

## On site observation of KPG oral examiners: Implications for oral examiner training and evaluation

*Evdokia Karavas and Xenia Delieza, University of Athens*

*The paper focuses on the development and main findings of an observation project carried out by the Research Centre for English Language Teaching Learning and Assessment (RCEL) of the Faculty of English Studies, University of Athens, within the context of the Greek state exams for foreign language proficiency, known as KPG exams. The project, which was launched in November 2005 and is ongoing, involved on site observation of oral examiners as a means of evaluating and monitoring the KPG oral exams for English; the project aimed at gathering information about, inter alia, the conditions of the oral exam administration, examiner discourse practices, examiner efficiency and conduct. The project involved the design of structured observation schemes used by trained observers during each exam administration (from November 2005 to May 2008). The observation schemes (the categories of which were refined after each exam administration based on the analysis of results) focused mainly on the discourse practices of examiners who assume the role of interlocutor in the KPG oral exams. More specifically, specific observation categories were gradually developed focusing on the changes examiners/interlocutors made to task rubrics and their interventions to candidates' language output. The observation scheme findings have provided a wealth of data concerning the frequency and nature of examiner interventions and have contributed significantly to the refinement and development of the KPG oral examiner training programme.*

### **Introduction: The oral examiner “variable” and its effect on validity and fairness in oral tests**

Oral assessment of language proficiency is a complex and largely subjective process in which many variables or facets have been found to affect the quality and quantity of candidate language output and the rating of their performance ultimately threatening validity, reliability and fairness of the oral test procedure. As a result, a major focus of oral assessment research has been on the actual oral assessment *process*, i.e. on the structure of the discourse and the language produced by the candidate and interviewer/interlocutor in an attempt to

---

Email of corresponding author: [evdokia@otenet.gr](mailto:evdokia@otenet.gr)

ISSN: 1457-9863

Publisher: Centre for Applied Language Studies, University of Jyväskylä

© 2009: The authors

<http://apples.jyu.fi>

understand how the participants' verbal behaviour may affect the quality of language production and the assessment outcomes. (Brown & Hill 2007; Lazaraton 1996a; Meiron & Schick 2000). The study reported in this paper represents a systematic attempt to identify the nature and effect of interlocutor discourse practices on candidates' language production within the context of KPG oral exams in English. The study, conducted through the use of structured observation forms, aimed at investigating fairness of the oral test procedure and at providing invaluable feedback for the further development of the oral examiner training programme.

The role and linguistic behavior of the interlocutor during the oral exam has been highlighted by many researchers (Bachman, 1990; Bachman et al. 1995; Bonk & Ockey 2003; Chalhoud-Deville 1996; Lazaraton 1996b, 2002; McNamara 1996; Milanovic & Saville 1996; O'Sullivan 2000) as a major variable which can potentially affect candidate output and examiner rating.

More specifically, as regards the role of the examiner/interlocutor, many studies (Berwick and Ross 1996; Brown and Hill 2007; Merrylees and McDowell 2007; Orr 2002; O'Sullivan 2000) have shown interviewer or interlocutor (depending on the nature of the oral test) variation and inconsistency a) in terms of overall patterns of leniency/harshness in assigning test scores, b) in the way examiners interpret rating scales and assessment criteria, c) in the way they accommodate to the level of the candidates, d) in the support examiners provide candidates in order to facilitate comprehension and production, and e) in the elicitation of communicative performance of candidates (see also Bachman and Palmer 1996; Fulcher 2003; McNamara 1996). This inconsistency and variation is a major factor affecting the candidate's language output, thus threatening validity and reliability in oral tests.

Variation in oral examiner discourse practices has been identified in a number of studies. For example, Ross and Berwick (1992), in their investigation of features of accommodation (i.e. requests for clarification, propositional reformulation, grammatical simplification) employed by interviewers during the OPI (Oral Proficiency Interview), found that interviewers tended to over-accommodate to candidates at lower levels of proficiency. Their findings point to the fact that in cases where interlocutors are inclined to over-accommodate to all non-native speakers regardless of their need, or contrastingly, are unaware of the propositionally complex probes they produce, validity and reliability of final marks are put at risk. In a later study, Berwick and Ross (1996), examining differences in the discourse practices of native and non-native speaker oral examiners during the OPI, found differences in the strategies the two groups of interviewers used in assisting interviewees to understand problematic material. O'Sullivan (2000) investigated the discourse practices of female and male native speaker interviewers during an OPI and found that the female interviewers affected the nature of the interview by showing support in a more emphatic ways than their male colleagues. Candidates also achieved higher scores when interviewed by female native speakers. Lazaraton's and Saville's (1993) study of interviewer behaviour in CASE also identified many instances of interviewer accommodation and support including supplying vocabulary to candidates, reformulating questions, drawing conclusions for candidates etc. Within the context of the IELTS interview, Brown and Hill (2007) found that interviewers fell into two categories: the difficult interviewers who were less supportive to candidates, used more complex language and pushed candidates to use more complex interactive skills and the easier interviewers who adopted a more

“teacher-like” behaviour by being more supportive and providing feedback using simple language and more straightforward questions. Similar conclusions were also arrived at in Brown’s (2003) study.

Various types of interlocutor support and accommodation have been identified in studies (see Berwick and Ross 1996; Lazaraton 1996b) such as: a) use of comprehension checks and clarification requests by the interviewer, b) grammatical, syntactical or lexical simplification of an utterance to facilitate comprehension, c) introducing a topic in order to set the scene for the candidate, d) elaborating and rephrasing the candidate’s utterance/response, e) rephrasing or repeating a question to facilitate comprehension, f) supplying vocabulary, g) completing a candidate’s utterance or drawing conclusions for the candidate, h) use of exaggerated pronunciation, i) articulating words/utterances more slowly and j) providing positive feedback.

Moreover, variation has also been identified in interlocutor’s adherence to examiner guidelines with examiners diverging from the suggested format of the exam, rating procedures and tasks rubrics. As a result examiners elicit discourse other than the discourse the task was designed to elicit (Brown and Hill 2007; Merrylees and McDowell 2007; Ross 1992).

Variation in interlocutor’s discourses practices during an oral exam and the tendency to accommodate to the candidate’s level and provide support is not surprising given the interactive nature of the oral exam and the occupational background (teachers) of most oral examiners. Examiners may unconsciously slip into their teacher role and adopt teacher like behaviours such as accommodating, supporting and facilitating candidates’ language production. However, this variation and inconsistency in interlocutor’s discourse practices introduces variation into the speaking test as a whole (i.e. some candidates may receive more help than others or changes in task rubrics may result in a particular task not eliciting the kinds of language that it was designed to elicit). Variability in support also means that the whole test taking experience may be different for different candidates: for some the test will be more fair than for others. This variability may also affect test scores i.e. some examiners take into account the help offered to the candidate in the assignment of the final mark while others do not, thus enabling less competent candidates to get the same score as more able ones. In general, variation in the discourse practices of oral examiners has a significant effect on the validity of the oral test as a whole and reduces our confidence in the inferences we draw on the basis of the final mark (Fulcher 2003; Lazaraton 1996a).

An effective way to reduce inconsistency in examiner behaviour and rating practices and variability in their discourse practices is through the provision of systematic training opportunities (i.e. initial and follow up training) which focus on a) the application of assessment criteria and what they mean (Bachman & Palmer 1996; McNamara 1996; Orr 2002) and b) on the role of the interlocutor in the assessment process and the forms of behaviour that are acceptable in this context based on empirical findings (Lazaraton 1996a). Moreover, recently many examination systems (see Fulcher 2003; Lazaraton 1996a; Merrylees & McDowell 2007) have introduced interlocutor scripts or frames (which take the form of a list of set phrases and questions to be used by the examiner throughout the oral test) in order to reduce variation in the speech of interlocutors and standardise interlocutor discourse. Calls have also been made for the need to monitor the behaviour of examiners during the oral exam and after training through the use of evaluation templates in order to reduce variation in examiner discourse

practices and deviations from oral test procedural guidelines (see Lazaraton 1996a, for such a development with CASE examiners).

The suggested measures were taken in the context of the KPG oral exam in order to reduce variability in examiners' discourse practices and in interventions to task rubrics, identified on the basis of systematic observation of examiner behaviour during the exam.

This paper presents the rationale and main findings of the observation of oral examiner discourse practices during the KPG oral exams in English. The paper first introduces the KPG examination system in general (sections 2 and 3) and the oral exam in particular (section 4), and briefly describes the oral examiner training programme (section 5). The rationale of the observation project and the main findings concerning the nature and frequency of interlocutor interventions on candidate language output are presented in sections 6 and 7. The paper closes by discussing the developments which this project has brought about in the oral test procedure and the oral examiner training programme.

Given that the oral exam observation project is ongoing and large scale (from 2005 to 2008 almost 2000 oral examiners have been observed "on the job"), this paper will focus on the main findings of examiner discourse practices during one observation phase (May 2007) which had considerable implications for the development of the oral exam and the oral examiner training programme.

### **The context of the study: The KPG state certification system of language proficiency**

KPG (*Kratiko Pistopiitiko Glossomathias*), which is a state certification system of language proficiency, was implemented in Greece for the first time in April 2003.<sup>1</sup> The actual exams are designed and developed by groups of foreign language assessment specialists at the Universities of Athens and Thessaloniki, appointed by the Ministry of Education.

Candidates can take exams in English and German, Spanish and Turkish (developed by groups of experts at the University of Athens) and French and Italian (developed by groups of experts at the University of Thessaloniki). At present, KPG exams in English are designed for candidates of A1, A2, B1, B2 and C1 levels according to the scale set by the Common European Framework of References for Languages. Exams are administered twice a year (November and May) for all levels and languages at exam centres throughout Greece which also serve as exam centres for the national university entrance exams.

### **Defining characteristics of the KPG examination system**

The KPG exams in all languages currently offered are designed on the basis of common test specifications. The exams represent a "proficiency assessment" (rather than diagnostic or competences measurement) examination system which

---

<sup>1</sup> The KPG has recently been credited by the Greek state and it is recognized as a work qualification, it does not have commercial interests and it is subsidized by the state. The Ministry of National Education and Religious Affairs is the legal copyright owner of all documents containing information about the KPG assessment system and is responsible for the organization and administration of the exam.

aims to test candidate's ability to make socially purposeful use of the target language in Greece and abroad. The KPG exams aim to measure candidates' competence in comprehending and producing oral and written discourse as well as their ability to act as mediators across languages and their awareness of how the target language works to create socially purposeful meanings. Its global scale descriptors and language use descriptors relate to those of the Common European Framework of References for Languages.

Each exam regardless of level and language consists of 4 modules:

Module 1: Reading comprehension and language awareness

Module 2: Free writing production and mediation skills

Module 3: Listening comprehension

Module 4: Free speaking production and mediation skills.

As the system is being developed, detailed descriptions of test specifications and of all relevant procedures are published online (see <http://www.ypepth.gr/kpg> and [www.uoa.gr/english/rcel](http://www.uoa.gr/english/rcel)), making the system transparent on both national and international levels.

### The content, structure and procedure of the oral exam in English (B1, B2 and C1 level)

Below is a brief description of the content and structure of the oral exam for the three levels. The oral exams for B1 and B2 level are similar in structure and content; they differ in terms of task difficulty and the linguistic and cognitive complexity of Greek texts provided for mediation.

**Table 1.** The content and structure of the KPG oral exam

B1 & B2	C1
<b>Duration of test</b>	
15 minutes (B1) 20 minutes (B2)	20 minutes
<b>Pattern of participation</b>	
Candidates are tested in pairs but do not converse with each other	Candidates are tested in pairs and converse with each other
<b>Content of oral test</b>	
a) <b>ACTIVITY 1: Dialogue</b> (3-4 minutes) between examiner and each candidate who answers questions about him/herself and his/her environment posed by the examiner. b) <b>ACTIVITY 2: One-sided talk</b> (5-6 minutes) by each candidate who develops a topic on the basis of a visual prompt. c) <b>ACTIVITY 3: Mediation</b> by each candidate who develops a topic based on input from a Greek text. (6 minutes for both)	a) <b>Warm-up</b> (not assessed - 1 minute) Examiner asks each candidate a few ice-breaking questions (age, studies/work, hobbies) b) <b>ACTIVITY 1: Open-ended response</b> (4 minutes): The candidate responds to a single question posed by the examiner expressing and justifying his/her opinion about a particular issue/topic. c) <b>ACTIVITY 2: Mediation and open-ended conversation</b> (15 minutes): Candidates carry out a conversation in order to complete a task using input from a Greek text.

The KPG oral test procedure involves two examiners and two candidates in the examination room. One of the two examiners is the 'Interlocutor', i.e. the one who conducts the exam by asking the questions, assigning the tasks and by participating in the speech event). The other is the 'Rater', i.e. the one who sits aside silent, observing the examination and rating the candidate's performance. The Interlocutor also marks candidates' performance, but only after the candidates have left the room. Examiners alternate in their role as Interlocutor and Rater every three or four testing sessions.

Examiners are trained for their roles as Interlocutor and Rater through seminars which are offered throughout the year. However, on the day of the oral exam, examiners at their respective exam centres are given an examiner pack which contains guidelines regarding the exam procedure and oral examiner conduct and the oral test material: questions, tasks and rating criteria. These are handed to the examiners at least two hours before the exam begins along with the Candidate Booklet, which contains the prompts for the exam (photos and/or Greek texts, since one of the activities involves mediation).

### **The training of oral examiners**

The training of oral examiners started just before the first exam administration (October 2003) and has been systematic and on going ever since. Up to November 2008, 9 rounds of training seminars for oral examiners throughout Greece have taken place and more than 3000 examiners have been trained in assessing candidate's oral performance at A1/A2, B1, B2, C1 levels.

The ultimate goal of the oral examiner training programme is to develop a database of 3000 oral examiners who are fully and systematically trained to assess candidates' oral performance at all levels offered by the KPG exam battery and whose performance has been observed and evaluated.

Oral examiner training materials (handbooks and audio/ video simulations of oral exams) are developed by a group of experienced teacher educators from the Faculty of English Studies, University of Athens; this group is also involved in the design of the oral exam. The oral examiner training seminars are conducted by this group and a body of 50 multipliers throughout Greece. Multipliers are highly qualified associates (all have completed postgraduate studies), with extensive experience in teaching EFL and in teacher education. All multipliers have been practicing oral examiners or observers in order to ensure that they can draw on their personal experience to deal with the concerns of oral examiners.

### **The KPG oral examiner observation project**

The KPG observation project was launched in November 2005 as a pilot study in an attempt to identify whether and to what extent examiners adhere to exam guidelines and the suggested oral exam procedure (see Delieza 2008, Delieza in progress).<sup>2</sup> The overall goals of the observation project were to gain information about the efficiency of the oral exam administration and the efficiency of oral

---

<sup>2</sup> Since audio or video recording of actual examinations is prohibited by law in KPG exams, observation during actual examination was the sole solution for data collection.

examiner conduct, of the applicability of the oral assessment criteria and of inter rater reliability. This information was and is considered essential for further development and refinement of the oral exam and for the training and evaluation of oral examiners.

To date, six rounds of observation (observation phases) have been carried out. For each, a new, refined observation scheme is prepared on the basis of previous observation scheme findings.

As can be seen from Table 2, throughout the six observation phases 1948 oral examiners have been observed examining 6755 candidates.

**Table 2.** The KPG observation project in numbers

		Examiners	Candidates	
PHASE 1: November 2005 Levels B2 & C1	25 observers	B2	138	470
		C1	98	288
PHASE 2: May 2006 Levels B2 & C1	33 observers	B2	155	540
		C1	118	418
PHASE 3: May 2007 Levels B1, B2 & C1	32 observers	B1	35	132
		B2	156	588
		C1	105	342
PHASE 4: November 2007 Levels B1, B2 & C1	42 observers	B1	50	201
		B2	177	753
		C1	100	339
PHASE 5: May 2008 Levels A1-2, B1, B2 & C1	48 observers	A1-2	45	184
		B1	60	193
		B2	182	612
		C1	136	440
PHASE 6: November 2008 Levels A1-2, B1, B2 & C1	41 observers	A1-2	51	113
		B1	55	154
		B2	187	659
		C1	100	329

Multipliers (i.e. the trainers of oral examiners) who have been specially trained to use the observations forms, carry out observations in a representative sample of randomly selected examination centres throughout Greece. They are instructed to visit all examination rooms and observe all oral examiners at their assigned examination centre as long as the examination sessions last: from morning to afternoon. All observers enter the examination rooms as third parties and observe the examination process without interfering with it in any way. The observation form is completed before, while and after they have observed the process.

Each form is designed so that it is used in one examination session, i.e. with one examiner examining a pair of candidates. Observers are instructed to observe each examiner twice, i.e. with two pairs of candidates.

After each exam administration, observers send their completed observation forms to the Research Centre for English Language Teaching, Testing and Assessment (RCEL: [www.uoa.gr/english/rcel](http://www.uoa.gr/english/rcel)) of the University of Athens. Once collected, observation forms are processed and results are analysed. A report of the main findings is then produced. Findings are taken into account for

the design and development of the next round of oral examiner training seminars.

### *Categories of the observation forms*

Observation of the oral exam procedure is carried out, as mentioned above, with the use of specially designed, structured observation schemes, which focus on various aspects of the oral exam administration and content as well as on oral examiner conduct (see below for a detailed description).

In terms of format, the observation forms are clearly defined, structured checklists, based on specific categories and subcategories having to do with rules and guidelines pertaining to the test procedure and examiner conduct (see Appendix 1 for a sample observation form). They are in the form of tables, where the observers circle YES/NO or tick for the presence of a particular feature, but there is also some space in some of the questions where the observers can take notes if necessary.

The observation forms elicit information on the seating arrangements in the examination room, procedure followed, the candidates' age and sex, the choice of questions and tasks made by the examiner/interlocutor, ratings by the two examiners and the observer, duration of the oral exam and time allocation to different activities and to the two candidates, overall assessment of the examiner's oral performance as either Excellent, Very Good, Good, Mediocre, Poor. Apart from these common observation categories, from the third observation phase onwards the observation forms focused more specifically on the linguistic behaviour of the oral examiners, i.e. whether and to what extent examiners altered task rubrics and interfered with candidates' language output. This change of focus was due to the findings of the first two phases of the observation project, where observers had noted that almost 20% of the examiners had a tendency to change task rubrics and to interrupt candidates while speaking in order to correct or provide support. Observers' comments also enabled us to identify categories of "intervention". For the purposes of the observation project, intervention was defined as

*a discourse practice whereby the interlocutor-examiner chooses to make a linguistic change to or linguistically interfere with the given rubric and/ or interrupt the candidate or interfere with his/her language output.*

The types of examiner interventions (which are very similar to types of accommodation and support found in other studies see section 1 above) identified by observers in the first two phases of the observation project are listed in Table 3 below.



**Table 3.** Types of examiner intervention identified by observers.

Changes to or interference with the rubrics	Interruption of the candidate or interference with his/her language output in order to
a) Use of an introductory question. b) Change of one-two words from the rubric. c) Expansion of the original exam question. d) Explanation of the rubric without being asked to by the candidate / after being asked to. e) Repetition of the rubric (more slowly) without being asked to/ after being asked to. f) Supplying a synonym for a word without being asked to /after being asked to. g) Using examples which have not been provided	<ul style="list-style-type: none"> <li>• redirect the candidate because s/he misunderstood something,</li> <li>• help the candidate continue by repeating his/her last words,</li> <li>• make some kind of correction,</li> <li>• repeat the question or part of it,</li> <li>• supply one or more words the candidate was unable to find,</li> <li>• ask a seemingly irrelevant question,</li> <li>• add something</li> </ul>

The categories of intervention became part of the 2007 observation form. The frequency of examiner intervention also motivated us to examine why these interventions take place and whether they seem to have an effect on candidate language output. Thus, the revised observation form was designed to address the following research questions:

1. Do examiners change the task rubrics in any way?
2. If yes, what is the nature and the frequency of these changes?
3. Do examiners interfere with candidate language output?
4. If yes, what is the nature and frequency of this interference?
5. Why are these interventions made?
6. What seems to be the effect of these interventions on candidate language output?

The following section presents the results of the 2007 observation phase with particular focus on the nature and frequency of interventions made by examiners.

### **Findings regarding oral examiners' discourse practices - 2007 observation phase**

The tables that follow below present the frequency of occurrence of each category of intervention for each activity of each level. Before commenting on the percentages, we include an example of the way in which these frequencies were calculated. The following table concerns the first type of change to the rubrics in Activity 1, at B2 level (i.e. examiner uses an introductory question). The total of 588 represents the total of candidates observed, 102 represents the times this category was selected (ticked) by observers while 486 represents the times it was left blank. All types of intervention were calculated in the same way.

**Table 4.** The frequency of use of (an) introductory question(s) by the examiner

	Frequency	Percent
NOT TICKED	486	82,7
YES	102	17,3
Total	588	100,0

Before proceeding to the presentation of results, it should be pointed out that this observation phase involved 32 observers who observed 35 examiners examining 132 candidates in the B1 level oral test, 156 examiners who examined 588 candidates in the B2 level oral test and 105 examiners examining 342 candidates in the C1 level oral test. The results of the B1 and B2 level oral test will be presented together since both tests are similar in structure and the findings of examiner discourse practices for both tests were similar as well.

As can be seen from the Table below, examiners mostly intervened with task rubrics in Activity 1 of both levels of the oral test. In this activity (as mentioned in section 4) candidates are required to respond to 2-4 questions about themselves and their immediate environment (e.g. *Tell us a few things about your best friend* or *How would you like to spend your Christmas holidays?*). The frequency of occurrence of intervention with task rubrics in Activity 1 may well be due to the fact that examiners feel the need to put the candidates at ease and to create comfortable interaction at the beginning of the test (when candidates are most anxious). What is also interesting is that in Activity 1 examiners appear to favour expanding the question or simplifying it by using examples or by using an introductory question despite the fact that these questions are straightforward and less cognitively challenging than other questions in the oral test.

**Table 5.** Frequency of examiner intervention with task rubrics in the B1 and B2 level oral test

Type of change to or intervention with the rubric	B1			B2		
	Activity 1	Activity 2	Activity 3	Activity 1	Activity 2	Activity 3
Used an introductory question	16.7%	4.5%	3.8%	17.3%	2.0%	2.9%
Changed one-two words and/or supplied a synonym for a word	9.8%	7.6%	6.1%	9.0%	7.1%	11.2%
Expanded the question and/or used examples to explain it.	18.9%	22.0%	7.6%	16.8%	14.1%	10.4%
Other	11.4%	3.0%	4.5%	14.3%	8.0%	8.5%

The category “other” also appeared to have a fairly high frequency in Activity 1. Observers noted that for 11.4% of observed candidates in the B1 level test and 14.3% in the B2 level test, examiners used an entirely new question (i.e. questions not included in the examiner pack). The examiners may have followed this practice in an effort to facilitate candidates’ language production by asking questions the candidate was bound to be able to answer. These percentages,

although not high, are somewhat alarming since they put at risk the validity of the first activity.

In Activity 2, where the candidates are required to carry out a task on the basis of one or more photos given to them<sup>3</sup>, the expansion or explanation of the question occurred much more often than the other three types of interventions, in both level oral tests. The tasks in this activity usually consist of a series of sub-questions; examiners may adopt this strategy of intervention by breaking down the initial question into its sub-parts in order to facilitate candidates' comprehension and to ensure that candidates respond to all parts of the task. In Activity 3 - where the candidates are required to carry out a task on the basis of a Greek text<sup>4</sup>- interventions to task rubrics were more frequently made in the B2 level test. Given the original nature of this task (which involves relaying information from Greek to English) and candidates' lack of appropriate test taking strategies for this task, examiners may opt to expand the question or simplify it in order to prompt candidates to continue speaking.

Concerning interruptions to candidates' language output (see Table 6) these occurred almost with the same frequency with B1 and B2 level candidates. Moreover, interruptions occurred slightly more frequently in Activities 2 and 3 for both levels; this may be due to the fact that these activities tend to be more cognitively challenging and require more complex language output than Activity 1. At B1 level, in Activities 1 and 3, the examiners mostly interfered by repeating the candidates' last words to help them continue. In B2, Activity 1, examiners most frequently redirected the candidate by repeating the question or part of it. Both courses of action are considered by the test development team as an acceptable way of dealing with the difficulty of the candidates to continue talking, since they do not provide them with the linguistic means to continue.

In Activity 2, as well as in Activity 3 in both levels, the 'other' category also has a fairly high percentage. According to observers, this percentage mostly refers to examiners interrupting the candidates by asking more questions in order to help the candidate to continue talking. Examiners may adopt this practice since tasks often appear a bit complicated and examiners feel they need to expand on them. However, given that we do not know what the nature and content of these "additional" questions is, we are concerned about the effect of these questions on the validity of Activities 2 and 3.

---

<sup>3</sup> For instance, the candidate is shown a picture of a 'Street Scene' (May 2007 - English KPG - B2 - Module 4, Candidate Booklet, page 4) and the relevant task is '*Look at photo 7 (or 8 or 9). Try to guess who these people are and where they are, what they are doing and what you think they are going to do next.*'

<sup>4</sup> For instance, the candidate is given a text in Greek with advice on surfing the Internet safely (May 2007 - English KPG - B2 - Module 4, Candidate Booklet, page 10) and the relevant task is '*Imagine I am your Italian friend who has just become an Internet user. Using information from Text 2, give me advice on what I should do so as to surf the net safely.*'

**Table 6.** Frequency of examiners' intervention with candidate language output in the B1 and B2 level oral test.

The examiner interrupted the candidate or intervened with their language output in order to...	B1			B2		
	Activity 1	Activity 2	Activity 3	Activity 1	Activity 2	Activity 3
redirect the candidate because s/he misunderstood something by repeating the question or part of it.	2.3%	5.3%	6.1%	7.1%	10.0%	7.0%
help the candidate continue by repeating his/her last words.	6.1%	3.8%	12.1%	3.4%	4.9%	5.8%
make some kind of correction or add something (e.g. supply one or more words the candidate was unable to find).	3.8%	7.6%	5.3%	3.4%	5.3%	5.8%
other	5.3%	22.0%	10.6%	6.6%	13.4%	10.0%

The findings of examiner intervention in the C1 level oral test are presented below. It should be pointed out that the C1 level test consists of two activities (see section 4): Activity 1 requires candidates to respond to a single question providing their opinion on a topic/issue<sup>5</sup>, while Activity 2 is a mediation problem solving task where candidates are provided with two similar texts (or two different parts of the same text) in Greek and they are required to read their respective texts and to relay in English relevant information from their text in order to solve a problem or reach a consensus<sup>6</sup>. What becomes immediately apparent when comparing the results of examiner interference with task rubrics in the B1/B2 level oral test with the results of the C1 level oral test, is that examiners generally appear to intervene less with task rubrics in the C1 level oral test. This tendency may be due to the higher language level of the candidates, i.e. examiners may feel that candidates at this level do not require as much support and prompting as do candidates of lower levels.

As can be seen from Table 7 below, in Activity 1 examiners mostly intervened with task rubrics by expanding the question or providing examples in order to simplify it. Observers also noted that in 7.3% of the observed cases (see "other" category) examiners embellished the original question with additional questions in order to prompt candidates to produce more language output. Once again, this practice raises doubts about the validity of Activity 1. It is interesting that, unlike in the B1 and B2 level oral test, examiners do not use introductory questions although the opinion question by nature is more complex

<sup>5</sup> For instance, 'Do you agree or disagree with the statement that technology has brought people closer?' or 'Many people believe that violence on TV has negative effects, especially on children. What do you think?', (May 2007 – English KPG C1 – Module 4, Examiner Pack, page 8)

<sup>6</sup> For instance, the candidates are given two different but related texts both dealing with different tourist sites in Madrid (May 2007 – English KPG C1 – Module 4, Candidate Booklet, pages 5 and 12) and the relevant task is 'Imagine you and your partner are planning to visit Madrid for the weekend. Exchange information from your texts and together decide which places you will and which you will not see.'

linguistically and cognitively and therefore could be supported by such an accommodation strategy. Presumably, examiners may feel that since the level is higher, candidates do not need such support at this level.

In Activity 2, which is the most cognitively and linguistically challenging one in the full suite of oral tests offered by the KPG, observers noted that in almost 10% of the observed cases examiners repeated and rephrased the original task rubric - a practice which is often necessary since candidates tend to forget the task or part of it because they have to read a Greek text before actually performing the task. Observers also noted that in slightly fewer cases (8.2%) examiners changed a few words of the original rubric or provided a synonym for a word in order to ensure comprehension of the task rubric.

**Table 7:** Frequency of examiner intervention with task rubrics in the C1 level oral test.

C1		
Type of change to or interference with the rubric	Activity 1	Activity 2
used an introductory question	4.7%	0.6%
changed one-two words and/ or supplied a synonym for a word	2.3%	8.2%
expanded the question and/ or used examples to explain it.	8.2%	4.7%
other	7.3%	9.9%

Regarding interruptions to candidate language output in the C1 level oral test, the second activity presented some categories not detectable in the first activity since the second activity is a task carried out collaboratively by two candidates with the examiner remaining a mere listener and observing the candidates' conversation. In Activity 1 (opinion question), the 'other' category presented by far the highest occurrence. Observers noted that in 21.6% of the cases, examiners added more questions to the original task presumably in order to prompt the candidates to continue speaking. In Activity 2, examiners mostly interrupted candidates in order to remind them to use information from their texts when interacting (observed in 16.4% of the cases) This type of interference is recommended in training seminars since there are often several small texts given to candidates and they often choose to relay information from only one or two neglecting to mediate information from the rest. Observers also noted that examiners interrupted candidates in order to remind them to reach a decision or draw a conclusion (which is required by Activity 2 tasks). Thus, the types of interference with candidate language output observed in Activity 2 of the C1 level oral test mainly aimed at reminding candidates of the activity requirements rather than at facilitating candidates' language production.

**Table 8.** Frequency of examiners' intervention with candidate language output in the C1 level oral test.

C1		
The examiner interrupted the candidate or interfered with their language output in order to...	Activity 1	Activity 2
Redirect the candidate because s/he misunderstood something by repeating the question or part of it.	7.6%	10.5%
Help the candidate continue by repeating his/her last words.	5.8%	-
Make some kind of correction or add something (e.g. supply one or more words the candidate was unable to find.	1.2%	2.9%
Remind the candidates of (part of) the task / their goal	-	2.3%
Remind them they have to interact	-	4.7%
Remind them to refer to all the texts	-	16.4%
Participate in the discussion? (Why?/How? Please specify.)	-	4.1%
Other	21.6%	14.6%

The aggregate results for all types of intervention for all activities of all level tests are presented in Table 9 below.

**Table 9.** Frequency of examiner intervention in the B1, B2 and C1 level oral tests.

Intervention	B1	B2	C1
Changes to the rubrics	47.0%	59.2%	31.5%
Interruptions and/or interventions	49.2%	47.1%	60.2%

The aggregate results presented in Table 9 clearly show that intervention occurred with almost the same frequency in all level oral tests, with the B1 and B2 level test being most "prone" to examiner intervention. In the B1 and B2 level oral test examiners were fairly consistent in their tendency to change rubrics (or add questions to the original task rubric) and interfere with candidates' language output. As reported in Tables 5 and 6 above, these interventions aimed at facilitating candidate language production. In the C1 level oral test, examiners' intervention mainly took the form of interrupting candidates language output (60.2%). As mentioned above though, examiner intervention with candidate language output was mainly done with the purpose of reminding candidates of test procedures and requirements. Thus it appears that examiners tend to vary their intervention practices according to the level of the candidates; the lower the level, the more they feel the need to support, facilitate and accommodate to candidates' level.

A final research question addressed through this observation phase related to the observers' perceptions of whether the examiners' interventions facilitated or obstructed candidate language production. Undoubtedly, analysis of transcriptions of oral exams would have constituted a much more valid approach to addressing this issue. However, given that by law no form of recording of the oral exam is permitted in the examination room, it was necessary to rely on our trained observers' perceptions of the effects of examiner intervention. The results presented below thus represent a value judgment on the part of the observers and cannot be empirically verified in actual exam conditions. The results therefore should be approached and interpreted with caution.

The table below presents the frequencies with which observers reported that examiner interventions influenced candidate language output (column “YES”) and the frequencies with which observers reported that examiner interventions facilitated (column “easier”) or obstructed (column “difficult”) candidate language output.

**Table 10.** Observers’ perceptions of the effect of examiner interventions on candidate language output

The interlocutor’s intervention (change of rubric/ interruption or interference) influenced the candidate’s language output - made it <u>easier</u> or <u>more difficult</u> for the candidate to answer.									
	Activity 1			Activity 2			Activity 3		
	Yes	Easier	More difficult	Yes	Easier	More difficult	Yes	Easier	More difficult
<b>B 1</b>	34.8	92.0	8.0	38.4	92.0	8.0	26.7	83.3	16.7
<b>B 2</b>	29.9	92.3	7.7	33.0	95.2	4.8	30.2	92.2	7.8
<b>C 1</b>	26.7	97.7	2.3	30.4	90.2	9.8			

In Activity 1, it appears that the lower the level is, the more the influence of examiner intervention on candidate language output becomes detectable. In most cases, it appears that interventions facilitate candidates’ language production rather than make it more difficult. Of course, any such effect on candidates’ performance threatens the validity of the test items and the reliability of the final mark.

Especially in Activity 1, the higher the level of the candidate, the more positive the effects of intervention seem to be on candidates’ language output.

In Activity 2 though, the positive influence of examiner interventions seem to decrease with the level; although they facilitate the candidates’ language production more in B2 than in B1, in the C1 level exam the percentage falls below that of B1, probably because the nature of Activity 2 in the C1 level exam is quite different and not at all comparable with the respective activity in the lower level exams. In other words, since candidates perform collaboratively the mediation task at C1 level, they do not seem to benefit as much from examiner interventions.

Finally, in Activity 3, once more the effect on the candidates’ language output is higher in the B2 level exam than in B1. B2 candidates seem to benefit more from examiner intervention than B1 candidates. As a general remark, it appears that the more advanced the level of candidates is, the more they benefit from the examiners’ interventions, probably because they have developed strategies for interaction and test wiseness strategies, which allow them to use all the support they get from their interlocutor in order to give a complete answer or to continue speaking.

## Discussion of results

The findings of the third observation phase proved valuable in many respects. Firstly, they verified our concerns about the frequency of examiner interventions and their potential effects on the validity, reliability and fairness of the test as a whole. Secondly, they highlighted the need to introduce changes in our oral examiner training programme with a view to limit such examiner intervention. Thirdly, from a research perspective the findings were particularly revealing since they made clear that examiners do not choose to intervene on an ad hoc basis but tend to adjust their strategies for changing task rubrics or intervening with candidate language output depending on the level of the candidates and the stage of the exam.

More specifically, the findings reveal that examiners most frequently choose to change task rubrics (by using an introductory question, adding their own questions or expanding the original question with added information) in the first activity of the lower level exams. These strategies are probably chosen by examiners in order to reduce candidates' anxiety and to help them produce more language output so that they feel some sense of achievement from the beginning of the exam. In the other two activities of the B1 and B2 level exams, examiners tended to intervene with task rubrics less frequently; interventions mainly took the form of expanding the original task rubric or simplifying it through the use of examples in order to facilitate candidates' comprehension of task requirements and to ensure that candidates respond to all parts of the task.

In the higher level oral test (C1), examiners chose to alter task rubrics more sporadically than in lower levels. Examiners' preferred strategy of rubric alteration was the expansion of the original question or its simplification (through the use of examples) in the first activity. In Activity 2, the most linguistically and cognitively challenging speaking task, examiners mainly reformulated rubrics in order to ensure task comprehension and to remind candidates of the requirements of the task. It appears that examiners chose to adjust the frequency and nature of support to candidates' language proficiency level and the stage of the exam. The higher the level of candidates, the less support they are offered (see also Berwick and Hill 2007; Ross and Berwick 1992 for similar findings). Additionally, examiners tend to intervene with task rubrics more frequently in the initial activity of all exam levels presumably to help put candidates at ease and boost their confidence.

As far as examiners' intervention with candidate language output is concerned, it appears that in lower level exams (B1 and B2), examiners tend to intervene more frequently with candidate language output in activities which are considered linguistically more complex with the intention of facilitating production. Interruptions of the language output of candidates at higher levels (C1) in linguistically complex activities do not seem to be made with the purpose of facilitating language production. More specifically, in the B1 and B2 level exam (especially in Activities 2 and 3) examiners mainly interrupted candidates by adding more questions to the original task rubric in order to help them continue speaking. In the second activity of the C1 level exam, interruptions were mostly made with a view to reminding candidates of exam procedures or task requirements.

Observers' reports of the effect of examiner interventions on candidate language output revealed that these interventions in the vast majority of cases seemed to facilitate candidate language production; the higher the level of the



candidates, the more facilitating the effect of examiner intervention seemed to be.

However interesting these findings are from a research perspective, they are nevertheless quite alarming considering their potential effect on the validity and reliability of the exam. Moreover, they raise a number of questions and doubts concerning exam fairness: Do all candidates of the same level receive the same kind of support? Is the frequency and nature of support provided to candidates taken into account in the assignment of the final mark? These concerns necessitated a number of decisions for measures to reduce examiner intervention, which are described below.

## **Developments in the training of oral examiners and monitoring of their performance as a result of the observation findings**

### *The introduction of an interlocutor frame*

There were three major developments deriving from the analysis of the data obtained through the first three phases of the observation project. First of all, it was decided that despite multipliers' insistence in training seminars to raise examiners' awareness of the negative repercussions of changing task rubrics and interfering with candidates' language output throughout the exam, the introduction of an Interlocutor Frame in the oral test should be made in order to tackle the problem of examiner discourse variation more effectively. The use of interlocutor frames has been adopted by the Cambridge Local Examinations Syndicate for most of its speaking tests in an effort to control and reduce the difference between the discourse practices of interlocutors. The consistent results of all our observations led us to the decision to introduce an interlocutor frame, incorporate it in the training handbooks and raise examiners' awareness of its use and purpose during the seminars. The use of the interlocutor frame and adherence to it became a focus of the May 2008 observation phase. Its use by oral examiners and its effectiveness will be monitored in future observation phases.

### *Development of a list of suggested examiner discourse practices*

Secondly, a detailed list of suggested examiner discourse practices was compiled and included in the seminar training materials. Acceptable and non-acceptable examiner interventions were spelt out and have become another focus of our training seminars in an effort to raise examiners' awareness of the harmful effects of their interventions on test validity and fairness. The tables below present the types of acceptable and non-acceptable examiner discourse practices.

**Table 11.** List of acceptable examiner interventions☉ **ACCEPTABLE INTERVENTION**

<p><b>It is acceptable to take action or intervene in order to:</b></p> <ul style="list-style-type: none"> <li>• repeat the rubric (more slowly) [if asked or if considered necessary]</li> <li>• repeat part of the rubric or the rubric in parts (e.g. to remind the candidate(s) of something they have forgotten to answer) [if asked or if considered necessary]</li> <li>• supply a synonym for a word after being asked to do so</li> <li>• remind the candidates of (part of) the task / their goal [using the wording of the rubric]</li> <li>• remind the candidate to use/ relay information from their text(s)</li> <li>• help the candidate continue by repeating his/her last words</li> <li>• give the candidates the opportunity to produce more output by saying 'is there anything else you want/ would like to add?' or simply 'anything else' or even 'and'?</li> <li>• use fillers, such as 'aha', 'uhm' etc to show that they are following the candidate</li> <li>• direct the candidates to a picture they have not used (<b>only in B1 &amp; B2</b>)</li> <li>• remind the candidates that they have to interact (<b>only in C1</b>)</li> <li>• remind the candidates to refer to all the texts (<b>only in C1</b>)</li> </ul>
---

**Table 12.** List of non-acceptable examiner interventions☉ **NON - ACCEPTABLE INTERVENTION**

<p><b>The following ways of changing, or interfering with the rubrics are non-acceptable :</b></p> <ul style="list-style-type: none"> <li>• changing one-two words, adding words to or expand the rubric</li> <li>• supplying a synonym for a word without being asked to</li> <li>• paraphrasing or using examples to explain the rubric</li> <li>• directing the candidates' answer by using an introductory question or a leading question</li> <li>• using their own questions</li> </ul>	<p><b>It is non-acceptable to interrupt the candidate(s) or interfere with his/her/their language output by:</b></p> <ul style="list-style-type: none"> <li>• making some kind of correction</li> <li>• supplying one or more words the candidate is unable to find</li> <li>• asking a seemingly irrelevant question</li> <li>• making suggestions, giving options or examples or by adding something</li> <li>• explaining a word by paraphrasing it and giving examples</li> <li>• making personal comments or providing personal information</li> <li>• finishing the phrase for the candidate</li> <li>• participating in the discussion (<b>only in C1</b>)</li> </ul>
---	--

Within the training seminars, examiners are provided with specific examples of acceptable and non-acceptable interventions and are then asked to identify instances of such interventions and their ensuing effects in the video recorded oral exam simulations which they view during the seminar.

*Development of an oral examiner evaluation form*

Finally, the observation findings and comments on examiner behaviour by observers enabled us to arrive at and develop a specific set of criteria on the basis of which examiners can be evaluated 'on the job'; the criteria were tried out (on a trial basis) in May 2008. The evaluation of KPG oral examiners' performance forms one of the aims of the oral examiner training programme and a prerequisite for our examiners to be included in the final database (see section 5). The evaluation criteria include: *Attitude towards candidates, Body language,*

*Pronunciation/ intonation, Linguistic competence, Ability to follow test procedure, Ability to handle trouble situations.* Each criterion is assessed on a 5 point scale. The criteria for oral examiner evaluation can be found in Appendix 2.

## Conclusion

The KPG observation project has produced valuable findings which complement related research on interviewer/interlocutor variation and its effect on oral test validity and reliability. Despite the inherent complexities of setting up and organizing such a large scale project, the whole procedure has provided a wealth of qualitative and quantitative data which has benefited the KPG oral exam in three main ways.

First and foremost, it has pointed to the factors which threaten the validity and reliability of the oral test by throwing light on categories of variation in the oral examination procedure. Secondly, observation has provided insights into the ways in which the oral test can be improved by more clearly defining the role of the oral examiners as interlocutors. Finally, it has provided the core ideas for the design of more focused training materials. These major contributions clearly prove the usefulness of observation as a method to evaluate and investigate the oral exam procedure, especially since recording of any type has been impossible.

## References

- Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F., Lynch, B. K., & Mason, M. 1995. Investigating variability in tasks and rater judgements in a performance test of foreign language speaking. *Language Testing* 12, 239-258.
- Bachman, L. F. & Palmer, A. S. 1996. *Language Testing in Practice*. Oxford: Oxford University Press.
- Berwick, R. & Ross, S. 1996. Cross-cultural pragmatics in oral proficiency interview strategies. In M. Milanovic and N. Saville (eds.) *Performance Testing, Cognition and Assessment, Studies in Language Testing 3, Selected Papers from the 15<sup>th</sup> Language Testing Research Colloquium, Cambridge and Arnheim*, Cambridge: UCLES/Cambridge University Press, 34-54.
- Bonk, W.J. & Ockey, G. 2003 A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing* 20 (1), 89-110.
- Brown, A. 2003. Interviewer variation and the co-construction of speaking proficiency. *Language Testing* 20 (1),1-25.
- Brown, A. & Hill, K. 2007. Interviewer style and candidate performance in the IELTS oral interview. In L. Taylor and P. Falvey (eds.) *Research in Speaking and Writing Assessment Studies in Language Testing 19, IELTS Collected Papers*, Cambridge: UCLES/Cambridge University Press, 37-62
- Chalhoub-Deville, M. 1996. Performance assessment and the components of the oral construct across different tests and rater groups. In M. Milanovic and N. Saville (eds.) *Performance Testing, Cognition and Assessment, Studies in Language Testing 3, Selected Papers from the 15<sup>th</sup> Language Testing Research Colloquium, Cambridge and Arnheim*, Cambridge: UCLES/Cambridge University Press, 55-73
- Delieza, X. 2008. Monitoring KPG examiner conduct. *Directions* Issue 1. RCEL publications, University of Athens

- Delieza, X. (in progress). *Interlocutor Performance Variability in Language Proficiency Testing: the case of the Greek State Certificate Examination*. PhD Dissertation in progress. Faculty of English Studies, University of Athens
- Fulcher, G. 2003. *Testing Second Language Speaking*. Harlow: Pearson Education Limited.
- Lazaraton, A. 1996a. Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing* 13 (2), 151-172.
- Lazaraton, A. 1996b. A qualitative approach to monitoring examiner conduct in CASE. In M. Milanovic and N. Saville (eds.) *Performance Testing, Cognition and Assessment, Studies in Language Testing 3*, Selected Papers from the 15<sup>th</sup> Language Testing Research Colloquium, Cambridge and Arnheim, Cambridge: UCLES/Cambridge University Press, 18-33.
- Lazaraton, A. 2002. *A Qualitative Approach to the Validation of Oral Language Tests. Studies in Language Testing 14*. Cambridge: Cambridge University Press.
- Lazaraton, A. & Saville, N. 1993. *Processes and outcomes in oral assessment*, Paper presented at the 16<sup>th</sup> Language Testing Research Colloquium, Washington DC.
- MacNamara, T. F. 1996. *Measuring Second Language Performance*. London: Longman.
- Merrylees, B. & McDowell, C. 2007. A survey of examiner attitudes and behavior in the IELTS oral interview. In L. Taylor and P. Falvey (eds.) *Research in Speaking and Writing Assessment Studies in Language Testing 19, IELTS Collected Papers*, Cambridge: UCLES/Cambridge University Press, 142-182
- Meiron, B. & Schick, L. 2000. Ratings, raters and test performance: An exploratory study. In A.J. Kunnan (ed.) *Fairness and Validation in Language Assessment, Studies in Language Testing 9, Selected Papers from the 19<sup>th</sup> Language Testing Research Colloquium, Orlando, Florida*, Cambridge: UCLES/Cambridge University Press, 153-176.
- Merrylees, B. & McDowell, C. 2007. A survey of examiner attitudes and behaviour in the IELTS oral interview. In M. Milanovic and C. Weir (eds.) *Research in Speaking and Writing Assessment Studies in Language Testing 19, IELTS Collected Papers*, Cambridge: UCLES/Cambridge University Press, 142-183.
- Milanovic, M., & Saville, N. 1996. *Performance Testing, Cognition and Assessment: Selected Papers from the 15<sup>th</sup> Language Testing Research Colloquium*. Cambridge: Cambridge University Press.
- Orr, M. 2002. The FCE speaking test: using rater report to help interpret test scores. *System* 30, 143-154.
- O'Sullivan, B. 2000. Exploring gender and oral proficiency interview performance. *System* 28, 373-386.
- Ross, S. 1992. Accommodative questions in oral proficiency interviews. *Language Testing* 9 (2), 173-186.
- Ross, S. & Berwick, R. 1992. The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition* 14, 159-176.

## Appendix 1: Sample observation form

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ  
ΤΜΗΜΑ ΑΓΓΛΙΚΗΣ ΓΛΩΣΣΑΣ ΚΑΙ  
ΦΙΛΟΛΟΓΙΑΣ  
Επιστ. Υπεύθυνη Αγγλικής: Καθηγήτρια Β.  
Δενδρινού

NATIONAL AND ΚΑΠΟΔΙΣΤΡΙΑΝ  
UNIVERSITY OF ATHENS  
FACULTY OF ENGLISH STUDIES  
English Project Director: Professor B.  
Dendrinou



THE GREEK STATE EXAMINATION SYSTEM TO CERTIFY FOREIGN  
LANGUAGE PROFICIENCY

May 2007

<b>THE B2 LEVEL ORAL TEST OBSERVATION FORM - INTRODUCTORY PART*</b>
---

<b>The setting for the Oral Test:</b> Have the examiners set the desks up properly?		
Are the candidates sitting side by side?		
Is the Examiner-Interlocutor at some distance from the candidates but sitting so that s/he's facing them?		
Does the desk of the Examiner-Assessor allow visual contact with the candidates and the Examiner-Interlocutor?		

\*Fill in this table ONCE, i.e. every time you enter a new examination room.

<b>THE B2 LEVEL ORAL TEST OBSERVATION FORM - THE EXAMINATION</b>
--

	<b>Name &amp; surname (in GREEK)</b>	<b>Code</b>
<b>Interlocutor-Examiner</b>		

<b>Candidate A</b>		<b>Candidate B</b>	
Approximate age		Approximate age	
Sex (Circle.)	MALE / FEMALE	Sex (Circle.)	MALE / FEMALE

<b>PART 1: ACTIVITIES</b>
---------------------------

<b>ACTIVITY 1 - DIALOGUE</b>
------------------------------

Questions no:

Questions no:

Candidate A		Candidate B		Candidate B
YES	NO	1. Are the questions asked by the examiner from <i>different</i> categories? (Circle.)	YES	NO
YES	NO	2. Did s/he take into account the candidate's profile when s/he chose the questions? (Circle.)	YES	NO
YES	NO	3. Did s/he change or interfere with the rubrics in any way? (Circle.) (If NO, move to question 4.)	YES	NO
(TICK ✓)		If your answer is YES, TICK (✓) below to indicate in what way s/he did so.	(TICK ✓)	
		a. S/he used an introductory question.		
		b. S/he changed one-two words and/ or supplied a synonym for a word.		
		c. S/he expanded the question and/ or used examples to explain it.		
		d. other (Please specify.) _____		
(TICK ✓)		4. Did s/he interrupt the candidate or interfere with his/her language output in order to...	(TICK ✓)	
		a. redirect the candidate because s/he misunderstood something by repeating the question or part of it?		
		b. help the candidate continue by repeating his/her last words?		
		c. make some kind of correction or add something (e.g. supply one or more words the candidate was unable to find)?		

		d. other?(Please specify.) _____		
YES	NO	5. Did the interlocutor's intervention (change of rubric/ interruption or interference) influence the candidate's language output in any way? If you circle <u>YES</u> , please also <u>circle</u> if s/he made things easier or more difficult for the candidate and provide any useful comments: _____	YES	NO
Easier / More difficult			Easier / More difficult	
YES	NO	6. Did the examiner ask as many questions as were necessary to assess the candidate in this activity?	YES	NO
YES	NO	7. Was each candidate given enough time to complete the task?	YES	NO
<b>ACTIVITY 2 - ONE-SIDED TALK</b>				
Photo(s): <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> - Task <input type="checkbox"/>			Photo(s): <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> - Task <input type="checkbox"/>	
<b>Candidate A</b>			<b>Candidate B</b>	
		1. Did the examiner take into account the candidate's profile when s/he chose		
YES	NO	a. the photo(s)? (Circle.)	YES	NO
YES	NO	b. the task? (Circle.)	YES	NO
YES	NO	2. Did s/he change or interfere with the rubrics in any way? (Circle.) (If NO, move to question 3.)	YES	NO
(TICK ✓)		If your answer is YES, TICK (✓) below to indicate in what way s/he did so.	(TICK ✓)	
		a. S/he used an introductory question.		
		b. S/he changed one-two words and/ or supplied a synonym for a word.		
		c. S/he expanded the question and/ or used examples to explain it.		
		d. other (Please specify.) _____		
(TICK ✓)		3. Did s/he interrupt the candidate or interfere with his/her language output in order to...	(TICK ✓)	
		a. redirect the candidate because s/he misunderstood something by repeating the question or part of it?		
		b. help the candidate continue by repeating his/her last words?		
		c. make some kind of correction or add something (e.g. supply one or more words the candidate was unable to find)?		
		d. other?(Please specify.) _____		
YES	NO	4. Did the interlocutor's intervention (change of rubric/ interruption or interference) influence the candidate's language output in any way? If you circle <u>YES</u> , please also <u>circle</u> if s/he made things easier or more difficult for the candidate and provide any useful comments: _____	YES	NO
Easier / More difficult			Easier / More difficult	
YES	NO	5. Was each candidate given enough time to complete the task?	YES	NO

ACTIVITY 3 - MEDIATION					
Text: <input type="checkbox"/> <input type="checkbox"/> - Task <input type="checkbox"/>			Text: <input type="checkbox"/> <input type="checkbox"/> - Task <input type="checkbox"/>		
Candidate A				Candidate B	
		1. Did s/he take into account the candidate's age and other relevant factors when s/he chose			
YES	NO	a. the text(s)? (Circle.)		YES	NO
YES	NO	b. the task? (Circle.)		YES	NO
YES	NO	2. Did s/he change or interfere with the rubrics in any way? (Circle.) (If NO, move to question 3.)		YES	NO
(TICK ✓)		If your answer is YES, TICK (✓) below to indicate in what way s/he did so.		(TICK ✓)	
		a. S/he used an introductory question.			
		b. S/he changed one-two words and/ or supplied a synonym for a word.			
		c. S/he expanded the question and/ or used examples to explain it.			
		d. other (Please specify.) _____			
(TICK ✓)		3. Did s/he interrupt the candidate or interfere with his/her language output in order to...		(TICK ✓)	
		a. redirect the candidate because s/he misunderstood something by repeating the question or part of it?			
		b. help the candidate continue by repeating his/her last words?			
		c. make some kind of correction or add something (e.g. supply one or more words the candidate was unable to find?)			
		d. other?(Please specify.) _____			
YES	NO	4. Did the interlocutor's intervention (change of rubric/ interruption or interference) influence the candidate's language output in any way? If you circle <u>YES</u> , please also <u>circle</u> if s/he made things easier or more difficult for the candidate and provide any useful comments:		YES	NO
Easier / More difficult		_____		Easier / More difficult	
YES	NO	5. Was each candidate given enough time to complete the task?		YES	NO

PART 2: COMMUNICATIVE COMPETENCE - BEHAVIOUR - BODY LANGUAGE					
Candidate A				Candidate B	
(TICK ✓)		1. Did the Examiner use the appropriate communicative strategies?		(TICK ✓)	
		a. S/he was polite, friendly and welcoming making the candidates feel comfortable.			
		b. S/he was too supportive.			
		c. S/he lacked eye-contact and/ or didn't use appropriate body language.			
		d. S/he appeared strict/ distant/ stiff/ shy/ indifferent. (Circle once or more and then tick.)			
		e. other (Please specify.) _____			
YES	NO	2. Do you think that the interlocutor's communicative competence/ behaviour/ body language influenced the candidate's language output in any way? If you circle <u>YES</u> , please also <u>circle</u> if s/he made things easier or more difficult for the candidate and provide any useful comments:		YES	NO
Easier / More difficult		_____		Easier / More difficult	
3. How would you rate the Interlocutor-Examiner's communicative competence as Interlocutor?					
<input type="checkbox"/> Excellent <input type="checkbox"/> Very good <input type="checkbox"/> Good <input type="checkbox"/> Mediocre <input type="checkbox"/> Poor					





**Appendix 2:** Criteria for the evaluation of oral examiner performance**DESCRIPTORS FOR THE EVALUATION OF ORAL EXAMINER PERFORMANCE (for Observation Form B)**

	<b>Excellent = 5</b>	<b>Very good = 4</b>	<b>Good = 3</b>	<b>Mediocre = 2</b>	<b>Poor = 1</b>
<b>Attitude towards candidates</b>	Is naturally polite, friendly and welcoming enough creating thus a comfortable atmosphere for the candidate.	Is polite and quite friendly or too friendly for the speech event and makes the candidates feel quite comfortable.	Is polite but strict and distant and not very welcoming and/or doesn't help candidates feel comfortable.	Is quite polite but stiff and/or shy and/or tense and makes the candidates feel tense.	Is very distant, strict, and almost indifferent and makes candidates feel quite or very tense.
<b>Body language</b>	His/her body language and manner is that of an interlocutor who is leading and controlling the speech event, keeping a safe distance but at the same time creating a positive environment for language production. Maintains eye-contact, naturally consulting his/ her notes and follows with head nodding, smiling or fillers (e.g. aha, uhum etc.) the candidate's speech.	His/her body language and manner is that of an interlocutor who seems to be leading and controlling the speech event, keeping some distance but at the same time creating an appropriate environment for language production. Maintains, for the most part, eye-contact, consulting his/ her notes and often follows with head nodding, smiling or fillers (e.g. aha, uhum etc.) the candidate's speech.	His/her body language and manner is that of an interlocutor who is leading and controlling the speech event in a rather authoritative way, is rather distant, thus creating a rather uncomfortable environment for language production. Maintains some eye-contact, often consulting his/ her notes and only sometimes follows with head nodding or fillers (e.g. aha, uhum etc.) the candidate's speech.	His/her body language and manner is that of an interlocutor who is leading and controlling the speech event in a rather mechanistic way and who is distant, thus creating a rather negative environment for language production. Has little eye-contact, invariably looking at and reading from his/ her notes and does not often show with head nodding or fillers (e.g. aha, uhum etc.) that she/he is following the candidate's speech.	His/her body language and manner is that of an interlocutor who is leading and controlling the speech event in an indifferent way, keeping a distance which is demotivating for the candidate, thus creating a very negative environment for language production. Seldom maintains eye contact with the candidate and reads from his/her notes, and does not show with head nodding or fillers (e.g. aha, uhum etc.) that she/he is following the candidate's speech.

<b>Pronunciation / intonation</b>	Speaks with a clear, natural (if not native-like) and fully intelligible pronunciation. Uses appropriate stress and intonation to explain the procedure and reads rubrics, highlighting words or phrases where necessary.	Speaks with a clear, and fully intelligible pronunciation. Sometimes uses appropriate stress and intonation to explain the procedure and often reads rubrics highlighting important points.	Has discernible Greek pronunciation, which does not impede intelligibility. Uses stress and intonation to explain the procedure but only locally and rarely reads rubrics highlighting important points.	Has discernible Greek pronunciation, which may interfere with intelligibility. Does not use stress or intonation to explain the procedure and simply reads out rubrics rather mechanically.	Has discernible Greek pronunciation, which often interferes with intelligibility. He uses stress and intonation inappropriately or not at all, sometimes impeding intelligibility.
<b>Linguistic competence</b>	(When and if s/he is not reading from the Interlocutor Frame) Uses language spontaneously, fluently and precisely. Maintains consistent grammatical control of complex language with no grammar and syntax errors and can self-correct any slips of the tongue. Uses vocabulary related to the topic discussed correctly and appropriately.	(When and if s/he is not reading from the Interlocutor Frame) Uses language, fluently and precisely. Maintains consistent grammatical control of complex language with almost no grammar and syntax errors and can often self-correct any slips of the tongue. Uses vocabulary related to the topic discussed correctly and appropriately.	(When and if s/he is not reading from the Interlocutor Frame) Uses language fairly fluently and precisely in most cases. Maintains good grammatical control of complex language with occasional grammar and syntax errors which are self-corrected. Uses vocabulary related to the topic discussed quite correctly and almost appropriately most times.	(If s/he is not reading from the Interlocutor Frame) Lacks fluency and spontaneity. Generally maintains grammatical control of complex language but makes some minor grammar and syntax errors. Uses vocabulary related to the topic discussed sometimes incorrectly and/ or inappropriately	(If s/he is not reading from the Interlocutor Frame) Lacks fluency and spontaneity. Maintains grammatical control of simple language and makes some grammar and syntax errors. Uses vocabulary related to the topic discussed often incorrectly and/ or inappropriately

<b>Ability to follow test procedure</b>	Follows the procedure according to instructions. <sup>7</sup>	Follows the procedure according to instructions for the most part.	Follows the procedure according to instructions only to some extent.	Follows the procedure according to instructions to a very a very small extent.	Does not follow the procedure according to instructions at all.
<b>Ability to handle trouble situations</b>	Can effectively handle trouble situations, following the suggested course of action suggested by rules for troubleshooting.	Can handle trouble situations, following the suggested course of action but with some effort.	Can partly handle trouble situations, following some of the suggested course of action. Sometimes comes up with his/her own solutions	Can hardly handle trouble situations using the suggested course of action and/ or s/he mostly comes up with his/her own solutions.	Cannot handle trouble situations using the suggested course of action and/or resorts to personal ways of dealing with troublesome cases.

---

<sup>7</sup> The procedure includes the following points: 1) s/he takes into consideration the profile of the candidate when choosing questions/ tasks/ photos/ texts, 2) s/he asks 2- 4 questions from different categories (B1 & B2), 3) s/he does not change the rubrics, 4) s/he does not interrupt the candidates or interfere with their language output, 5) s/he allocates time equally to both candidates, 6) s/he keeps to the time provided by instructions for each activity and for the whole examination, 7) s/he follows the interlocutor frame faithfully etc.