COMPUTER CAMP (ORIGINAL VERSION)

Role-card for Candidate A

On your flight back to Hungary from the States you meet another student. After some introductory polite phrases you end up exchanging information about the camp you have been to.

Mention facts and/or your opinion about:

- place, location
- type of camp
- types of activities
- your special interest in computer sciences

The information in the following advertisement will help you:

Computer-Ed High-Tech Camps High-Tech Learning at its Best!

Boston, San Fransisco & Chicago Areas

Internet & World Wide Web * Build & Repair a PC * Windows 95 *
Networking * RC Cars * Computer Arts & Graphics * CADD *
Image Processing * Animation * Web Broadcasting * Digital Photography *
Recreation & Sports Activities

Call 1-888-2 COMPED
Out of U.S.A. (1-781-933-7681)
Email: camp@computered.com
URL: www.computered.com

Computer-Ed Camps Trade Center Park
100 Sylvan Rd. G 500 Woburn, MA 01801

Role-card for Candidate B

On your flight back to Hungary from the States you meet a student who has been to an interesting camp. After some introductory polite phrases you end up exchanging information about the camp.

Ask him/her about:

- location of the camp
- ways of application or getting information
- any special requirements
- his/her special interest in computer sciences

In *Role-play Task 6* we would like to demonstrate that bad task design has a negative effect on the interaction. In the original version of the task only Candidate A has a text prompt (and advertisement). This prompt is very long and full of information. It takes a long time for the candidate to read. The other candidate has only the role-card. The rubrics are unnecessarily complicated and are not standardised. The candidates' roles are unequal and unbalanced, since one of them has all the information, while the other is reduced to a role in which the only type of contribution to the interaction can be questions. The aim of the role-play is not clear: the expression "*exchanging information*" in the rubrics is misleading, since one of them has all the information, while the other has nothing. The candidate who has the prompt is in an advantageous position.

ROLE-PLAY TASK 6: COMPUTER CAMP (REVISED VERSION)

Interlocutor's Card

A classmate of yours has been to an interesting camp and has told you about it in the English lesson. You cannot remember all the details. Ask your partner about the missing information and tell him/her what you know. Together you can collect all the information. Study the information on your card.

(after some seconds) *Let's start now.*

Role-card for Candidate A

Use the information in the following advertisement :

Computer-Ed High-Tech Camps

High-Tech Learning at its Best!

* Build & Repair a PC *

* Digital Photography *

Recreation & Sports Activities

Ask about: – location of the camp
– ways of application or getting information on the phone or in any other form
– the exact address

Role-card for Candidate B

Use the information in the following advertisement :

Boston, San Fransisco & Chicago Areas
Call 1-888-2 COMPED
Out of U.S.A. (1-781-933-7681)
Email: camp@computered.com
Computer-Ed Camps Trade Center Park
100 Sylvan Rd. G 500 Woburn, MA 01801

Ask about: – name of the camp
– type of camp
– special activities and spare time activities

In the revised version of the task the prompt is shortened, and there is a much more balanced information gap to fill in for the candidates. The rubrics are simplified and read out by the Interlocutor, the role-cards have the prompts. The information in the advertisement is divided between the two candidates, and the task requires a sort of jigsaw activity to fill in the gaps. The situation is life-like, the candidates take roles which are familiar.

ROLE-PLAY TASK 7: HAMPTON PLAYHOUSE (ORIGINAL VERSION)

Role-card for Candidate A

You would like to go to a summer camp in the USA. You have some information about a special camp and your friend has the rest of it. Tell him/her what you know

and ask him/her about: – activities in the camp

– performances

– accommodation and facilities

– ways of applying and getting more information

You can use the information in the following advertisement:

THE HAMPTON PLAYHOUSE THEATER ARTS WORKSHOP

(In Hampton, New Hampshire, on the beautiful
New England Seacoast) **Co-ed 13 through 18**

The most dynamic theater training program in America. Operates with
Hampton Playhouse the nationally famous professional union
theater, now in its 50th year.

Structured and supervised, but never regimented.

Sessions: 3 sessions available. One 8-week session
(June 28 to August 22)

Two 4-week sessions (June 28 to July 25 and July 26 to August 22)

Role-card for Candidate B

You would like to go to a summer camp in the USA. You have some information about a special camp and your friend has the rest of it.

Tell him/her what you know

and ask him/her about:

- location of the camp
- age of participants
- organising institution
- time and length of camp

You can use the information in the following advertisement:

Daily classes: Acting (modern, classical, scene study), directing, singing, voice and speech, dance, body movement, body building, set and costume design and construction, etc.

Productions: Each week on the professional stage, two full-scale performances of different plays for large, public audiences

Superb accommodations: Carpeted motel units fronting an Olympic-sized pool, private dining room, scene shop, costume shop, recreation room.

All sports available.

Write: A.H.Christie, 405 East 54th St, New York, NY 10022

Call: (212) 759-7977

In *Role-play Task 7* a similar task design problem occurs. In the original version of the task the text prompt (and authentic advertisement) is too long, over-packed with information. The advertisement is divided between the two candidates, and the task requires a sort of jigsaw activity to obtain the missing bits of the information. The situation, however, is very artificial. The rubrics are too complicated and are not standardised.

ROLE-PLAY TASK 7: HAMPTON PLAYHOUSE (REVISED VERSION)

Interlocutor's Card

You have heard about an interesting camp in Hampton, New Hampshire, USA. You know some of the details and are interested in other details. Please ask your partner what he/she knows and answer each other's questions. Before you start you can read the information on your card.

(after some seconds) Let's start now.

Role-card for Candidate A

Use the information in the following advertisement:

THE HAMPTON PLAYHOUSE THEATER ARTS WORKSHOP

on the beautiful New England Seacoast
Co-ed 13 through 18
8-week session starting June 28
4-week sessions starting June 28 or July 26

Ask your partner about: – activities in the camp
– accommodation and facilities
– ways of applying and getting more information

Role-card for Candidate B

Use the information in the following advertisement:

Daily classes.

Productions: Each week on the professional stage
Superb accommodation: fronting an Olympic-sized pool,
All sports available.

Write: A.H.Christie, 405 East 54th St, New York, NY 10022
Call: (212) 759-7977

Ask your partner about: – location of the camp
– age of participants
– organising institution
– time and length of camp

In the revised version of the task the prompt is shortened, and there is a much more manageable information gap to bridge for the candidates. The rubrics are simplified and read out by the Interlocutor. The situation is more life-like, as the candidates take roles which are familiar.

The following task demonstrates various task design problems.

ROLE-PLAY TASK 8: AU PAIR PLUS

Role-card for Candidate A

You are aupairing in England and owing to changes in the host family, you are no longer required. You see the following advert in a magazine and telephone to inquire.

AU PAIR PLUS, ISLINGTON. – Start
immediately. Look after girl 11, house, cats.
Must swim. Live in. – Tel. 0171 226 5711
after 7 pm.

Introduce yourself. Ask about your tasks, about the girl and about the cats. Ask about wages, free time and the accessibility of the city centre, receiving visitors etc.

Role-card for Candidate B

You are Mrs Richards. You put the following advert into a magazine.

AU PAIR PLUS, ISLINGTON. – Start
immediately. Look after girl 11, house, cats.
Must swim. Live in. – Tel. 0171 226 5711
after 7 pm.

You have a 4-bedroom house with a garden. You need help with cleaning and ironing. No cooking required. Your daughter is well-behaved. She enjoys swimming every day and likes playing in the water with a companion. You have 3 cats who are friendly. You pay £60 per week and provide a private room with shower. Weekends are free, plus Thursday afternoons. Easy access to the city by public transport. No visitors allowed.

Role-play Task 8 demonstrates unsuccessful task design for several reasons. While the role for Candidate A is realistic, the other candidate has to take an unfamiliar role: it is very unlikely that a young examinee will have experienced being an employer. The layout of the rubrics does not help the candidates to remember what kind of contribution is expected from them throughout the role-play. Bullet points would be much more suitable than continuous text. The roles are extremely guided, especially for Candidate B. This task is very unlikely to produce a successful life-like interaction.

Recommended role-play tasks in the paired mode

The Hungarian Examinations Reform Teacher Support Project has tried out a number of role-play tasks in the paired mode with real students in mock exam situations. In addition to *Role-play Task 1*, we recommend the following tasks as appropriate speaking activities to assess candidates performing in pairs with an interlocutor.

ROLE-PLAY TASK 9: NEW NEIGHBOURS

Interlocutor's Sheet

Your families have recently moved into a new block of flats. You have become next door neighbours. You meet for the first time in the corridor. Try to get to know each other.

On your task sheets there are some further ideas.
Start when you are ready.

Prompts for Candidate A

- Greet your new neighbour.
 - Introduce yourself.
 - Describe your old home.
 - Discuss why you like the new flat/neighbourhood.
 - Invite your neighbour to look at your room.

Prompts for Candidate B

- Greet your new neighbour.
- Introduce yourself.
- Describe your old home.
- Discuss why you like the new flat/neighbourhood.
- Invite your neighbour to a party next week.

Role-play card for Candidate A*New Neighbours*

- Greet your new neighbour.
- Introduce yourself.
- Describe your old home.
- Discuss why you like the new flat/neighbourhood.
- Invite your neighbour to look at your room.

Role-play card for Candidate B*New Neighbours*

- Greet your new neighbour.
- Introduce yourself.
- Describe your old home.
- Discuss why you like the new flat/neighbourhood.
- Invite your neighbour to a party next week.

ROLE-PLAY TASK 10: AT A BIRTHDAY PARTY

Interlocutor's Sheet

Imagine that both of you are at a friend's birthday party. You have just noticed each other. Discuss how you like the party.

On your task sheets there are some further ideas.

Start when you are ready.

Prompts for Candidate A

- Greet your friend.
- Say what you like about the party.
- Complain about the music.
- Describe the best party you have been to.
- Discuss how to make the party better.

Prompts for Candidate B

- Greet your friend.
- Say what you hate about the party.
- Complain about the food.
- Describe the best party you have been to.
- You would like to leave soon.

Role-play card for Candidate A

At a Birthday Party

- Greet your friend.
- Say what you like about the party.
- Complain about the music.
- Describe the best party you have been to.
- Discuss how to make the party better.

Role-play card for Candidate B

At a Birthday Party

- Greet your friend.
- Say what you hate about the party.
- Complain about the food.
- Describe the best party you have been to.
- You would like to leave soon.

ROLE-PLAY TASK 11: LOST KEYS

Interlocutor's Sheet

Imagine that you (Candidate A) have invited your friend (Candidate B) to your home to watch a video together. It is 3 o'clock in the afternoon and nobody is at home yet. You are standing in front of your door and you (Candidate A) cannot find your keys.

On your task sheets there are some further ideas.

Start when you're ready.

Prompts for Candidate A

- say where the keys may be
- discuss what to do:
 - call somebody?
 - wait?
 - anything else?
- discuss when/how to watch the video

Prompts for Candidate B

- help your friend to find the keys
- describe what you did in a similar event
- discuss what to do:
 - go somewhere?
 - anything else?
- discuss when/how to watch the video

Role-play Card for Candidate A

Lost Keys

- say where the keys may be (lost? at home?)
- discuss what to do:
 - call somebody?
 - wait?
 - anything else?
- discuss when/how to watch the video

Role-play Card for Candidate B

Lost Keys

- help your friend to find the keys (bag? pockets?)
- describe what you did in a similar event
- discuss what to do:
 - go somewhere?
 - anything else?
- discuss when/how to watch the video

Role-play Tasks in the Individual Mode

In role-play tasks in the individual mode the candidate is always expected to play familiar roles only. The examiner can play familiar as well as less familiar roles. Role-play tasks should always have a clear information and/or opinion gap between the examiner and the candidate. This can only be achieved by designing the roles of the two participants in such a way that they necessarily represent different points of views. The interlocutor's role description should clearly specify what the outcome of the exchange should be. The interlocutor can be asked to agree/disagree fully or partially with the candidate's arguments.

While the interlocutor's contributions should be kept to the minimum (since it is the candidate whose language product is assessed), s/he should always try to challenge the candidate as much as possible in order to elicit the maximum possible assessable language from the candidate. The prompts for the interlocutor can instruct him/her to ask for further justifications from the candidate; to argue against what the candidate has suggested; or to present opposite points of views.

Role-play tasks should be guided but not fully controlled, i.e. candidates should have a chance to add something of their own to the exchange. It is always the candidate who has to open the conversational exchange while it is the interlocutor's duty to close it.

Role-play tasks conducted in the individual mode with an examiner-interlocutor can replicate real-life interaction to a far less extent than the ones conducted in the paired mode. In case of standardised examinations the interlocutor's contributions to the exchange are closely guided and limited. They cannot hold the floor for long as that will automatically reduce the time for the candidate to perform. The interlocutor also has to contribute to the interaction in such a way that elicits assessable language output during life-like interaction from the candidate.

The following sample tasks demonstrate some typical task design problems.

ROLE-PLAY TASK 12: HOMEWORK

Interlocutor's Card

Read your role-card, then begin a conversation with the candidate.

Role-card

You had no time in school today to write down the homework questions. It is now ten fifteen. Call your friend and ask him/her to dictate the questions to you. You have to go to bed at eleven o'clock. (Tomorrow you will have to get up early.)

Role-card for Candidate

Read your role-card, then play a conversation with the examiner.

Role-card

A classmate calls to get the homework. You are expected to dictate ten questions to him/her. You are in the middle of watching an exciting TV programme, and you don't want to miss any of it. The programme is over at eleven o'clock. Try to suggest some solutions.

In *Role-play Task 12* the interlocutor has to take a rather unnatural role, s/he has to play the role of a student (differences in age and social status). The role-play situation would be acceptable (there is an information gap and a difference between the participants' plans) if it was not a telephone conversation with participants facing each other. It is really difficult to simulate talking on the phone in such a situation, when the examination environment adds to the fact that it is awkward to pretend that one is talking on the phone while facing the partner. The interlocutor's contributions are not guided enough, there is no script for the interlocutor controlling who starts and who finishes the conversation, and what kind of contribution the interlocutor is supposed to make. In this way the interlocutor is very much left alone and has to rely on imagination and previous experience in conducting role-play tests. Needless to say that this will not result in a standardised performance and equal treatment for each candidate.

ROLE-PLAY TASK 13: KEEPING A PET

Role-card for Candidate

You would like to have a pet, preferably a dog. Unfortunately your mother is very much against this idea since you live in a flat, which is quite small and might not be the ideal living place for an animal. Try to convince her that keeping a pet (dog) has its advantages. Below you will find a few ideas you might use.

- protection
- keeping fit – walking
- sense of responsibility
- other friends
- sharing of work
- future possibility of moving into a house

Interlocutor's Card

Use the following ideas as appropriate.

- ♦ *Our flat is small enough without an animal and it would be a torture for the poor thing, too.*
- ♦ *Even if you walk the dog there's so much more to be done, you have to feed him, take him to the vet, etc.*
- ♦ *Yes, it's true that keeping a pet would develop a very strong sense of responsibility, but are you sure you could cope with it? We couldn't travel anywhere without it, we would have to plan all our lives differently, we would be the slaves of that animal.*
- ♦ *We already have two locks on the door and the flat is insured.*
- ♦ *Because your friends live in houses. It's totally different there.*
- ♦ *OK, that seems a reasonable compromise. If we move into a house, you can get a dog. Until then you can still have goldfish or a turtle.*

The above role-play situation is realistic: many families have similar problems. The roles are well-designed, the candidate has to take a very familiar role, while the interlocutor's role is unfamiliar, but suitable for the situation.

However, this task was meant for A2/B1 level candidates, so the rubric is quite challenging and difficult to internalise (*preferably, very much against this idea, might not be the ideal living place for an animal, convince, has its advantages, sense of responsibility*). It would need considerable simplification. It would be better if the interlocutor read the rubric, so that the candidate would not have to read so much.

The candidate's role-card has all the necessary information and guidance for carrying out a successful interaction in the form of bulleted short text prompts. However, there seems to be no requirement for the candidates to produce their own ideas.

The interlocutor's card has six bullet points but with very detailed and specified ideas in full sentences. There is no space for flexibility if the candidate offers something else. The interlocutor's hands are tied, which can easily lead to a breakdown in communication. There should be no need to script the Interlocutor's contribution word by word.

ROLE-PLAY TASK 14: WALKING WORLD-WIDE

(For a sample candidate performance see DVD Sample 6.7)

Interlocutor's card

WALKING WORLDWIDE

Small group 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.

Role-card 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.

The advertisement from The Walking Magazine, the USA:

WALKING WORLDWIDE

Small group 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.

Role-play Task 15 suffers from several serious task design problems. The situation of the role-play is quite unnatural, the candidate has to take a totally unfamiliar role. Playing the parent of the examiner-interlocutor can be strange and, for some candidates, may be shocking. The situation and the roles are not sufficiently well defined and the instructions are not sufficient, especially the interlocutor's, who has no guidelines at all. There is no Interlocutor Frame, and the candidate has to do a lot of reading. Both the candidate and the interlocutor find themselves in a very difficult situation (see the DVD sample), and the interaction cannot develop into any meaningful conversation. The task ends quite quickly, simply because neither the candidate nor the interlocutor can add any new elements to develop it further. More guidelines and different roles would be needed to turn it into a suitable task for speaking examinations.

ROLE-PLAY TASK 15: MICHAEL'S RESTAURANT

(To view a candidate performing this task, see DVD Sample 6.8)

Role-card 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.

MICHAEL'S RESTAURANT

An excellent varied menu.

A totally unique dining experience.

3 Crown Street, Bolton,

Tel: (01204) 373325

Interlocutor's card

Inquire about: – number of people
– time
– what table (by window?)

Give info: – no window tables available between 6 p.m. and 8 p.m.
– no chicken dishes available tonight

In *Role-play Task 15* the basic task design problem lies in an insufficient prompting of the candidate. The situation is life-like and could work apart from the fact that this is a telephone conversation 'in face-to-face mode' again. It would be advisable to avoid these telephone situations and create a face-to-face encounter instead. Apart from the advertisement, the candidate's role-card has no prompts for ideas at all, the candidate is left to her own devices completely. There is no way for the candidate to dominate the interaction with so little information and guidance available. Strangely enough, the interlocutor has much more to rely on, he talks too much, and his contributions (see the DVD sample) totally dominate the scene. The lesson that can be learnt from this task is that without sufficient prompting the candidate fails to fully participate in the interaction, and the shift of balance towards the interlocutor kills the desired interaction.

Recommended role-play tasks in the individual mode

The Hungarian Examinations Reform Teacher Support Project has tried out a number of role-play tasks in the individual mode with real students in mock exam situations. We recommend the following tasks as appropriately designed speaking activities to assess candidates performing individually with an interlocutor. You can also judge how these tasks work as we provide a sample performance recorded on the DVD for each of them.

ROLE-PLAY TASK 16: A WEEKEND IN HUNGARY

(For a sample candidate performance see DVD Sample 6.9)

Interlocutor's Card

To be read out for the candidate:

I am a foreign visitor in Hungary. Give me advice on where to spend a weekend in your country. On your sheet there are some ideas but you can suggest other ones too. You must start the conversation when you're ready.

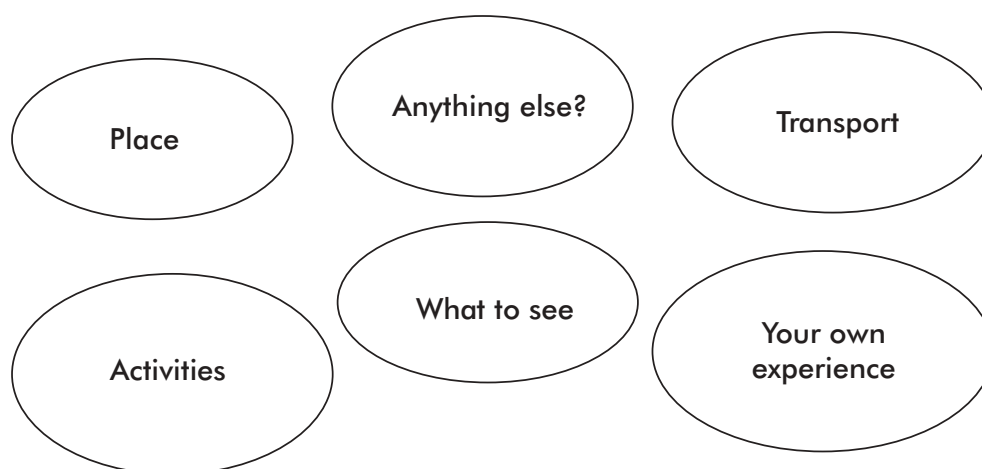
Prompts for the candidate:

- Place
- What to see
- Activities
- Transport
- Your own experience
- Anything else?

For the examiner only:

- Ask for some clarification you consider relevant.
- Challenge the candidate in a polite manner so that s/he should be forced to defend
- his/her arguments.
- Say you can't drive.
- You don't like walking for hours.
- You must finish the conversation.

Role-card for Candidate



ROLE-PLAY TASK 17: CONCERT**(For a sample candidate performance see DVD Sample 6.10)****Interlocutor's Card**

To be read out to the candidate:

I am a foreign tourist. You and I are in a ticket office in Hungary. Both of us want to buy a ticket for a concert. I don't know which concert to go to. Give me advice. On your sheet there are some ideas but you can suggest other ones, too. Say when you're ready.

Prompts for the candidate:

- Type of music: rock / pop / classical, etc.
- Time of the concert
- Where to sit
- Musicians
- Anything else?

For the examiner only:

- Ask for some clarification you consider relevant (e.g. ticket price, place of the concert).
- Don't accept the first two suggestions the candidate makes. Ask for further possibilities.
- You must start and finish the conversation.

Role-card for Candidate

Type of concert:
pop/rock/ classical,
etc.

Anything else?

Musicians

Time of the
concert

Where to sit

ROLE-PLAY TASK 18: INTERNATIONAL QUIZ

(For a sample candidate performance see DVD Sample 6.11)

Interlocutor's Card

To be read out to the candidate:

You will participate in an international quiz on Europe. I am the team leader and I have to organise how the team prepares for the quiz. Suggest at least three topics that you would like to prepare for. Try to convince me why you think the topics would suit you. On your sheet there are some ideas but you can suggest other ones too. You must start the conversation when you're ready.

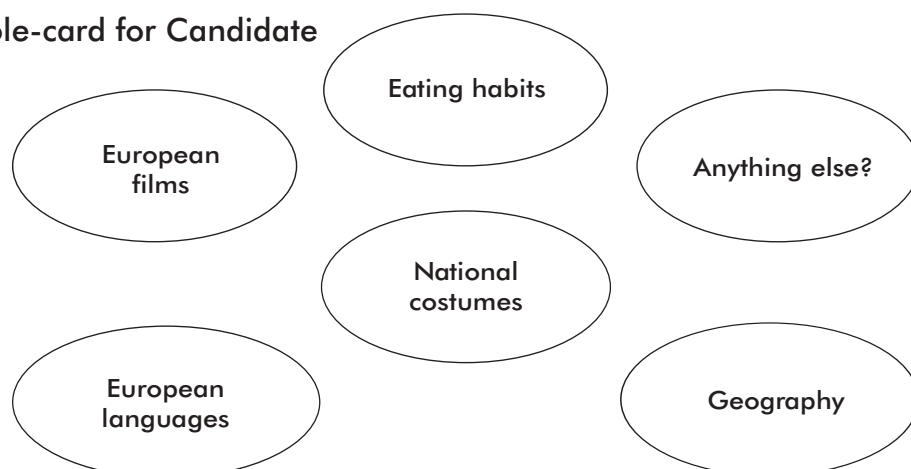
Prompts for the candidate:

- Geography
- Eating habits
- National costumes
- European languages
- European films
- Anything else?

For the examiner only:

- Ask the candidate about his/her personal experience in connection with the topics.
- Inquire how s/he is planning to prepare for them.
- Ask the candidate to select three topics that would suit him/her best.
- Accept two of them and say that you are not sure about the third one.
- Ask the candidate to convince you that s/he can be well prepared for that topic.
- You must finish the conversation.

Role-card for Candidate



ROLE-PLAY TASK 19: SCHOOL ANNIVERSARY

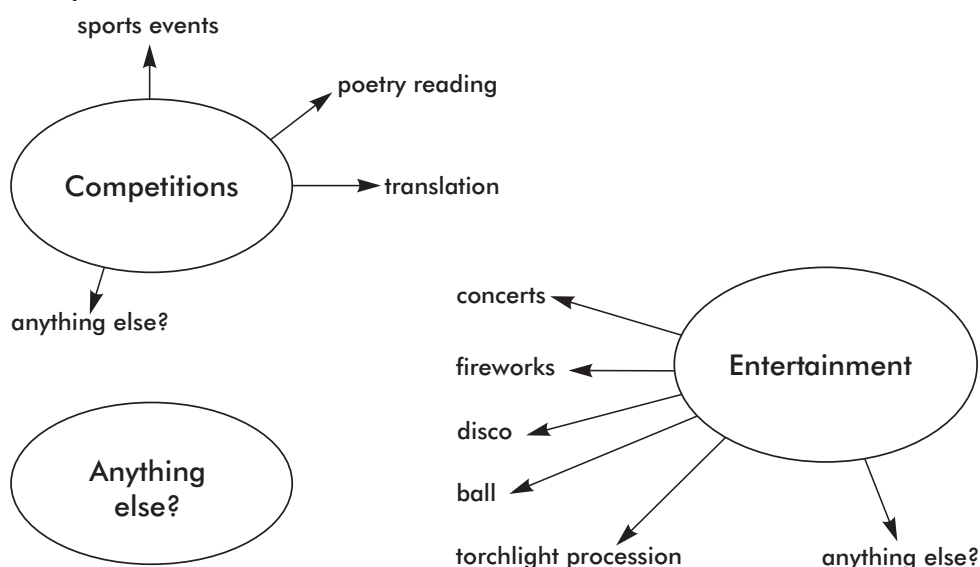
(For a sample candidate performance see DVD Sample 6.12)

Interlocutor's Card

To be read out to the candidate:

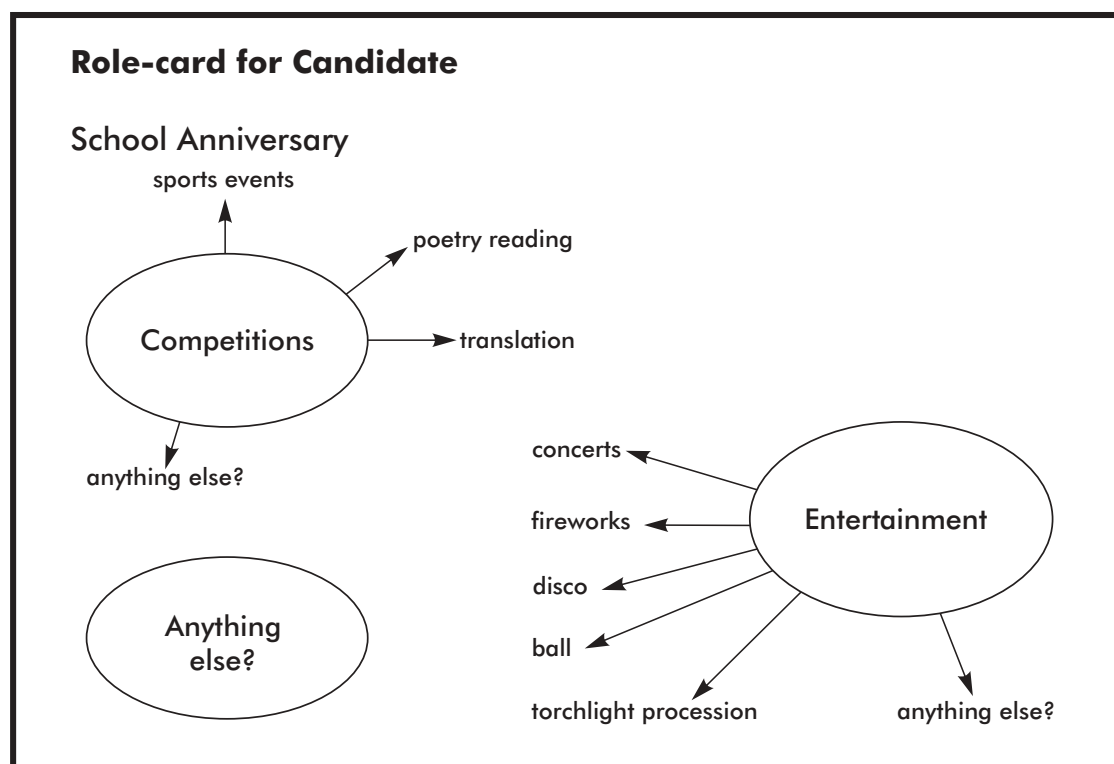
Your school is going to celebrate its anniversary and would like to find a sponsor. I am a teacher in the organizing committee and I will help you to get some money. First, you have to tell me about your plans. Explain why your ideas are good and what you would like to spend the money on. On your sheet there are some ideas but you can suggest other ones too. Your plan should have at least three ideas. You must start the conversation when you're ready.

Prompts for the candidate:



For the examiner only:

- It is very important for you that the events attract as many people as possible and they are appropriate and memorable ways of celebrating the school anniversary.
- For each idea ask about the following:
 - Who and how many people would be involved (students, teachers, parents, ex-students or other guests)?
 - Why would it be a memorable event?
 - Why would it be an appropriate way of celebrating the anniversary?
- Accept the ideas that satisfy your requirements.
- Reject the ideas that do not satisfy your requirements and either suggest changes or ask for other ideas.
- You must finish the conversation.



Based on the lessons learnt from piloting many role-play tasks, we recommend the following set of general guidelines for designing role-play activities. The guidelines were originally written to help secondary school and university teachers who had been trained in language testing, to write speaking test tasks for the Hungarian Examinations Reform Project. We believe that they are also useful for English teachers in general to help them write better speaking tasks for classroom assessment purposes.

GENERAL GUIDELINES FOR DESIGNING ROLE-PLAY TASKS

Prompts

- Select prompts/input text and design the task so that it generates the appropriate amount of varied language at the required level.
- Use prompts that are clear, black and white, photocopiable (in size max. A4) and appropriate for the target age.
- Limit input text so that candidates can process it in 30 seconds.
- The language level of verbal prompts should be below the tested level.
- Do not use distressing, offensive, violent or taboo topics. Treat sensitive topics with care.
- Do not use surreal, abstract, puzzling or symbolic pictures at lower levels.

Situations, roles, interaction

- Create situations that are life-like, suitable for the target age-group and in which it is realistic for them to speak in English.
- Candidates should not be required to take unfamiliar roles.
- Use different questions and tasks of different complexity if using the same prompt for more than one level.
- The task should not require students to use their imagination or creativity as these are not to be tested.
- Role-plays should involve real interaction and have an outcome.
- Create an information-gap or opinion-gap in role-play tasks.
- In the individual format, when the student has to speak to the interlocutor, make sure the interlocutor does not have to speak more than the student.
- In the paired format, when two candidates have to talk to each other, make sure they have equal roles.
- Restrict the use of yes/no and alternative questions.
- Always try out the task yourself to see whether you can speak for the required length of time.

Rubrics

- Use standardised instructions.
- Instructions for candidates should be given orally by the interlocutor, so they should appear only on the interlocutor's sheet. Short, simplified versions only may appear on the candidate's sheet.
- Instructions must be given in English.
- Instructions should be simple.

Chapter 7:

Training Interlocutors

The aims of the interlocutor training

The aim of the interlocutor training developed by the Hungarian Examination Reform Teacher Support Project is to provide participants with sufficient information about the model Speaking Examination developed for training purposes. Participants are provided detailed information about the outline, task types, and mode of the examination model. They also gain insight into different other possible modes of conducting speaking examinations. The training course familiarises participants with the main principles and procedures of administering the model Speaking Examination and enables them to develop the necessary interlocuting skills through simulation and practice.

The outline of the interlocutor training

The interlocutor training has three main stages. Stage 1 is a distance phase. This means participants carry out pre-course tasks in a self-study mode at home in preparation for Stage 2. The pre-course tasks include detailed studying of the *Introductory Training Pack* and accomplishing the pre-course tasks. The Introductory Training Pack is accompanied by a pre-course video, which contains videoed sample performances to observe.

Stage 2 is a face-to-face phase, where a series of workshop sessions are held to discuss the experiences of the first distance phase and to get more guided hands-on practice in conducting speaking examinations. During the live interlocutor training sessions participants familiarise themselves with the *Interlocutor Frame* and the administration procedure through analysing DVD samples of standard and non-standard interlocutor behaviour. The standardisation of the administration procedure takes place through simulated examination situations. Finally, alternative modes and techniques of conducting oral examinations are demonstrated.

Stage 3 is a distance phase again. It is done in participants' own environment (schools, examination centres) after the course. Participants' post-course tasks include practical application of the acquired skills through conducting mock speaking tests in a standardised manner, and writing a report about the results. Feedback on the usefulness of the training is also provided by them to the course trainers.

Stage 1: Pre-course distance learning

The basis for the pre-course stage is the *Introductory Training Pack* (henceforth referred to as *Intropack*). The main body of the *Intropack* contains all the necessary information about the speaking test with the pre-course tasks organised as follows:

1. General Introduction
2. Outline of the Interlocutor Training
3. Overview of the model Speaking Examination
4. Guidelines for Interlocutors with the *Interlocutor Frame*
5. Pre-Course Tasks

This is followed by a section containing *Self-Assessment Questions* (SAQs) with an answer key. The SAQs help participants internalise all the information they have studied in the *Intropack*.

The *Appendices* section contains sample speaking examination tasks and useful reference materials in connection with testing speaking, for example the Council of Europe Framework scales, a glossary of testing terminology and recommended literature on testing speaking.

The *Intropack* is accompanied by a *Pre-Course Video*. This contains sample performances recorded during pilot examinations for training purposes.

Sample materials from the Introductory Training Pack

In this section we provide an insight into the materials of the Introductory Training Pack. Sample 1 outlines the model Speaking Examination. The information provided by the overview below is necessary for course participants.

SAMPLE 1: AN OVERVIEW OF THE MODEL SPEAKING EXAMINATION

Part	Task	Timing	
		A2/B1 level (10-15 minutes)	B2 level (20 minutes)
Part 1	Interview	2-3 minutes	4-6 minutes
Part 2	Individual long turn	4-6 minutes	7-8 minutes
Part 3	Simulated discussion task (role-play)	4-6 minutes	7-8 minutes

Future interlocutors need general guidelines for behaving and acting in a standardised manner while conducting the speaking examinations. Sample 2A below demonstrates guidelines for examinations conducted in the individual mode.

SAMPLE 2A:
GUIDELINES FOR INTERLOCUTOR BEHAVIOUR FOR
EXAMINATIONS CONDUCTED IN THE INDIVIDUAL MODE

Throughout the whole exam	DO's	DON'Ts
The Environment	Well-lit, friendly, peaceful, undisturbed.	No other candidate, apart from the one being examined, is allowed to be present in the examination room.
	Arrange chairs for the exam so that the candidate and the interlocutor sit facing each other. The assessor should be sitting outside this arrangement but ideally s/he should be able to see the candidate's face.	Avoid positions suggesting confrontation or inequality (e. g. the interlocutor sitting higher than candidates). Don't forget that you should not dominate the conversation but facilitate it.
The Interlocutor	BE NICE, POLITE, PLEASANT, ENCOURAGING AND ATTENTIVE	DON'T BE UNPLEASANT, IRONIC, DISAPPROVING, DISAGREEABLE OR PATRONISING.
	Keep to the suggested time limits. Give the candidate enough time/opportunity to speak.	Don't start off 'prattling' to the candidate. Never ask more than one question at a time. Don't talk more than necessary: refrain from making unnecessary comments.
	Maintain eye contact with the candidate when talking to him/her. Remember and use the candidate's first name.	Don't look at the task sheet for a long time or shuffle your papers when the candidate is talking to you.
	Speak clearly.	Never raise your voice.
	Receive mistakes and weaknesses without comments.	Don't give any indication of assessment of candidate performance. Don't correct or give 'lectures'. Don't let your face show very strong emotions (either positive or negative), resulting from the candidate's performance.
	Give encouraging and positive backchannel signals (e. g. umm, aha, right, etc.).	Don't interrupt or finish what the candidate wants to say.
	Be discrete when making notes or checking the time.	Never remind the candidate that they are being examined.

In Part 3	DO's	DON'Ts
The Interlocutor	Use global questions for elicitation.	Do not ask questions that require background knowledge. Avoid ambiguous and embarrassing questions.
	Use referential/genuine questions.	Avoid asking display questions (i.e. questions to which the answer is already known).
	Ask <i>why</i> -questions whenever possible. For example, ask the candidate to justify his/her opinion (e.g.: Why do you think...?).	Avoid asking <i>yes/no</i> questions.
	Interrupt the candidate's talk as and when necessary to make the conversation suitably interactive.	Don't let candidates give a monologue.
	In case of incomprehension, increase the comprehensibility of input by using various meaning-negotiating strategies (e.g. use paraphrasing).	Do not just repeat what you said if/when you identify incomprehension. Do not over-accommodate, i.e. simplify your language too much to help the candidate.

Interlocutors conducting speaking examinations in the paired mode face special challenges. Sample 2B provides additional guidelines which cater for these situations.

**SAMPLE 2B:
ADDITIONAL GUIDELINES FOR INTERLOCUTOR BEHAVIOUR
FOR EXAMINATIONS CONDUCTED IN THE PAIRED MODE**

DO's
Arrange chairs for the exam so that the two candidates and the interlocutor could sit in a triangle. The assessor should be sitting outside this triangle but ideally s/he should be able to see the candidates' faces.
Make sure both candidates are given an equal opportunity to speak.
Maintain eye contact with the candidate when s/he has to address you (the interlocutor). Speak clearly.
DON'Ts
Avoid positions suggesting confrontation or inequality (e. g. the interlocutor sitting higher than candidates). Don't forget that you should not dominate the conversation but facilitate it.
Don't keep eye contact in the paired task to prevent the candidates addressing you rather than their peer partner. Don't look for eye contact with the assessor during the exam as it may embarrass the candidate.

Apart from the general guidelines for behaviour, future interlocutors have to use standardised wording for beginning the examination, giving instructions, providing transitions from one part of the examination to another, intervening and rounding off the examination. The *Interlocutor Frame* scripts these for the interlocutors. (Sample 3B below demonstrates the short version of the *Interlocutor Frame* for the individual mode B2 level exam.)

The Intropack is accompanied by a pre-course video which contains sample performances. Participants are asked to observe the interlocutor's behaviour and comment on it. Sample 3A shows a pre-course task participants have to accomplish. The reader can do this task using DVD Sample 7.1 (attached to this book). Use the short version of the *Interlocutor Frame* for the individual mode B2 level examination provided in Sample 3B.

SAMPLE 3A: PRE-COURSE TASK

PRE-COURSE TASK 1: OBSERVING INTERLOCUTOR BEHAVIOUR

- Watch DVD sample 7.1.
- While watching it, **observe the interlocutor's behaviour**. Report on how the interlocutors managed to apply the "*Guidelines for Interlocutor Behaviour*" and the *Interlocutor Frame*.
- Report on your findings **by completing the following Pre-Course Task Report Sheet**.

PRE-COURSE TASK REPORT SHEET

1. Try to remember and put down the things the interlocutor said that **were not included in the *Interlocutor Frame***.

2. What were your general impressions of the interlocutor's behaviour?

SAMPLE 3B:
SHORT VERSION OF THE INTERLOCUTOR FRAME FOR THE
INDIVIDUAL MODE B2 LEVEL EXAMINATION

PART 1 (5-6 min)

Hello. My name's ... *(the Interlocutor introduces him/herself)*

My colleague is just going to listen. *(the Interlocutor refers to the silent Assessor)*

What's your name? *(to the candidate)*

First, I'd like to ask you some questions. *(the interlocutor says the candidate's first name)*

What do you like about your school?

I see, thank you.

Let's talk about something else.

Nowadays many people learn foreign languages, even after school.

- Why do you think it is important to learn foreign languages?
- Which are the most popular foreign languages? Why?
- In what kinds of jobs is it essential to know foreign languages?
- What are the advantages of speaking more than one foreign language?
- What are the best ways of learning languages?

OK, that's enough, thank you.

Let me ask you now about advertisements.

It seems that people buy everything that is well advertised.

- Have you ever bought anything because of an advertisement you saw/read/heard? What was it?
- Which products are advertised too often? Why?
- What do you find the most effective way of advertising?

AFTER 5-6 MINUTES

That will do, thank you.

PART 2 (5-6 minutes)

Now I'm going to give you some pictures. I'd like you to talk about them.

Please draw a number, ...

It is number ...

Here are your pictures.

READ OUT THE TASK SPECIFIC INSTRUCTIONS, WAIT 5-10 SECONDS.

Shall I repeat what you have to do?

WAIT MAX. 30 SECONDS

Let's start now.

AFTER 5-6 MINUTES OR WHEN THE CANDIDATE HAS FINISHED
That's enough. Thanks. / Thank you.

PART 3 (5-6 minutes)

Now let's talk about something else.

I'd like you to draw a number again.

It is number ...

Here is your task sheet.

AFTER GIVING THE TASK SHEET TO THE CANDIDATE
Please listen carefully.

READ OUT THE TASK SPECIFIC INSTRUCTIONS, WAIT 5-10 SECONDS.
Shall I repeat what you have to do?

WAIT MAX. 30 SECONDS

Let's start now.

If there is a communication breakdown, the interlocutor should

- repeat all or part of the rubric
- invite the candidate to talk about one specific aspect of the task

AFTER 5-6 MINUTES OR WHEN THE CANDIDATE HAS FINISHED THE TASK

Thank you. / That will do, thank you.

ROUNDING OFF THE EXAMINATION

Thank you, ... (candidate's name) That's the end of the test for you.

Thank you very much. (Good luck.)

The Self-Assessment Questions help participants internalise the information provided by the Intropack. With the help of the attached Key they can evaluate how well they have internalised the content of the Intropack. Sample 4 provides some of these self-assessment questions. These questions relate to examinations conducted in the individual mode.

SAMPLE 4: SAMPLE SELF-ASSESSMENT QUESTIONS (SAQS) WITH KEY

SAQ 1

What is the general aim of the Speaking Examination?

To assess candidates' ability to interact in conversational English in a range of contexts & to use spoken language to perform different tasks related to everyday life.

SAQ 2

Give reasons why the proposed Speaking Examination is standardized.

Candidates should feel secure in the perception that their performances have been fairly judged in appropriate conditions.

SAQ 3

What kind of questions can the Interlocutor ask in Part 1 of the Speaking Examination?

S/he can select the questions from a list of options provided in the shortened version of the Interlocutor Frame.

SAQ 4

How can the Interlocutor prevent the candidate from giving a "pre-prepared speech" in Part 1 of the Speaking Examination?

By interrupting in good time and asking further questions that will change the focus of the conversation.

SAQ 5

Can the Interlocutor make up his/her own questions to help the candidate talk about the pictures in Part 2 of the Speaking Examination?

No. The Interlocutor is not allowed to divert from the question prompts given on the Interlocutor's sheet.

SAQ 6

How should the Interlocutor assist in Part 3 if the interaction fails to develop or breaks down completely?

If there is a communication breakdown, the interlocutor should

- repeat all or part of the rubric,
- invite the candidate to talk about one specific aspect of the task.

SAQ 7

Which of the following guidelines should the interlocutor observe? Please put a tick in the right boxes.

- ☐ Ask several questions to make sure the candidate will understand at least one of them.
- ☒ Give encouraging feedback by using back channel signals.
- ☒ Keep eye contact with the candidate all throughout the exam.
- ☒ Tolerate mistakes and weaknesses without commenting on them.
- ☐ Keep eye contact with the assessor during the exam.

SAQ 8

What is the purpose of using the Interlocutor Frame?

With the help of the Interlocutor Frame, the Interlocutor can

- adopt a standard manner during the examination,
- adopt standard procedures in conducting the examination.

SAQ 9

What should the interlocutor do if the candidate makes a mistake during the test?

Nothing at all. There should be no comments, let alone corrections on the part of the interlocutor.

SAQ 10

What question types should the Interlocutor *avoid* using in Part 3?

- questions that require background knowledge,
- questions that are ambiguous and embarrassing,
- display questions,
- yes/no questions.

SAQ 11

When do the assessor and the interlocutor discuss the candidates' marks?

They do not discuss them at all. It is the assessor's sole responsibility to assess the performances and to award the marks.

Stage 2: Live interlocutor training course

The one-day live interlocutor training course consists of four 90-minute workshop sessions:

- Session 1: Introduction, Pre-course Tasks
- Session 2: Familiarising participants with the methods and procedure of the model Speaking Examination.
Demonstrating standard and non-standard interlocutor behaviour.
Practising conducting Part 1 tasks and applying the *Interlocutor Frame*.
- Session 3: Practising conducting Part 2 and Part 3 of the Speaking Examination.
Internalising the relevant sections of the Interlocutor Frame.
Preparing future interlocutors for handling difficult exam situations.
- Session 4: Comparing and contrasting alternative approaches to oral examinations.

The basic material for these sessions is provided by recorded standard and non-standard sample interlocutor performances.

Session 1 deals with the *Introductory Training Pack* and the *Pre-course Task*. Participants compare and discuss their experience with the *Pre-course Task* and ask any questions they may have concerning the material of the *Introductory Training Pack*.

In the first part of the session participants go into pairs and discuss their impressions about the interlocutors' behaviour on the pre-course DVD. They can consult and compare their Task Report Sheets. They are asked to list positive features of the standardized interlocutor behaviour as well as to raise issues and share their worries in connection with the Interlocutor's role.

Trainers are advised that the following issues may come up during the discussion:

- It's difficult to challenge "passive" students.
- Students may be embarrassed and anxious.
- We would like to ask further questions.
- It's difficult to use global questions only.
- It seems unnatural to stick to the questions in the Frame.
- You have to adjust to the role of interlocutor.
- It is difficult not to correct.
- It may be difficult to keep the time limit.
- Some students are not at the level.
- Students don't know about the topic.

Participants and trainer reflect and react. Opening up the floor, the trainer answers any questions that have come up in the group discussions concerning the *Introductory Training Pack*, or the speaking exam, and draws the discussion to a close.

Background information for the trainer:

- Interlocutor should challenge passive and anxious students by positive backchannelling.
- Candidate training is very important, practice can help passive and anxious students.
- No further questions are allowed, only Part 3 provides some freedom.
- A global question is likely to provide a wider scope for candidates to respond than a question focusing on specific details or lexical elements.
- The following verbal backchannelling signals are possible: Yes, Right, Aha, All right, OK, etc. Anything that is neutral and brief.
- An exam situation is not necessarily natural, a test is always artificial.
- The role of an interlocutor and a teacher is different: the former should refrain from any verbal or non-verbal evaluation and error correction; the aim is to elicit as much language as possible.
- Interlocutors should practise their roles, as it is not an easy job to conduct an exam; in order to minimize subjectivity, a standardized examination procedure should be followed.
- Students should be prepared for the exam properly: the topics are listed in the specifications and they should also practice exam situations as well.

Session 2 familiarises participants with the methods and procedure of the model Speaking Examination, demonstrates standard and non-standard interlocutor behaviour, and provides opportunities for practising conducting Part 1 tasks and applying the *Interlocutor Frame*.

First, participants observe standard interlocutor behaviour (see DVD Sample 7.2). They are asked to use their *Introductory Training Pack* to monitor the Interlocutor's contributions. Participants watch the DVD and follow the interlocutor's mode of applying the *Guidelines for Interlocutor Behaviour* and the *Interlocutor Frame*. (The reader can use the short version of the *Interlocutor Frame* for the paired mode A2/B1 level examination provided in Sample 5). Participants are asked to focus on this question: '*How well did the interlocutor apply the Interlocutor Frame?*' The question is also put on the flipchart/OHT. Participants are requested to list three things that they like the most about the interlocutor's behaviour.

Participants are provided the following observation points to guide the DVD viewing:

- manner of conducting the exam
- asking questions in Part 1 and 2
- elicitation techniques in Part 3
- setting the tasks
- intervening
- eye contact

**SAMPLE 5: SHORT VERSION OF THE INTERLOCUTOR FRAME
FOR THE PAIRED MODE A2/B1 LEVEL EXAMINATION****PART 1**

Hello. My name's ... My colleague is just going to listen.

What's your name? *(Candidate A)*

And what's your name? *(Candidate B)*

CANDIDATE A

Let me first ask you some questions.

1. *Do you have a large family or a small family?*
2. *What kinds of shops are there near the place where you live?*
3. *Do you like shopping?*
4. *Is learning languages important for you? Why/Why not?*
5. *How do you usually spend your holidays?*
6. *Do you go out in your free time? (If yes) Where do you go?*

Thank you.

CANDIDATE B

Now ... let me ask you some questions.

1. *Can you tell me about the job you would like to have in the future?*
2. *Where are you from? Do you like living in this town?*
3. *Have you got any hobbies? (If yes) How did you become interested in it / them?*
4. *Where would you like to go on holiday and why?*
5. *What do you and your friends like wearing at school?*
6. *What do you enjoy most about learning English?*

Thank you.

PART 2

Now I'm going to give each of you some pictures. I'd like you to talk about them.

Here are your pictures.

I'll give you your pictures, Candidate B, when s/he has finished but I will ask you a question about these pictures in a few minutes.

Candidate A, can you show the pictures to Candidate B?

READ OUT THE TASK SPECIFIC INSTRUCTIONS, WAIT 5-10 SECONDS.

Shall I repeat what you have to do?

WAIT MAX. 30 SECONDS.

I'd like you to start now.

AFTER CANDIDATE A HAS FINISHED

Thank you.

REPEAT LAST QUESTION OR ASK CANDIDATE B:

And what is your opinion about this? OR And what about you?

Thank you.

Here are your pictures.

Can you show the pictures to Candidate A? I will ask you, too, a question about these pictures in a few minutes

READ OUT THE TASK SPECIFIC INSTRUCTIONS, WAIT 5-10 SECONDS.

Shall I repeat what you have to do?

WAIT MAX. 30 SECONDS

I'd like you to start now.

AFTER CANDIDATE B HAS FINISHED

Thank you.

REPEAT LAST QUESTION OR ASK CANDIDATE A:

And what is your opinion about this? OR And what about you?

Thank you.

PART 3

Now I'd like you to talk about something together. I'm just going to listen.
Here are your task sheets.

AFTER GIVING TASK SHEETS TO CANDIDATES
Please listen carefully.

READ OUT THE TASK SPECIFIC INSTRUCTIONS, WAIT 5-10 SECONDS.

Shall I repeat what you have to do?
WAIT MAX. 30 SECONDS
I'd like you to start now.

If there is a communication breakdown, the interlocutor should

- *repeat all or part of the rubric;*
- *invite candidates to talk about one specific aspect of the task;*
- *invite the candidate whose contributions seem to be unsatisfactory (i.e. too short) to talk about one specific aspect of the task.*

AFTER 5-6 MINUTES OR WHEN CANDIDATES HAVE FINISHED THE TASK

Thank you / That will do, thank you.

ROUNDING OFF THE EXAMINATION

Thank you Candidate A and B. That's the end of the test for you.
Thank you very much. (Good luck.)

After watching the DVD, the trainer initiates a short plenary discussion, in which s/he tries to elicit comments in relation to the points listed earlier.

Participants' feedback on standard interlocutor behaviour will probably include the following features:

- The Interlocutor seems organized, calm, encouraging.
- The Interlocutor sticks to the questions in the Frame.
- The Interlocutor uses both verbal and non-verbal backchannelling signals, there is sufficient waiting time, elicits not only with questions, but makes comments as well.
- In the individual mode the Interlocutor attempts to maintain a natural flow of discussion in Part 3, though doesn't take the floor.
- The instructions are given at a slower than natural speed and repetition is offered.
- The Interlocutor keeps constant eye contact with the candidate, which shows encouragement and interest in what he says.

After having seen and discussed standard interlocutor behaviour, participants have the opportunity to compare it with non-standard performances. Participants watch DVD samples that show non-standard interlocutor behaviour. In both samples the interlocutor exaggerated the features of non-standard interlocutor behaviour. (For sample interlocutor performances see DVD Sample 7.3 for the individual mode and DVD Sample 7.4 for the paired mode.)

Participants write down examples of non-standard interlocutor behaviour with special regard to the instances where the interlocutor's behaviour deviates from the *Interlocutor Frame*. They are requested to note down the negative features of the interlocutor's behaviour. After watching the DVD, participants work in groups and draw up a list of the negative examples on OHT.

Participants' feedback on non-standard interlocutor behaviour will probably include the following features:

- The interlocutor is disorganized, unprepared and impatient.
- At times she shows unnatural interest, at other times she shows boredom.
- In the paired mode she keeps mixing up the two candidates.
- The Interlocutor talks at length and expresses her own ideas without listening to the candidate.
- She talks about personal things and opinions.
- She corrects the candidates.
- The questions in Part 1 and 2 are asked quickly, there is not enough waiting time and the order is not followed either.
- During the Part 3 task, the interlocutor goes back to the previous task, which confuses the candidate(s) entirely.
- In some cases the task is not properly set as the candidate doesn't know what to do.
- In the individual mode the Interlocutor fails to provide eye contact, which is highly disturbing as this expresses indifference and boredom.
- In both examples the test is ended in an unnatural way.

In the next phase of the live training participants simulate real examination situations.

First, they are asked to role-play Part 1 tasks in groups of 3. The materials for these role-plays are role-play cards (for the Interlocutor and the Candidate) and observer's task sheets. Sample 6 demonstrates the materials for the Part 1 role-plays.

SAMPLE 6: ROLE-PLAY CARDS AND OBSERVER'S TASK SHEET FOR SIMULATING PART 1 IN THE INDIVIDUAL MODE

Version A

Candidate

You are a shy, not very talkative candidate who tends to wait for guiding questions. You often reply with one or two short sentences only.

Version B

Candidate

You are a candidate who tends to answer the interlocutor's questions by presenting a rehearsed text at excessively unnatural speed. You avoid eye contact with the interlocutor.

Interlocutor (for A & B versions)

(Please remember to ask the Qs listed in the Interlocutor Frame)

You are the interlocutor who asks the questions of the first part of the speaking test. You have to elicit as much speech from the candidates as possible.

Observer's task sheet for Part 1 tasks

While observing the interlocutor's way of conducting the Speaking Examination, please refer to the following questions. When the mock exam is over, give feedback in your group.

1. Does the interlocutor only ask questions listed on the task sheet?		
Yes / No	What else?	
2. How does the interlocutor react to the candidate's answers?		
With short but encouraging responses.	With unscripted comments.	
3. Does the interlocutor ever interrupt?		
Yes / No	If yes, why?	How?
4. Does the interlocutor keep to the allotted time of 2-3 minutes?		
Yes / No	Finish earlier / Finish later	

Trainers are advised that the following issues may come up in the plenary discussion about conducting Part 1 tasks:

- It is difficult to stop a talkative student – interrupt with a non-evaluative comment.
- Sometimes the candidate gives a short answer – wait patiently, indicate that you are waiting.
- Some candidates may come up with rote-learned material – stop them immediately and ask them about something else.
- The interlocutor should maintain eye contact all the time.

Session 3 aims to practise conducting Part 2 and Part 3 of the speaking test, encourage participants to internalise the relevant sections of the Interlocutor Frame, and prepare future interlocutors for handling difficult exam situations. Throughout this session, participants role-play in groups of three, practising Part 2 and Part 3 tasks. Below see sample role-play cards and observer's task sheets for this session.

SAMPLE 7: SAMPLE ROLE-PLAY CARDS AND OBSERVER'S TASK SHEET FOR PART 2

Version A Candidate

You can describe what can be seen in the pictures but you cannot compare and contrast them. You are hesitant, looking for the right words. You make rather long pauses.

Version B Candidate

You try to compare and contrast the two pictures but you're constantly searching for language and expect the examiner to help you. Sometimes you ask for words in English.

Interlocutor (for A & B versions)

You are the interlocutor conducting the second task of the speaking test. You have to elicit as much speech from the candidate as possible. Please remember to ask the questions listed on the Interlocutor's Sheet.

Observer's task sheet for Part 2 tasks

While observing the interlocutor's way of conducting the Speaking Examination, please refer to the following questions. When the mock exam is over, give feedback in your group.

1. Does the interlocutor only ask questions listed on the task sheet?	
Yes / No	Give examples:
2. Does the interlocutor elicit enough speech?	
Yes / No	Why not?
3. Does the interlocutor keep to the suggested time (4-6 minutes)?	
Yes / No	Why not?
4. Does the interlocutor interrupt the candidate?	
Yes / No	If yes, why and how?

Trainers are advised that the following issues may come up during the plenary discussion about conducting Part 2 tasks:

- In case the candidate is hesitant or asking for language, the interlocutor should say, "I am sorry, I can't help".
- If the candidate doesn't seem to understand, the Interlocutor can repeat the question.
- The Interlocutor should try to elicit language from the candidate.
- The Interlocutor should intervene in case of misunderstanding, and should try to put the candidate back on track.

SAMPLE 8: SAMPLE ROLE-PLAY CARDS AND OBSERVER'S TASK SHEET FOR SIMULATING PART 3 IN THE INDIVIDUAL MODE

Version A**Candidate**

You are generally good at English but this time you don't have the language to talk about the topic on the Task Sheet, so you start talking about other things.

Interlocutor

You must try to get the candidate to focus on the task instead of talking about other things. You may need to interrupt to make sure that the candidate responds to the task.

Version B**Candidate**

You are a candidate whose English is not at the required level. You keep repeating yourself or the examiner's words. You cannot contribute to the discussion effectively.

Interlocutor

You must make sure that the candidate contributes sufficiently to the conversation and therefore you try to use different question types and paraphrasing.

Observer's task sheet for Part 3 tasks

While observing the interlocutor's way of conducting the Speaking Examination, please refer to the following questions. When the mock exam is over, give feedback in your group.

1. Does the interlocutor provide adequate opportunities for the candidate to perform?		
Yes	How?	
No	Why not?	
2. Does the interlocutor interrupt the candidate in an appropriate manner?		
Yes / No	If yes, why?	How?
3. Does the interlocutor use appropriate forms of elicitation (e.g. question types)?		
Yes	How?	
No	Why not?	
4. Does the interlocutor keep to the suggested time limit (4-6minutes)?		
Yes / No	Why not?	

The following issues may come up during the plenary discussion about conducting Part 3 tasks:

- Talkative candidates are definitely better, but sometimes it is difficult to stop them.
- It is not an interview, the Interlocutor should interrupt, reflect on what the candidate says and comment on it – Part 3 is meant to be an interactive task.

- The interlocutor should avoid limiting the scope of the candidate's answer.
- Some questions that the interlocutor can ask to elicit an extended response from the candidate:
 - What do you think about ...?
 - What type of film do you like?
 - What's your opinion about ...?
- If you have already asked all the suggested question prompts and there is still no response, the candidate is probably not good enough so that is the end of the test for him/her.

SAMPLE 9: SAMPLE ROLE-PLAY CARDS AND OBSERVER'S TASK SHEET FOR SIMULATING PART 3 IN THE PAIRED MODE

Candidate 1

(pretends to be very poor at English)

You are a weak candidate, you hardly put any questions to your partner because you are worried about making mistakes. You hesitate a lot and expect your partner to help you.

Candidate 2

(pretends to be over-talkative)

You are very talkative and you decide to use your own ideas instead of the ones on the card. You come up with new suggestions, not listed on your card.

Interlocutor

Give the task instructions and withdraw from the interaction. Intervene as and when necessary (check the guidelines for intervening).

Observer's task sheet

While observing the interlocutor's way of conducting the Speaking Examination, please refer to the following questions. When the mock exam is over, give feedback in your group.

1. Does the interlocutor clearly set the task?		
Yes / No	Why not?	
2. Does the interlocutor give equal opportunity to both candidates to perform?		
Yes	How?	
No	Why not?	
3. Does the interlocutor interrupt or prompt either of the candidates?		
Yes	Why?	How?
No		
4. Does the interlocutor keep to the suggested time limit (5 – 6 minutes)?		
Yes / No	Why not?	

After having simulated all parts of the speaking examination, participants discuss their comments, questions and concerns in small groups. A plenary discussion follows and the trainer answers the questions that have been raised. Below, we present some frequently asked questions (FAQs) with possible answers.

How can the Interlocutor balance the contributions of candidates of different proficiency levels in Part 3 in the paired mode?

The Interlocutor should intervene by all means when there is a clear imbalance in the two candidates' contributions to the interaction. The actual wording of the intervention cannot be pre-specified but the Interlocutor is recommended to clearly indicate to the candidate who is contributing inadequately to the task that s/he should try harder. An easy way of trying to involve the candidate is to ask "And what do you think?". Alternatively, the Interlocutor may ask the candidate to talk about a specific point that has already been mentioned.

How many times should the Interlocutor repeat questions in Part 1?

The Interlocutor can repeat the interview questions in Part 1 and the guiding questions in Part 2 once or maximum twice.

What should the Interlocutor do when candidates misunderstand the task?

The Interlocutor is expected to intervene if candidates have seriously misunderstood the task. If this happens, the Interlocutor is advised to repeat the whole or just specific parts of the instructions, depending on what candidates misunderstood about the task.

How should the Interlocutor intervene in the case of very talkative candidates in the paired mode?

The Interlocutor should interrupt the talkative candidate and specifically ask the other candidate to join in. The actual wording of this intervention cannot be pre-specified but the question "And what do you think?" is an easy way to involve the candidate as well as to indicate to the talkative partner that it is time to let the other candidate contribute to the interaction, too. When the Interlocutor has to intervene, s/he is advised to re-establish eye contact with the candidates, but when the intervention is over, the Interlocutor has to withdraw eye contact once again.

What should the Interlocutor do when one of the candidates is continuously trying to keep eye contact with him/her in the paired mode?

The Interlocutor should ignore it and should look down (e.g. at the papers in front of him/her) to indicate that s/he is not the target audience although s/he is also listening.

What if the candidate doesn't respond at all?

In a real examination situation this is highly unlikely to happen, candidate training is essential.

After these sessions participants will probably feel more confident about using the questions in the *Interlocutor Frame*; DO's and DON'Ts for Interlocutor behaviour; being patient while examining; how and when to interrupt/intervene; how and when to finish the conversation; using non-verbal signals; the different ways of

eliciting language; handling different types of students and receiving mistakes without comments.

Some possible doubts from participants and possible reactions to them from the trainer:

- Timing – the interlocutor should practise keeping to the time-limits.
- Topics – all topics used in the Speaking Examination are described in the Examination Specifications in detail: candidate training is essential.
- Candidate anxiety: the interlocutor should create and maintain a stress-free atmosphere during the exam.

Session 4 is devoted to comparing and contrasting **alternative approaches** to the oral examination. The aims of the session are to raise participants' awareness of how scripted vs. unscripted contributions by the interlocutor may influence candidates' performance and to familiarize participants with the different possible modes of conducting speaking examinations (paired vs. individual formats).

The following steps and activities take place in this session:

1. Discussing the advantages and disadvantages of using an Interlocutor Frame.

Plenary discussion.

2. Demonstrating speaking tests conducted without an Interlocutor Frame

Participants watch 3 DVD samples and fill in an observation sheet (see Chapter 3, Table 3). These samples demonstrate how the lack of an Interlocutor Frame may influence examiner behaviour and in turn candidate performance. Two of the DVD samples show the same interlocutor, talking to two different candidates but using the same tasks (see DVD Samples 7.5 & 7.6). The third DVD sample shows another interlocutor, who uses the same tasks as the previous interlocutor (see DVD Sample 7.7). The DVD samples demonstrate how different interlocutor styles may influence the elicitation and eventually the assessments if the administration procedure is not standardised.

3. Observing the paired format

Participants watch 2 DVD samples of speaking tests, conducted in the paired mode showing good examples of interlocutor behaviour (see DVD Samples 7.8 & 7.9). Participants fill in an observation sheet. First, they discuss their observations in small groups, which is followed by a plenary discussion.

The trainer should

- ✓ dispel misconceptions and unfounded worries in connection with the paired format;
- ✓ emphasize the importance of appropriate task design;
- ✓ point out the interlocutor's key role in intervening whenever candidates seem to get stuck or need redirection;
- ✓ provide detailed guidelines for the above.

Stage 3: Post-course task

In the distance post-course phase future interlocutors are asked to try out the acquired interlocuting skills in mock speaking examinations in their own context. The practical application of the principles of good practice they learnt during the training in a familiar environment helps them become confident interlocutors and get used to conducting speaking examinations in a standardised manner. Participants usually carry out the post-course task in co-operation with another trainee interlocutor colleague, thus they can observe each other and share their experiences. They have to report on their experience in detail. They are also asked to give feedback on the usefulness of the training in the light of their practical experience. These post-course mock examinations are monitored by the members of the training team. Post-course feedback shows that this final phase is the real trial: after this, future interlocutors are ready to conduct live examinations.

The interlocutor training outlined above equips participants with interlocuting skills which will be useful for conducting many different kinds of modern European language examinations. The training course demonstrated in this chapter is currently being offered to teachers in Hungary by the British Council (see www.britishcouncil.hu).

Chapter 8: Training Assessors

The aims of the assessor training

The main aims of the assessor training developed by the Hungarian Examination Reform Teacher Support project are to provide participants with sufficient information about the Speaking Examination they are going to be trained for (outline, task types, mode), and to familiarise participants with the main principles and procedures of assessing speaking performances. The training also aims at enabling participants to develop the necessary assessing skills and introduces the idea and practice of using analytic rating scales for assessing oral performances. Valid and reliable assessment of live performances is ensured through standardisation.

The outline of the assessor training

Similarly to the interlocutor training, the assessor training has three main stages. Stage 1 is a distance phase. This means participants carry out pre-course tasks in a self-study mode at home in preparation for Stage 2. The pre-course tasks include detailed studying of the *Introductory Training Pack* and accomplishing the pre-course tasks. The Introductory Training Pack is accompanied by a pre-course video, which contains videoed sample performances to assess.

Stage 2 is a face-to-face phase, where a series of live workshop sessions are held to discuss the experiences of the first distance phase and to get more guided hands-on practice in different modes and techniques of assessing oral performances and using analytic rating scales for assessment. Standardisation of the assessment procedure and comparison of performances at different levels are also important elements of this stage.

Stage 3 is a distance phase again. It is done in participants' own environment (schools, examination centres) after the course. Participants' post-course tasks include practical application of the acquired skills through assessing mock speaking tests in a standardised manner, and writing a report about the results. Feedback on the usefulness of the training is also provided by them to the course trainers.

Stage 1: Pre-course distance learning

The *Introductory Training Pack* (henceforth referred to as *Intropack*) is the basis of the first, pre-course, stage. The main body of the *Intropack* contains all the necessary information about the speaking test, together with the pre-course tasks in the following sequence:

1. General Introduction
2. Outline of the Assessor Training

3. Overview of the model Speaking Examination
4. Guidelines for Assessors
5. Pre-Course Tasks

The main part is followed by a section containing *Self-Assessment Questions* (SAQs) with an answer key. The SAQs help participants internalise all the information they have studied in the *Intropack*.

The first part of the *Appendices* contains all the documents assessors need for carrying out the pre-course tasks:

- I. Sample Speaking Examination Tasks
- II. Assessment Scale
- III. Mark Sheet
- IV. Marks & Justifications for the Sample Speaking Tests
- V. Examples of Candidate Language for Interpreting the Assessment Scale

The second part of the *Appendices* contains useful reference materials in connection with evaluation and assessment:

- VI. Council of Europe Framework Scales
- VII. Glossary of Testing Terminology
- VIII. Literature on Testing Speaking

Excerpts from these will be provided later on in this chapter.

The *Intropack* is accompanied by a *Pre-Course Video*. This contains sample performances recorded in pilot examinations for training purposes. These performances have been benchmarked (see *Chapter 1*). The benchmarks and justifications produced by the judges in the benchmarking sessions are used for supporting the pre-course tasks. Each participant tries out the assessor's role in this first stage of the training by carrying out marking at home.

Sample materials from the Introductory Training Pack

This section provides sample materials from the *Intropack*. The first sample (Sample 1) is a set of guidelines which contains all the necessary information about the model speaking assessment scale, the usage of this scale and the assessor's behaviour. This document provides the basic knowledge for further live practice.

SAMPLE 1: GUIDELINES FOR ASSESSORS

THE FRAMEWORK OF THE SPEAKING ASSESSMENT SCALE

The Assessor uses an analytic rating scale for assessing the candidates' performance. There are four criteria, each consisting of 8 bands. Five of these bands (0, 1, 3, 5, 7) are defined by band descriptors, 3 of them (2, 4, 6) are provided for evaluating performances which are better than the level below, but worse than the level above, in other words performances which are in between two defined levels.

The criteria for assessment are the following:

Communicative impact

This criterion refers to

- the candidate's ability to take an active part in the interaction and express themselves effectively in fulfilling the task;
- the candidate's ability and willingness to contribute actively and positively to the development of the task and move it towards a conclusion (rather than supplying only minimal responses), to initiate and respond adequately, at a natural speed;
- the candidate's ability to use interactive strategies to maintain and / or repair communication (asking for clarification, paraphrasing, etc.);
- the amount of assistance (additional prompting) required from the interlocutor for the candidate in the course of fulfilling the task;
- hesitations and pauses that appear in the candidate's speech.

Grammar and coherence

This criterion refers to

- the range of grammatical structures the candidate uses during the performance;
- the accurate and appropriate use of these structures;
- the frequency and gravity of errors;
- the candidate's ability to express themselves coherently by using appropriate linking devices;
- the candidate's ability to maintain a coherent flow of language over several utterances (these utterances should range from those consisting of only one word to longer ones consisting of several words or even sentences).

Vocabulary

This criterion refers to

- the range of vocabulary the candidate uses;
- the appropriacy of the vocabulary the candidate uses;
- the candidate's ability to convey the intended meaning by using alternative words and / or phrases without excessive repetitions;
- the candidate's ability to use appropriate style and register.

Sound, stress, intonation

This criterion refers to the quality of the candidate's speech, i.e. his/her ability to produce comprehensible utterances. This criterion focuses on

- the candidate's production of individual sounds;
- the candidate's use of word and sentence stress;
- the candidate's ability to use intonation in order to convey the intended message effectively.

HOW TO USE THE RATING SCALE

Familiarising with the scale and the mark sheet

- Assessors must be thoroughly familiar with all the assessment criteria and the assessment procedures before they start marking.
- Assessors must study the mark sheet carefully before using it.

Arriving at a mark

- Assessors use a mark sheet for recording the candidates' marks. Each mark sheet consists of four boxes for each candidate. The boxes refer to the four criteria in the rating scale, and have the same headings.
- All three parts of the examination (Part 1: Interview, Part 2: Individual long turn and Part 3: Simulated discussion task) are assessed together by giving a score for each criterion only **ONCE**. By the end of each speaking examination session, the Assessor should decide and record the four separate marks for the candidate in the appropriate boxes of the mark sheet: one mark for Communicative Impact, one mark for Grammar and Coherence, one mark for Vocabulary and one mark for Sound, Stress and Intonation. A total score should also be calculated for each candidate after the individual scores for all the four criteria have been decided on.

- Initially, assessors must refer to the descriptors of a satisfactory performance (Band 3 for each criterion), and decide whether the candidate's performance is below, above or exactly at this level. It is suggested that assessors start marking immediately after Part 1, and adjust their scores in the light of the candidate's performance in Parts 2 and 3.
- Assessors must reach a decision on the mark they award by the end of the Speaking Examination. It is not acceptable to remain undecided between two bands, so they cannot put a mark as 4/5.
- The awarded marks must always be justifiable, i.e. they should reflect candidate performance in accordance with the assessment criteria.
- A candidate's final mark (i.e. total score) results from summing up the individual marks for each criterion. In this way, a top performance is awarded with
 $7 + 7 + 7 + 7 = 28$ points.

If assessors do not find it easy to decide at first, they should not worry. By the end of the whole training process they will feel more comfortable and be able to apply the scale more consistently and with greater confidence than they could at the start.

THE ASSESSOR'S BEHAVIOUR

During the examination the Assessor should

- preferably sit outside the candidate's range of vision so that s/he does not feel an urge to include him/her in the conversation, but in a position to see and hear both the candidate and the interlocutor clearly;
- not engage in eye contact with the candidate at all;
- talk neither to the candidates, nor to the interlocutor;
- fill in the mark sheet discreetly.

The Assessor's Post-Examination Tasks

After all the examinations on one day the Assessor must

- check that the mark sheet has the overall mark for each candidate as well as the individual mark for each criterion;
- check that each candidate's name and code number are properly recorded.

The Intropack is accompanied by a pre-course video which contains sample performances to assess. Participants are asked to assess the performances and comment on the experience. Sample 2 shows the pre-course tasks participants have to accomplish. The reader can do these tasks using DVD Samples 8.1 and 8.2 (attached to this book). The B2 Level Speaking Assessment Scale and Mark Sheet can be found in Appendix 3. The benchmarks and justifications for DVD Sample 8.1 can be found in Appendix 4.

SAMPLE 2: PRE-COURSE TASKS

PRE-COURSE TASK 1: FAMILIARISING **YOURSELF** WITH THE ASSESSOR'S ROLE

- **Study the B2 Level Speaking Assessment Scale** carefully. This document is attached to this training pack
- **Watch sample speaking test 8.1.** You will find the sample attached to this training pack.
- Take the assessor's role and **evaluate** the candidate's performance **by applying the B2 Level Speaking Assessment Scale**. **Follow** the procedures and instructions described in the "Guidelines for Assessors". Fill in the appropriate boxes of the attached *Mark Sheet* (see Appendix 3). Take notes to justify your marks.
- **Study the final benchmarks and their justifications** attached to this training pack carefully. **Compare** the marks you have given with the ones given by trained assessors.
- **Watch** the performance **again** to observe the candidate's performance **in the light of the justifications**.
- **Report** on how you managed to apply the scale **by completing the following Pre-Course Task Report Sheet below**.
- This task is expected to take ca. 60 minutes.

PRE-COURSE TASK REPORT SHEET (2)

1. What were your general impressions as an assessor?
2. In what ways was it different to assess the candidates' performance according to the speaking assessment scale from your usual assessment techniques?
3. Did any of the criteria cause difficulty? If yes, which one(s) and why?
4. Were your marks different from the ones given by trained assessors?
5. How did the justifications help you to internalise the assessment criteria?

PRE-COURSE TASK 2: TRYING OUT THE ASSESSOR'S ROLE

- **Watch sample speaking test 8.2.** You will find the sample attached to this training pack.
- Take the assessor's role and **evaluate** the candidate's performance **by applying the B2 Level Speaking Assessment Scale. Follow** the procedures and instructions described in the "*Guidelines for Assessors*". Fill in the appropriate boxes of the attached *Mark Sheet* (see Appendix 3). Take notes to justify your marks.
- Bring your completed *Mark Sheet* with your marks and notes for the live training sessions.
- This task is expected to take ca. 25 minutes.

Stage 2: Live assessor training course – a series of workshop sessions

The one-day live assessor training course consists of four 90-minute workshop sessions. The basic material for these sessions is provided by further benchmarked pilot examination performances.

Session 1 deals with the *Intro pack* and the pre-course tasks. Participants compare and discuss their experience about the pre-course task and ask questions concerning the material of the *Introductory Training Pack*.

Next, participants watch and mark Sample 8.2 individually. They are given the Justifications for Sample 8.2 (see Appendix 5), and have the chance to review their home marks. They discuss the marks in small groups of 3 or 4. Finally, the trainer reveals the benchmarks given by a selected team of expert assessors (see Appendix 6). The trainer introduces the idea of using recorded examples of candidate language for interpreting the scale (see Appendix 7) to help the assessment procedure.

In **Session 2**, future assessors have the opportunity to raise their awareness of the marking process and practise using an analytic rating scale for assessing speaking performances.

The trainer tells participants that he/she is going to stop the DVD after each part of the test to allow the participants to give/modify their marks after each part of the examination. Participants are also reminded that it is only a training technique; it won't be like this in real test situations. Participants do the marking individually. Participants watch a sample A2/B1 level performance. (The reader can watch DVD Sample 8.3, and use the A2/B1 Level Speaking Assessment Scale and Mark Sheet in Appendix 2.)

While participants are watching the DVD, the trainer writes the following questions on the board to generate ideas:

How did you arrive at your final mark?

Did you change / modify the mark you gave after the 1st, the 2nd and the 3rd part? Why? / Why not?

Which criterion of the scales did you primarily focus on while watching Part 1, 2, 3 of the examination? Why?

After having seen and marked the sample, participants are asked to report on how they gave the marks. Conclusions are drawn in a plenary discussion. Justifications (see Appendix 8) are handed out in the light of which participants discuss and agree on marks within their small groups. Finally, benchmarks (see Appendix 9) are revealed by the trainer on OHT.

In the next phase of Session 2 participants mark another sample performance individually. (The reader can watch DVD Sample 8.4, and use the A2/B1 Level Speaking Assessment Scale and Mark Sheet in Appendix 2.) Participants discuss their marks in small groups; they justify their individual marks, but finally have to come to an agreement on final marks within the group. Small groups reveal their agreed marks and the trainer enters them in a comparative chart. The trainer reveals Benchmarks and Justifications for Sample 8.4 (see Appendix 10) and gives feedback on the groups' achievements.

In **Session 3** participants get more practice in marking and standardise their assessment. They internalise the analytic rating scale by writing justifications. Finally, they compare tasks and performances at different levels.

Participants mark another performance (The reader can watch DVD Sample 8.5, and use the A2/B1 Level Speaking Assessment Scale and Mark Sheet in Appendix 2.) individually. Before they start watching the DVD, they are strongly encouraged to take notes. After the viewing, the trainer enters the individual marks into a comparative chart on the flipchart. Finally, benchmarks are revealed (see Appendix 11).

If the individual marks differ markedly from each other or if they turn out to be greatly different from the benchmarks, the trainer must lead a brief discussion. In this s/he should keep referring to the band descriptors of the scales, and remind participants of the function of the empty bands. If needed, s/he might read out some paragraphs of the Guidelines for Assessors (*Introductory Training Pack*).

In groups of 4 participants are invited to write justifications on OHTs for all 4 benchmarks of the sample performance. The trainer should make it clear that in real examinations it is not part of the assessor's role to write justifications. This activity primarily serves training purposes, which is to help participants internalise the scale. That is why the trainer should encourage the participants to use the band descriptors of the scales as well as their own notes while writing their justifications.

A volunteer from each group reports their justification for one criterion of the scale. Finally, the trainer reveals justifications written by experts (see Appendix 12).

The last part of Session 3 is devoted to a comparison of performances at two different levels, A2/B1 and B2. Participants watch another sample performance, which is this time at level B2 (The reader can watch DVD Sample 8.6, and use the B2 Level Speaking Assessment Scale and Mark Sheet in Appendix 3.). While watching the DVD, participants take notes. Afterwards they discuss the differences in the tasks and in the performances in small groups.

With the help of an OHT a spokesperson from each small group tells the others what they found out in connection with the differences between the two levels. The trainer comments on the presentations and underlines the main differences,

referring to the Council of Europe Framework Scales (see Chapter 1) and the examination specifications.

In **Session 4** trainees raise problems and offer solutions in the framework of a so-called “assessors’ speaking clinic”, then summarise what they have learnt during the training in a plenary discussion.

Participants are asked to write “What shall I do..?” questions in pairs, bringing up the problems that they still have, pointing out areas that still need further clarification. The questions/problems are written on large sheets of paper. When ready, they are stuck on the wall for other participants to read. Each participant is invited to write suggestions/answers to them.

Each small group gets a few posters with problems and possible solutions. They discuss them and prepare an OHT. A volunteer reports from each group. The trainer reacts, gives their own comments, answers or explanations. The main aim is to reach a common understanding within the whole group.

Finally, participants are asked to draw a suitcase on an OHT and ‘pack in it’ everything that they think they have learnt, or which they have become more confident about during the assessor training sessions. A volunteer from each group shows their suitcase to the whole group. It is important to note that this is the last opportunity for the trainer to add comments or offer solutions if needed.

Stage 3: Post-course task

The participants’ post-course task is a practical application of the acquired skills in mock speaking tests in their own environment. This enables trainees to get more practice and confidence in assessment.

Following the live training course trainees are asked to try out the examiner’s role in mock examinations organised at their own school or at another secondary school in their area. If possible, they try out the examiner’s role in pairs with another participant on the course. They are asked to conduct and assess two tests in turns with a colleague. After the exams they must write a report on their experiences using the provided Mark Sheet (see Sample 3) and Feedback Sheet (see Sample 4).

SAMPLE 3: MARK SHEET FOR THE POST-COURSE SPEAKING EXAMINATION

Candidate’s name	Booklet	Communi- cative impact	Grammar & coherence	Vocabulary	Sound, stress, intonation	Comments (Candidate’s behaviour and performance, inter- locutor’s behaviour, exter- nal assessor present, observer present, difficul- ties, etc.)

SAMPLE 4: POST-COURSE FEEDBACK SHEET

FEEDBACK ON POST-COURSE SPEAKING EXAMINATIONS FROM ASSESSORS / INTERLOCUTORS

Name:

School where you acted as an examiner:

1. How successfully did the Interlocutor / Assessor training prepare you for your roles at the examinations? Please circle the appropriate rating.

1.1. Interlocutor's role

1	2	3	4	5
unsatisfactorily			very well	

PLEASE COMMENT ON YOUR RATINGS.

1.2. Assessor's role

1	2	3	4	5
unsatisfactorily			very well	

PLEASE COMMENT ON YOUR RATINGS.

2. Please comment on the administration of the speaking examinations.

3. How did you manage to follow the Interlocutor Frame?

4. Was there anything that you were supposed to do according to the Guidelines for Interlocutors but you were unable to? If yes, what was it? Please, give reasons.
5. Did you have any problems with timing? If yes, what exactly?
6. Did you have any difficulties in using the rating scale? If yes, please be as specific as possible.
7. How did you go about giving a score for each assessment criterion?
8. In the light of your experiences with the Speaking Examination, are there any aspects concerning the training of interlocutors and assessors that you would like to see changed?
9. Any other comments:

Conclusion

It is impossible to become a trained assessor without formal training. Training should involve both distance and face-to-face elements to ensure that future assessors go through every phase of the difficult and complex standardisation process. However, one training course is not enough: much practice is needed to acquire this special skill. This Handbook, together with the recorded sample performances, provides a good starting point for those interested in becoming an assessor for a speaking examination, but live, face-to-face training is essential if reliable assessments are to be made.

Recommendations for Good Practice

1. Classroom Assessment

As this Handbook is intended to help teachers to develop and conduct modern English oral examinations, we have presented guidelines for designing good speaking tasks and discussed ways of ensuring that the administration of speaking exams yield valid and reliable assessments of candidates' oral abilities. Although the context for all our discussion has been oral proficiency testing, we firmly believe that many of the ideas presented in this Handbook are transferable to classroom contexts as well. For example, we recommend the use of analytic assessment scales in classroom oral tests but at the same time we also wish to underscore the importance of applying the scales consistently in order to ensure assessor reliability. Teachers are also advised to consider carefully how they will elicit language from their learners as specific examiner behaviour has been shown to affect candidates' performance negatively. The guidelines for designing different speaking tasks may also be useful when teachers wish to write speaking tests for their own classrooms or would like to provide practice activities for their students who are preparing for a specific modern European speaking exam. Exam preparation or candidate training is very important as it is believed to enhance candidates' performance for at least two reasons: they will be clear about what is involved in doing a specific test task while at the same time familiarity with the task type may reduce their anxiety.

1.1 General Guidelines for Designing Speaking Tasks

We would like to emphasize that the item writing guidelines included in this Handbook were developed gradually over six years as part of the Hungarian Examination Reform Teacher Support Project, which piloted a number of speaking tasks (some of them more than once) not merely to try them out but also to check whether the guidelines for item writers needed any modification. When evaluating how tasks worked and whether the guidelines were appropriate and suitably detailed, we always bore the following criteria in mind:

Speaking tasks should

- measure the intended proficiency level and skills;
- provide the widest scope for candidates to perform;
- provide candidates with an opportunity to contribute equally to the exchange (in the paired format);
- be realistic in the sense that they should replicate real life language use as closely as possible;
- explore themes that candidates can relate to and/or on which they can express an opinion;

- be fair in the sense that they should not be biased towards a specific group of candidates;
- generate a rateable language sample from candidates.

1.2 Differences between Test Tasks and Teaching Tasks

As has already been suggested, different task types will elicit different aspects of oral proficiency, therefore it is recommended that a range of tasks should be used in order to assess candidates' overall speaking abilities. However, we have also argued that the task type in itself cannot guarantee that candidates will be able to perform to the best of their abilities, as specific task features may negatively affect their performance. These may include, for example, using inappropriate visual prompts in picture description activities, or the lack of a suitable task frame in discussions, or giving too much guidance for the interactants in role plays. If teachers wish to use speaking tasks for assessment purposes in their own classrooms, they should be aware that not all teaching tasks can be used for testing purposes. In fact, test tasks should be distinguished from teaching tasks as students' performance on the former will have certain consequences (they will get a pass grade or will be failed), and so test tasks need to satisfy specific criteria such as the ones above. Teaching tasks, however, are usually intended only to provide language practice opportunities for learners, and so they can range from very controlled to free production activities. Controlled practice activity often involve repetition and provide a (fairly) limited scope for original contributions from the students. For instance, this is the case when students are asked to memorize and recite different characters' parts of a dialogue from the course book. A less controlled activity will allow more freedom to students as to what they say exactly and how long they speak. For example, students can be asked to act out a specific situation in pairs (e.g. 'making a complaint in a restaurant'), in which the task instructions

- clearly describe the setting (Student A is the waiter, Student B is the customer);
- specify what the customer wants to achieve (have the cold soup replaced by hot soup and have the dirty table cloth replaced with a clean one);
- specify what the outcome of the exchange should be (waiter apologizes and brings a plate of hot soup).

If there is also a set of lexical chunks (e.g. "Would you mind ..."; "I'm sorry to say this, but ..."; "I'm extremely sorry for ...";) given for students to incorporate into the above role play, the activity will clearly be a teaching task as the purpose of the activity is to provide contextualized practice for the selected phrases that can be used to express requests, complaints and apologies.

Because language practice for learners is usually maximized through pair and group work activities in the classroom, it is important to note that not all pair or group work activities will be suitable for assessment purposes. Although many

speaking activities may allow for students' own contributions (e.g. when students are asked to discuss the most useful technical inventions of the last 50 years), they will not ensure automatically balanced contributions by the participants, which is an essential feature of paired-task performance. They may equally fail to help participants perform for a pre-specified time limit in order to get a rateable amount of language sample. If teachers follow the guidelines for item writers suggested in this Handbook, they will have a better chance of designing speaking tasks that are capable of eliciting an adequate sample of varied and life-like language performance from their learners.

2. Washback

It is important for teachers to be aware of the washback effects of modern European language exams: they should have a positive impact on the quality of English language teaching. Standardised examinations carefully designed in line with current principles in language assessment are expected to reinforce good teaching practice. Therefore, when preparing students for them, teachers should not change their teaching routine provided it is based on a balanced approach to the development of the four skills and learners are encouraged to use the language in meaningful and relevant contexts in order to realize their own communicative goals. If classroom practices do not reflect such an approach to language teaching, high quality exams may make teachers change what they do in the classroom, especially if they are made to understand the rationale for the examination and receive training in terms of how to teach towards the exam. In such cases, teachers need to learn how to provide their learners with practice targeted at those skills, language use contexts and tasks that will feature in the examination.

3. Measures of Ongoing Quality Control

Modern European language examinations should not only generate positive washback in the classroom, they should also ensure that learners can pass through borders freely and enter the European educational and labour market. Modern European examinations provide recognised language qualifications because candidates' achievements are related to well-defined levels of foreign language proficiency, the Common European Framework of Reference. However, modern European language examinations have to employ a set of rigorous quality control procedures that can guarantee that a language certificate issued after passing a specific examination is valid and meaningful as it gives precise information for the user (e.g. future educational institution or employer) about the actual foreign language abilities of its owner.

In the case of speaking abilities, it is essential to employ ongoing quality control measures as the degree of subjectivity involved in the elicitation and assessment procedures has to be minimised. In other words, candidates should be given equal and fair treatment both in terms of test administration and assessment. Based on the experience of the Hungarian Examinations Reform Teacher Support Project, we

recommend the following quality control procedures that are fully in line with the requirements set for modern European speaking examinations:

As compulsory elements of quality control, Test Specifications should

- a) clearly indicate the level the examination is aimed at. This level should be specified with reference to the Common European Framework of Reference;
- b) specify the assessment procedure in detail, including the marking scheme used for assessment;
- c) give detailed instructions and guidelines for conducting the speaking examinations (use of an *Interlocutor Frame*).

Trained item writers should be commissioned to design speaking tasks for all parts of the examination and they should be provided with detailed guidelines for item writing. These guidelines should contain precise information about all aspects of the tasks such as standardised rubrics, task frame, topic areas, quality and quantity of verbal/visual prompts, potential roles for participants in role plays, layout requirements, etc.

Speaking tasks should undergo a thorough evaluation process done by experts (called the 'editing committee'). The editing committee should check whether the tasks conform with all aspects of the item writer guidelines and Test Specifications, and decide whether they are suitable for the speaking examination in question. If there are any problems, they should offer instructions for revising the tasks, or – if the task is completely inadequate – refuse to accept it.

All speaking task should be piloted with real language learners before using them in live examinations. For high-stakes language examinations, such as a national school-leaving examination, the number of candidates taking part in the pilot exams should be representative of the whole candidate population. It is very important to have some of the pilot examinations video-recorded for interlocutor/assessor training purposes.

The results of the pilot examinations together with feedback collected from test takers should be processed and analysed. Only those tasks should be selected for use in future live exams that have proved to elicit adequate language samples at the intended level. Problematic tasks will have to be revised and piloted again.

A team of experienced trainers should select videoed sample performances demonstrating different levels of candidate performance, standard and non-standard interlocutor behaviour for training purposes. Training materials may need to be updated and revised from time to time, especially if there are changes in the Test Specifications.

The sample performances to be used for standardising the assessment procedure should undergo benchmarking. The benchmarking exercise is an essential element of the quality control system for speaking examinations, since only through benchmarked performances is it possible to standardise the assessment procedure.

All examiners – interlocutors and assessors – should be properly trained for their roles. Examiner training is a key element of quality control, since standardised procedures for both conducting the speaking examination and assessing the performances can be assured only through detailed and thorough training (see *Chapters 7 and 8*). This is the only way to acquire the necessary interlocuting and assessing skills.

Before administering and assessing live exams, future examiners should have some practice in structured and monitored mock examinations. Professional feedback on examiner performance is a stage in the quality control system of speaking examinations that cannot be left out since it provides invaluable opportunities for fine-tuning examiner behaviour in order to achieve the best possible standardised examiner performance.

Live examinations should be monitored continuously. A chief examiner should observe a sample of live exams and/or a random sample of exams can be recorded on tape/video for later inspection or analysis. Scores given by assessors should also be processed and analysed for examiner consistency. If some interlocutors are found to deviate from the standardised way of conducting exams or some assessors are found to score performances erratically and inconsistently, they will have to be retrained. However, even experienced examiners will benefit from further training opportunities as these will increase their self-awareness as well as help maintain the standardisation of the speaking exam.

It is hoped that the recommendations listed above will help language testers to adhere to modern European standards as certificates from exams that can be shown to be valid and reliable will be of considerable value to learners in the future Europe.

APPENDICIES

Appendix 1.

Self-Assessment Statements for Speaking¹

Overall Speaking Scale

Level	Statement
A1	I can use simple expressions and sentences to describe where I live and people I know. I can communicate in a simple way if the other person is willing to help me and repeat or say things again more slowly. I can ask and answer simple questions about familiar things in everyday situations.
A2	I can use expressions and sentences to describe in simple terms family and friends and other people, living conditions (where and how I live), my education, and my present or most recent job. I can communicate in everyday situations which involve short and simple exchanges about familiar things, but I may not know enough to keep the conversation going.
B1	I can describe in a simple way my experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can tell a story or describe the plot of a book or film and describe my reactions. In conversation, I can talk unprepared on topics of personal interest or everyday life (e.g. family or friends, hobbies, work, travel and current events).
B2	I can present a clear, detailed description on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue and give the advantages and disadvantages of various options. I can take an active part in discussion in familiar situations and can express and give reasons for my views.
C1	I can describe complex subjects clearly and in detail, paying attention to the various aspects of the subject. I can develop particular points and round off with a conclusion. I can talk fluently and spontaneously, having to search for expressions relatively seldom. I can use the language effectively both socially and for job purposes. I can adjust what I say to things other people say and express my thoughts precisely.
C2	It is very easy for me to take part in any conversation and discussion and I am familiar with the right expressions and also colloquial language. I can talk fluently and express the things I want to say very precisely. I can describe and explain things clearly and smoothly and in a way that is appropriate to the situation. I can structure what I say in such a way that it helps the listener to notice and remember important points. If I have a problem, I can say the same thing in a different way without the listener even noticing it.

¹ The Self-Assessment Scales were originally developed and calibrated by Brian North and were further developed within the framework of the DIALANG project supported by the European Commission under the EU's education programme SOCRATES.

Main Speaking Scale

Level	Statement
A1	I can handle simple numbers, quantities, cost and time.
A1	I can ask and answer simple direct questions about myself and other people, about where I live, people I know, and things I have.
A1	I can reply in an interview to simple direct questions spoken very slowly and clearly in direct simple speech about personal details.
A1	I can buy things in uncomplicated situations where pointing and other gesture can support what I say.
A1	I can ask people for things and give people things they ask for.
A1	I can make an introduction and use basic greeting and leave-taking expressions.
A1	I can ask and answer simple questions, formulate and respond to simple statements in areas of immediate need or about very familiar things.
A2	I can greet people, ask how they are and react to simple things people are telling me.
A2	I can discuss in a simple way what to do, where to go and make arrangements to meet.
A2	I can handle very short social exchanges but I am rarely able to understand enough to keep a conversation going.
A2	I can discuss everyday practical issues in a simple way when the other person(s) speak(s) clearly and slowly.
A2	I can say in a simple way what I think about things when I'm addressed directly in a formal meeting, if I can ask for repetition of important points if I need it.
A2	I can ask and answer questions about what people do at work and in the free time.
A2	I can make and respond to invitations and apologies.
A2	I can make and respond to suggestions.
B1	I can express and respond briefly to feelings such as surprise, happiness, sadness, interest and indifference.
B1	I can keep up a conversation or discussion but I may sometimes be difficult to follow when I am trying to say exactly what I would like to.
B1	I can initiate, keep up and close simple face-to-face conversation about things that are familiar or of personal interest to me.
B1	I can take part in conversations about familiar things even if I have not had a chance of preparing them.

Level	Statement
B1	I can manage to handle most situations likely to arise when making travel arrangements abroad through an agent or when actually travelling.
B1	I can handle less routine situations on public transport, for instance asking a passenger where to get off for an unfamiliar destination.
B1	When I can't think of a word I want, I can use a simple word meaning something similar and ask for the correct word.
B1	I can give personal accounts of experiences, describing in some detail feelings and reactions.
B1	I can briefly explain and give reasons for my plans, intentions and actions.
B2	I can talk to native speakers without amusing or irritating them, if I don't mean to, or without requiring them to behave other than they would with a native speaker.
B2	I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
B2	I can speak the language naturally, fluently and effectively in familiar situations.
B2	I can express and support my opinions in discussion about familiar matters by providing relevant explanations, arguments and comments.
B2	I can take active part in extended conversation about most things of general interest.
B2	I can describe or define something concrete for which I cannot remember the word, for example 'a parking ticket', 'a credit card', or 'an insurance policy'.
C1	I know the language well enough to be able to find a solution to a dispute (e.g. an undeserved traffic ticket, financial responsibility for damage in a flat, for blame regarding an accident).
C1	I can express myself fluently and spontaneously on a broad range of topics, almost without an effort.
C1	I can use the language flexibly and effectively for social purposes, including emotional, joking usage and references to well-known literary or other sources.
C1	I can express my ideas and opinions clearly and precisely, and can present and respond to complex lines of reasoning convincingly.
C1	I have a good command of a wide variety of words and phrases, allowing me to overcome any difficulty with words I may not know.
C1	I can keep up with a debate even if it is about abstract and complex things which I am not familiar with.

Level	Statement
C2	I can go back and reword a difficult point so smoothly that the other person(s) is/are hardly aware of it.
C2	I can produce clear and smoothly flowing speech which is structured in an effective and logical way.
C2	I can present a complex topic confidently and clearly to listeners who are not familiar with it, organising and adapting what I say in a flexible way to meet the needs of the listeners.
C2	I can speak the language comfortably and correctly, and it does not hinder me in any way in my social and personal life.
C2	I speak the language almost as well as I speak my mother tongue.
C2	I can express myself in such a way that even native speakers think I must have lived in the country for a long time.

Appendix 2 A2 B1 Level Speaking Assessment Scale

	COMMUNICATIVE IMPACT	GRAMMAR AND COHERENCE	VOCABULARY	SOUNDS, STRESS, INTONATION
	Candidate...	Candidate...	Candidate's vocabulary...	Candidate
7	<ul style="list-style-type: none"> • makes natural hesitations when searching for ideas • requires no additional prompting • makes relevant contributions 	<ul style="list-style-type: none"> • uses wide range of structures • uses mostly accurate grammar • makes coherent contributions 	<ul style="list-style-type: none"> • has wide range is fully appropriate 	<ul style="list-style-type: none"> • is understood with ease • uses mostly accurate and appropriate sounds and stress • uses a wide range of intonation to convey meaning effectively
6				
5	<ul style="list-style-type: none"> • makes hesitations when searching for language • requires no additional prompting • generally makes relevant contributions 	<ul style="list-style-type: none"> • uses adequate range of structures • makes frequent minor mistakes only • makes mostly coherent contributions 	<ul style="list-style-type: none"> • has adequate range • is generally appropriate with isolated inappropriacies 	<ul style="list-style-type: none"> • is understood easily with isolated difficulties • makes mistakes in sounds and stress which occasionally affect comprehensibility • uses an adequate range of intonation to convey meaning mostly effectively
4				
3	<ul style="list-style-type: none"> • makes frequent hesitations • requires some additional prompting • occasionally makes irrelevant contributions 	<ul style="list-style-type: none"> • uses limited range of structures • makes occasional major and frequent minor mistakes • makes contributions with limited coherence 	<ul style="list-style-type: none"> • has limited range • is generally appropriate with occasional disturbing inappropriacies 	<ul style="list-style-type: none"> • is understood with some strain • makes mistakes in sounds and stress which seriously affect comprehensibility • uses a limited range of intonation to convey meaning
2				
1	<ul style="list-style-type: none"> • makes long intrusive hesitations • requires major additional prompting • makes mostly irrelevant contributions • no assessable language 	<ul style="list-style-type: none"> • uses very limited range of structures • makes frequent major and minor mistakes • makes mainly incoherent contributions • no assessable language 	<ul style="list-style-type: none"> • has very limited range • is frequently inappropriate • no assessable language 	<ul style="list-style-type: none"> • is understood with constant strain • mostly uses sounds and stress that are difficult to understand • makes little use of intonation to convey meaning • no assessable language
0				

Mark sheet for A2/B1

No.	DVD SAMPLE	COMMUNICATIVE IMPACT	GRAMMAR & COHERENCE	VOCABULARY	SOUND, STRESS, INTONATION	TOTAL SCORE	COMMENTS
1.	8.3						
2.	8.4						
3.	8.5						

Appendix 3 B2 Level Speaking Assessment Scale

	COMMUNICATIVE IMPACT	GRAMMAR AND COHERENCE	VOCABULARY	SOUNDS, STRESS, INTONATION
	Candidate...	Candidate...	Candidate's vocabulary...	Candidate...
7	<ul style="list-style-type: none"> • makes entirely natural hesitations when searching for ideas • participates with ease without requiring additional prompting • contributes fully and effectively to the communication 	<ul style="list-style-type: none"> • uses wide range of structures • uses accurate grammar • makes fully coherent contributions 	<ul style="list-style-type: none"> • has wide range • is fully appropriate 	<ul style="list-style-type: none"> • is understood with ease • uses accurate and appropriate sounds and stress • uses a wide range of intonation to convey meaning effectively
6				
5	<ul style="list-style-type: none"> • makes mostly natural hesitations when searching for ideas • equires no additional prompting • in general contributes effectively to the communication 	<ul style="list-style-type: none"> • uses adequate range of structures • makes occasional minor mistakes only • makes adequately coherent contributions 	<ul style="list-style-type: none"> • has appropriate range • is generally appropriate with isolated inappropriacies 	<ul style="list-style-type: none"> • is understood easily with isolated difficulties • uses mostly accurate and appropriate sounds and stress • uses an adequate range of intonation to convey meaning mostly effectively
4				
3	<ul style="list-style-type: none"> • often makes hesitations in order to search for language • in general requires no additional prompting • contributes sufficiently to the communication, occasionally making irrelevant contributions 	<ul style="list-style-type: none"> • uses limited range of structures • makes occasional major and minor mistakes • makes mostly coherent contributions with occasional inconsistencies 	<ul style="list-style-type: none"> • has sufficient but somewhat limited range • is generally appropriate with occasional disturbing inappropriacies 	<ul style="list-style-type: none"> • is understood with some strain • makes mistakes in sounds and stress that occasionally affect comprehensibility • uses a limited range of intonation to convey meaning
2				
1	<ul style="list-style-type: none"> • frequently makes intrusive hesitations when searching for language • requires some additional prompting • frequently makes irrelevant contributions or contributes little to the communication 	<ul style="list-style-type: none"> • uses very limited range of structures • makes occasional major and frequent minor mistakes • makes mainly incoherent contributions 	<ul style="list-style-type: none"> • has limited range • is frequently inappropriate 	<ul style="list-style-type: none"> • is understood with constant strain • makes mistakes in sounds and stress that seriously affect comprehensibility • makes little use of intonation to convey meaning at all
0	<ul style="list-style-type: none"> • no assessable language 	<ul style="list-style-type: none"> • no assessable language 	<ul style="list-style-type: none"> • no assessable language 	<ul style="list-style-type: none"> • no assessable language

Mark sheet for B2

No.	DVD SAMPLE	COMMUNICATIVE IMPACT	GRAMMAR & COHERENCE	VOCABULARY	SOUND, STRESS, INTONATION	TOTAL SCORE	COMMENTS
1.	8.1						
2.	8.2						
3.	8.6						

Appendix 4

Benchmarks and Justifications for DVD Sample 8.1

Communicative impact	Grammar and coherence	Vocabulary	Sound
4	5	5	5

COMMUNICATIVE IMPACT: 4

The Candidate makes **natural hesitations** when he is searching for ideas, at times language. In Part 3 he is not really successful in generalisation, he prefers to talk about his personal experience. **In general, he requires no additional prompting**, but in Part 3 he completely misunderstands the question concerning book adaptations. Though he is a slow speaker, he manages to maintain his flow of speech.

GRAMMAR AND COHERENCE: 5

The Candidate uses an **adequate range** of structures appropriately: tenses, modal auxiliaries, conditionals, passive and combinations of these: *Whenever I travel to an English speaking country; This picture might be from the early 50s; They need to know the expression to be able to listen to the problem of the person who visit ... visits them*. He makes only **occasional minor mistakes**: *this different kinds of drinks; being in a nature; he watches the TV*. Though there are examples of major mistakes, too: *today's children not read as much as children in the past used to; he get to know many new things of the world*, these mistakes are considered as samples of lack of careful language use, since several times during his test he demonstrates accurate use of these structures. He **makes coherent contributions**, though when he is searching for ideas, he happens to give short utterances in response: Interlocutor:

Which products are advertised too often?

Candidate: *Too often? I think cigarettes.*

VOCABULARY: 5

The **range** of vocabulary items the Candidate uses is **appropriate** for the level: *for instance; a challenge to understand; sophisticated; reserved; It could depict*, but there are **isolated inappropriacies**: *by themselves; between the break of two series; I play the computer; movement is very important*. At times he seems to be pausing in order to find the words he needs but then manages perfectly by circumscribing it: *the person who visits them* [instead of patient]; *the speakings*[conversations] *with English-speaking people*.

SOUND, STRESS, INTONATION: 5

The Candidate's pronunciation is understood easily with **isolated inappropriacies in stress or sound**: *industry* [in'dʌstri]; *picking* [pɪ:kɪŋ]; *events* [ɪ:vənts]. He is a slow speaker but manages to **convey meaning mostly effectively** in his speaking test.

Appendix 5

Justifications for DVD Sample 8.2

COMMUNICATIVE IMPACT:

The Candidate takes an active part in the conversation, and **contributes fully and effectively to the communication** throughout the speaking test. She communicates with ease and **requires no additional prompting**. Her contributions come in a very natural manner, she uses a number of conjunctions and fillers to make her contributions sound life-like: *actually; definitely; besides;* as well as to emphasise the main idea: *I think people do need each other.*

GRAMMAR AND COHERENCE:

The Candidate uses a **wide range of structures** and demonstrates good use of tenses, modals, passive and conditional structures, as well as their combinations: *I might choose university; They should definitely help their mother; People do need each other* but makes **major mistakes** in some of them – mistakes that impede understanding: *almost all women works; If I will become a doctor; I had to make friends* [instead of I would have to make friends]. Quite a number of minor mistakes are committed: *teenagers should see that their parents are working; I was always interested in European history.* Her contributions are **not fully coherent**, but there are only rare examples of that: *It's hard to be, for example, for a doctor to be with children.*

VOCABULARY:

The **range** the Candidate uses is **wide** and enables her to meet the requirements of the tasks: *you have to adjust to that; cope with new situations; the state can't afford to pay doctors; according to their origins.* There are **isolated inappropriacies**: *out there* [instead of abroad], *economical university; funner; look after* [instead of look for]; *I would found* [start] *a family; a pair of people* [a couple]. Although she produces enough language to assess, owing to the number and nature of the mistakes this performance does not fall in the top band.

SOUND, STRESS AND INTONATION:

The Candidate is **understood with ease**, she uses **accurate and appropriate sound and stress** with only very few isolated examples of the opposite: *faculty* ['fækjʊlti]; *cuisine* ['kɪziːn]. She uses a **wide range of intonation** to convey meaning effectively.

Appendix 6

Benchmarks for DVD Sample 8.2

Benchmark for communicative impact	Benchmark for grammar and coherence	Benchmark for vocabulary	Benchmark for sound, stress, intonation
7	5	6	7

Appendix 7

Examples of Candidate Language for Interpreting the A2/B1 Level Speaking Assessment Scale (selected from the sample DVD performances)

GRAMMAR AND COHERENCE

wide range of structures

confident use of the tenses, the passive and conditional sentences, modal auxiliaries, comparatives and superlatives

the child is being helped by his mother; she is studying by herself; if it was me who lived here; it's worth going there.

major mistakes

constant inappropriate use of tenses, inappropriate form of agreement, inappropriate forms of the comparative adjective

we are walked; two people loves; they are watch the flowers; we not wear; the teachers are want the best about; they can talking; we are talking about everything; we are going to Balaton [the latter two meaning regular activities]; more smaller; I put pictures / posters [instead of I would put]; I will want a desk because she hasn't got;

minor mistakes

wrong word order, the wrong use of articles, uncountable nouns and prepositions

I don't know where will I go; on these pictures; in the autumn; like you see; advise going; : to play with our [us]; I don't know which kind of music do you like; a luggage; in downstairs;

VOCABULARY

wide range

it won't work; get to know; break up with; by herself; kind of like my room; get together; scattered; the ship sank;

inappropriacies which do not disturb understanding

telephone line [telephone cord]; in Monday; they feel very well themselves; they are in the nature; classic music

disturbing inappropriacies

lecture [tutor or teacher]; name [title]; they speak Hungarian [the film is dubbed]; with his woman [instead of mother]; something more culture; trousers made from jeans; economical [instead of economist]; your conditional [instead of physical condition]

SOUND, STRESS, INTONATION

mistakes in sounds and stress which occasionally affect comprehensibility

bilingual ['bɪlɪŋwɒl]; event ['ɪ:vənt]; to [tʊ]; very [wæɪ]; third [tə:d]; weather ['vedər]

Appendix 8

Justifications for DVD Sample 8.3

COMMUNICATIVE IMPACT:

The Candidate **communicates mostly effectively**. He **makes hesitations only sometimes** when searching for a word, **but it's not disturbing**. However, **he makes some irrelevant contributions** and **needs some additional prompting**: *I think the sport is the most healthy in Hungarian, because your conditional will be good*. In Part 2 the comparison could have been more elaborated (the main difference was not mentioned, he just described the pictures). In Part 3 he makes some attempt to give advice: *you can; you should*. Some utterances are difficult to judge because of language problems: *I like the economical or the other thing is to be a doctor; let's see the different of the pictures*.

GRAMMAR AND COHERENCE:

The Candidate's performance is rather extreme with regard to grammar and coherence. He uses the right tense most of the time (e.g. he always uses the present continuous for picture description), whereas **some of his utterances lack any kind of coherence**: *something more culture; doing something meal; the other one just draw*. He uses a **limited range of structures with occasional major and frequent minor mistakes**: *That's very big and not so thin so it's very thick; I would talk about the picture and say the different of the pictures; he's don't help her*. He uses wrong word order: *I don't know what's the time and the wrong preposition: in the same time*.

VOCABULARY:

The Candidate is equally extreme with his use of vocabulary. The range the Candidate tries to use is considerable – and so is the **number of mistakes he makes**. In many cases he searches for the word and uses a description: *fashion style* [instead of trend]; *something more culture; trousers made from jeans*. He frequently confuses verb forms and noun forms of the same lexical item, as well as nouns with adjectives: *economical* [instead of economist]; *the wife will have an argue with his husband; your conditional* [instead of physical condition].

SOUND, STRESS, INTONATION:

The Candidate **is understood easily with isolated difficulties**. These isolated difficulties mainly come from the unclear, hasty articulation of words at the end of sentences. He **makes some mistakes in sounds which affect comprehensibility**, the influence of German can be felt. Though he attempts to use intonation patterns, his intonation tends to be rather 'flat'.

Appendix 9

Benchmarks for DVD Sample 8.3

Benchmark for communicative impact	Benchmark for grammar and coherence	Benchmark for vocabulary	Benchmark for sound, stress, intonation
4	3	3	4

Appendix 10

Benchmarks and Justifications for DVD Sample 8.4

Benchmark for communicative impact	Benchmark for grammar and coherence	Benchmark for vocabulary	Benchmark for sound, stress, intonation
7	7	6	7

COMMUNICATIVE IMPACT: 7

The Candidate communicates with ease and **makes relevant contributions** throughout the whole test. Her responses are quick and natural: *the buses are really rare. She makes entirely natural hesitations when searching for ideas* using adequate fillers: *well; kind of; so; actually; you see* in the meantime. **She requires no additional prompting.** The Candidate initiates and responds adequately.

GRAMMAR AND COHERENCE: 7

The Candidate **uses a very wide range of structures** appropriately. She is a confident user of the tenses, the passive and conditional sentences. There are examples of subtle use, which are certainly considered top performance at the intermediate level: *the child is being helped by his mother; she is studying by herself; if it was me who lived here; it's worth going there.* **She uses mostly accurate grammar,** although some minor errors occur: *on these pictures; in the autumn; like you see; advise going.* **All her contributions are coherent.**

VOCABULARY: 6

The Candidate uses **a wide range** of vocabulary: *it won't work; get to know; break up with; by herself; kind of like my room; get together; scattered.* However, **there are some isolated inappropriacies.** She uses Hungarian words twice: *nazarénusok; gimnázium* and makes a minor mistake when using: *telephone line* (telephone cord). Apart from these slips her performance demonstrates an attempt to go beyond the vocabulary resource required at intermediate level.

SOUND, STRESS, INTONATION: 7

The Candidate **is understood with ease.** She uses **mostly accurate and appropriate sounds and stress.** Only a few inappropriately pronounced words occur, which are still comprehensible: *bilingual* ['bɪlɪŋwɒl]; *event* ['ɪvənt]. She uses the American accent consistently. **She uses a wide range of intonation to convey intended meaning effectively.**

Appendix 11

Benchmarks for DVD Sample 8.5

Benchmark for communicative impact	Benchmark for grammar and coherence	Benchmark for vocabulary	Benchmark for sound, stress, intonation
2	2	3	4

Appendix 12

Justifications for DVD Sample 8.5

COMMUNICATIVE IMPACT: 2

The Candidate's participation in all 3 parts of the test is passive and reluctant. She makes long intrusive pauses and **requires major additional prompting** on the part of the interlocutor. Some of her **contributions are irrelevant**:

Interlocutor: Why did you choose to study at the school where you study?

Candidate: *At Deák Ferenc High School.*

In a number of cases her responses are limited to *Yes/No* or she merely repeats the interlocutor's words.

She is reluctant to elaborate on a topic, produces short and simple utterances. She gives up her passive attitude in Part 3 where she shows some interaction asking *Do you like discos or cafes?*

GRAMMAR AND COHERENCE: 2

The Candidate **uses a limited range of structures** and **makes major mistakes**: lack of use of conditional structures: *I put pictures / posters* [instead of *I would put*]; *I will want a desk because she hasn't got*; inaccurate use of tense: *it's depends on what the person like*. There are **frequent minor mistakes** as well: misused aspects (*I'm kayaking 5 time a week*); prepositions (*Shall I talk the furniture?; on the picture*). **Coherence is extremely limited**, most contributions are very short, she uses chunks not whole utterances, she doesn't speak much.

VOCABULARY: 3

The Candidate uses a **limited range** of vocabulary, which is **generally appropriate** (*settee; squash; temple*) **with occasional disturbing inappropriacies** (*with his woman* [instead of *mother*]; *culture centre*). She probably does not understand the word '*tidy*' as she gives an inadequate answer when saying '*yes*' as the room is obviously untidy. She cannot finish a sentence owing to missing vocabulary.

SOUND, STRESS, INTONATION: 4

The Candidate's delivery is much slower than the natural flow of speech, therefore intonation loses its function to convey meaning effectively. However, she is **understood with ease** and her mistakes in sounds only occasionally affect comprehensibility: *th* sound in *three*, the pronunciation of *squash, furniture*. Her intonation is flat, and she tends to raise the end of words.

Appendix 13

List of Reference Books and Recommended Readings

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.
- Alderson, J. C., Nagy, E. & Öveges, E. (2000). *English Language Education in Hungary Part II. – Examining Hungarian Learners' Achievements in English*. Budapest: The British Council Hungary.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford University Press.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing* 20, 1–25.
- Csépes, I. (2002) Is testing speaking in pairs disadvantageous for students? A quantitative study of partner effects on oral test scores. *novELTy* 9(1), 22–45.
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. (2001). Council of Europe, CUP.
- Együd, Gy., & Glover, P. (2001). Oral testing in pairs – a secondary school perspective. *ELT Journal*, 55, 70–76.
- Együd, G., Gál, I. & Glover, P. (Eds.) (2001). *English Language Education in Hungary Part III. – Training Teachers for New Examinations*. Budapest: The British Council Hungary.
- Fekete, H., Major, É. & Nikolov, M. (Eds.) (1999). *English Language Education in Hungary. A Baseline Study*. Budapest: The British Council Hungary.
- Fulcher, G. (2003). *Testing Second Language Speaking*. Longman.
- Fulcher, G. & Reiter, R. M. (2003). Task difficulty in speaking tests. *Language Testing* 20, 321–344.
- Heaton, J. B. (1988). *Writing English Language Tests*. Longman.
- Hughes, A. (1989). *Testing for Language Testers (2nd edition)*. Cambridge University Press.
- Luoma, S. (2004). *Assessing Speaking*. J. C. Alderson & L. F. Bachman (Series Eds.) Cambridge Language Assessment Series. Cambridge University Press.
- McNamara, T. F. (1996). *Measuring Second Language Performance*. Longman.
- Underhill, N. (1987). *Testing Spoken Language*. Cambridge University Press.
- Weir, C. (1990). *Communicative Language Testing*. Prentice Hall International Ltd.

Appendix 14

Contents of DVD

CHAPTER 3: THE INTERVIEW

Samples 3.1 and 3.2 demonstrate how the lack of Interlocutor Frame may influence the Interlocutor's behaviour and the candidate's performance.

Sample 3.3 shows a sample performance on A2/B1 level interview questions.

Sample 3.4 shows a sample performance on B2 level interview questions.

CHAPTER 4: THE INDIVIDUAL LONG TURN

Samples 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 and 4.7 demonstrate some common problems with the selection of picture prompts and task design.

In Samples 4.8, 4.9, 4.10, 4.11, 4.12 and 4.13 well-designed individual long-turn tasks with recommended picture prompts are presented in use.

CHAPTER 5: DISCUSSION TASKS

Sample 5.1 shows the result of the regular procedure of item production (item moderation, editing, piloting and revision) through a sample performance on the final version of a paired discussion task.

Sample 5.2 presents a paired discussion task performance in which the lack of guidelines for interlocutor intervention prevented the interlocutor from facilitating the candidates' performance successfully.

Samples 5.3, 5.4, 5.5 and 5.6 demonstrate recommended tasks for paired discussion activities.

Sample 5.7 shows a well-designed discussion task for the individual mode.

CHAPTER 6: ROLE-PLAY TASKS

Samples 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7 and 6.8 show role-play tasks with problematic task design issues.

In Samples 6.9, 6.10, 6.11 and 6.12 role-play tasks reflecting good practice are presented.

CHAPTER 7: TRAINING INTERLOCUTORS

Sample 7.1 shows standard interlocutor behaviour in an examination conducted in individual mode at B2 level.

Sample 7.2 shows standard interlocutor behaviour in an examination conducted in paired mode at A2/B1 level.

Sample 7.3 shows non-standard interlocutor behaviour in an examination conducted in individual mode at A2/B1 level.

Sample 7.4 shows non-standard interlocutor behaviour in an examination conducted in paired mode at B2 level.

Samples 7.5, 7.6 and 7.7 present unscripted interlocutor behaviour.

Sample 7.8 shows standard interlocutor behaviour in an examination conducted in paired mode at A2/B1 level.

Sample 7.9 shows standard interlocutor behaviour in an examination conducted in paired mode at B2 level.

CHAPTER 8: TRAINING ASSESSORS

Samples 8.1, 8.2 and 8.6 contain sample candidate performances to assess at B2 level.

Samples 8.3, 8.4 and 8.5 contain sample candidate performances to assess at A2/B1 level.