Analysing and Evaluating Written Protocols by Raters of the KPG Speaking Test

Xenia Delieza,
PhD Candidate

Supervising Committee:
Professor B. Dendrinos
Professor M. Sifianou
Assistant Professor E. Karavas

PhD research: aim & context

- oral examiners and how they have differentially enacted the role they have been assigned - variation in examiner conduct ultimately results in unfair assessment

- *if, when* and *how* oral examiners, taking part in the KPG exams in English operate according to training instructions and guidelines
-'silent interlocutors', who mainly deliver questions and tasks
- if they involve themselves in what the candidate says, in his/her oral performance

- the consequences of such involvement in assessment - what sort of impact different degrees and types of involvement have on how the candidate's oral performance is evaluated and marked

- Context = the KPG oral exam in English at B2 (Independent User) and C1 (Proficient User) levels → difference in activities and examiner role

Examiner Involvement

- A key term

"Any instance of linguistic interference on the part of the interlocutor-examiner other than mere delivery of questions or tasks or test process description; it is a discourse practice which varies in quality and quantity depending on the context of occurrence; it is a discourse practice which presents a type of linguistic addition or change to the given question or task and/or the candidate's language production. This linguistic interference seems to affect in some way the candidate's linguistic output; such an effect also varies in quality and quantity."

Methodology of the PhD study

- Three methods for collecting data:
  1. Observation:
     - Observing oral examiners at work during the actual examination procedure - completing especially prepared forms.
     - The analysis of the data revealed different types, degrees and frequency of involvement.
     - Why examiners felt the need to get involved and what sort of effects the different types of involvement had on the candidate performance.

  2. Simulated speaking tests: recorded and transcribed
     - A combined method of discourse and conversation analysis was used: instances of examiner involvement - causes or reasons for examiner involvement - effects on the candidate's linguistic output.
     - Systematicity of co-occurrence of specific types of cause-involvement-effect was also sought and recorded.

3. Written protocols by raters of the KPG speaking test in English

- The third set of data consisted of written protocols produced by trained raters who assessed and marked the simulated speaking tests.

- These protocols were analysed with a view to understanding the ways in which examiner involvement may influence assessment.

- "*What do raters observe when they assess candidates? Do they internalise examiner involvement in the allocation of their final ratings? If yes, how and/ or to what extent?*"

The methodology and data collection

- Introspective method – stimulated recall: written protocols = a variant
- 6 B2 simulated interviews and 6 C1 simulated interviews
- 12 raters each assessing 1 B2 and 1 C1 simulated interview
- Each rater viewed each interview once, assigned a mark and produced the written protocol.
- For each interview = two marks and two written protocols

Mode of analysis

- Study of the reports → devise an analytical framework
- Segmenting the data into units of analysis: separate comments = evaluative comments
- Coding scheme
- Categorising also conducted by another coder-judge independently

Evaluative comments: five basic categories

- **Task**

e.g. "*What made the problem bigger, though, was that some of the tasks were probably beyond the 'culture' of the particular candidate, e.g. the film she had seen (She didn't seem much of a cinema-goer), changes that she would make in her university, etc.*"

- **Candidate**

e.g. "*Keera seems to be hesitant in either deciding which place she'd like to visit or whether 'France' as an answer would be an appropriate one.*"

- **Examiner**

e.g. "*The examiner engages in personal small talk ('Have you read the book? – You should. It's good fun').*"

- **Rating**

e.g. "*I gave her a 3 in linguistic and pragmatic competence.*"

- **Training**

e.g. "*Perhaps the examiners should be trained to handle situations like those.*"

The total of evaluative comments

	Comment about the…					
	task	candidate	examiner	rating	training	TOTAL
B2						
Number	4	207	407	40	1	659
Percentage	1%	31%	62%	6%	0%	100
C1						
Number	8	226	249	41	8	532
Percentage	2%	42%	47%	8%	2%	100

- More comments in B2 than in C1 speaking tests.
- In B2, comments about the examiner = almost twice as many as for the candidate
- In C1, examiner-candidate = almost the same frequency
- The role of examiner in each test
- The level in each test

The "Candidate": sub-categories

Category: *Candidate*	B2		C1	
The rater…	No of comments	%	No of comments	%
uses assessment criteria to comment on the candidate's achievement.	167	81	118	52
refers to an action by the candidate.	0	0	33	15
comments on the level of the candidate.	20	10	26	12
compares the candidates.	11	5	22	10
comments that the candidate helps his/her colleague.	1	0	10	4
comments on something other than the above.	8	4	17	8
Total	207	100	226	100

- At B2 level, most comments about candidates (81%) = on their degree of achievement relative to the set questions or tasks, based on the assessment criteria for the level

- At C1 level, raters most frequently comment on the candidates' achievement by use of the assessment criteria (52%), followed by comments relating to actions by the candidate (15%)(=0 in B2)

The "Examiner": sub-categories

Category: *Examiner*	B2		C1	
The rater…	No of comments	%	No of comments	%
comments on the examiner's involvement (and its effect on the candidate or on the rater's point of view).	208	51	133	53
comments on how the examiner handled procedural issues.	96	24	65	26
comments on something the examiner does that the rater seems to approve of.	59	14	29	12
comments on the examiner's personal behaviour and/ or how it affects candidates.	34	8	18	7
comments on how the examiner behaves as a teacher.	10	2	4	2
Total=	407	100	249	100

- When raters comment on examiner action in both levels, they refer to examiner involvement in more than half of their remarks (51% at B2 level and 53% at C1 level).

Comments per candidate – B2 Level

B2	Name	Rater	Mark	Comment on... (presented in percentage %)				
SIM.	Name	Rater	Mark	task	candidate	examiner	rating	training
2	A. Gregory	S.P.	6	0	20.3	78.5	1.3	0
	A. Gregory	P.B.	11	0	11.1	**88.8**	0	0
	B. Demy	S. P.	17	1.1	19.6	73.9	5.4	0
	B. Demy	P. B.	17	0	9.5	85.7	4.8	0
3	A. Mina	V.M.	12	4.7	51.2	32.6	11.6	0
	A. Mina	M.S.	16	0	11.1	77.8	11.1	0
	B. Gina	V. M.	11	2.6	21.1	65.7	7.9	2,6
	B. Gina	M.S.	15	0	10.5	78.9	10.5	0
4	A. Vera	E.C.	7	0	41.2	29.4	29.4	0
	A. Vera	I.P.	7	0	**69.7**	30.3	0	0
	B. Sheila	E.C..	15	0	47.1	23.5	29.4	0
	B. Sheila	I. P.	11	0	64.9	35.1	0	0
5	A. Keera	A.B.	19	0	18.5	81.4	0	0
	A. Keera	P.Bo.	15	0	21.4	78.6	0	0
	B. Nolan	A.B.	20	0	16.7	78.3	8.3	0
	B. Nolan	P.Bo.	18	0	25	75	0	0
6	A. Pat	M.H.	18	0	15.4	69.2	15.4	0
	A. Pat	V. Z.	18	0	58.3	33.3	8.3	0
	B. Valerie	M.H.	20	0	52.6	47.4	0	0
	B. Valerie	V.Z.	20	0	53.8	38.5	7.7	0
7	A. Vanessa	M.D.	11	0	25.0	68.8	6.3	0
	A. Vanessa	V.O.	14	0	37.5	62.5*	0	0
	B. Effie	M.D.	13	0	31.3	62.9*	6.3	0
	B. Effie	V.O.	12	0	66.7	33.4	0	0
	C. Eleanor	M.D.	6	0	43.5	52.1*	4.3	0
	C. Eleanor	V.O.	7	0	20.0	70.0	10.0	0

Comments per candidate – C1 Level

C1				Comment on…		(presented in percentage %)		
SIM.	Name	Rater	Mark	*task*	*candidate*	*examiner*	*rating*	*training*
2	A. Jim	A.B.	12	0	21.1	78.9	0	0
	A. Jim	P.Bo.	1	0	23.1	38.5	7.7	30.8
	B. Elena	A.B.	15	0	27.8	72.2	0	0
	B. Elena	P.Bo.	8	0	16.7	41.7	8.3	33.3
3	A. Aline	S.P.	9	3.8	52.8	37.7	5.7	0
	A. Aline	P.B.	8	0	24.1	72.4	3.4	0
	B. Denis	S. P.	11	2.4	70.7	24.4	2.4	0
	B. Denis	P. B.	13	0	18.8	78.1	3.1	0
4	A. Dina	V.M.	12	8	32	40	20	0
	A. Dina	V.B.	12	0	53.8	38.5	7.7	0
	B. Stacey	V.M.	13	0	50	38.5	11.5	0
	B. Stacey	V.B.	12	0	28.6	71.4	0	0
5	A. Gina	E.C.	7	5.3	47.4	42.1	5.3	0
	A. Gina	I.P.	8	0	31.3	62.5	6.3	0
	B. Alicia	E.C..	16	0	53.3	46.7	0	0
	B. Alicia	I. P.	12	0	36.8	47.4	15.8	0
6	A. Myriam	M.H.	8	0	50	46.2	3.8	0
	A. Myriam	V. Z.	9	5.3	66.7	20.8	8.3	0
	B. Jamie	M.H.	12	0	67.7	32.3	0	0
	B. Jamie	V.Z.	11	4.2	63.2	26.3	5.3	0
7	A. Mathew	M.D.	16	0	20	40	40	0
	A. Mathew	V.O.	14	0	50	50	0	0
	B. Velma	M.D.	12	0	16.7	41.7	41.7	0
	B. Velma	V.O.	12	0	44.4	44.4	11.1	0

Results

- 5 categories of evaluative comments were found.
- The most popular were on *candidates* and *examiners*.
- In B2, *examiners* were more frequently commented upon.
- In C1, comments on *examiners* were found to be balanced with the ones on *candidates*.
- *Task* and *rating* are not frequently commented on, while *training* is very rarely so.
- It is not clear exactly how raters internalise what they observe when assigning marks.
- Marks could not be correlated with types of comments in none of the two levels.
- There seems to be enough evidence that raters *are* aware of the role examiners play and the ways they may get involved in the candidates' performance.
- Findings from the written protocol analysis combined with insights from observation and analysis of simulated speaking tests led to the design of a model depicting the examiner's impact on the assessment of candidates in the KPG oral exam in English context.
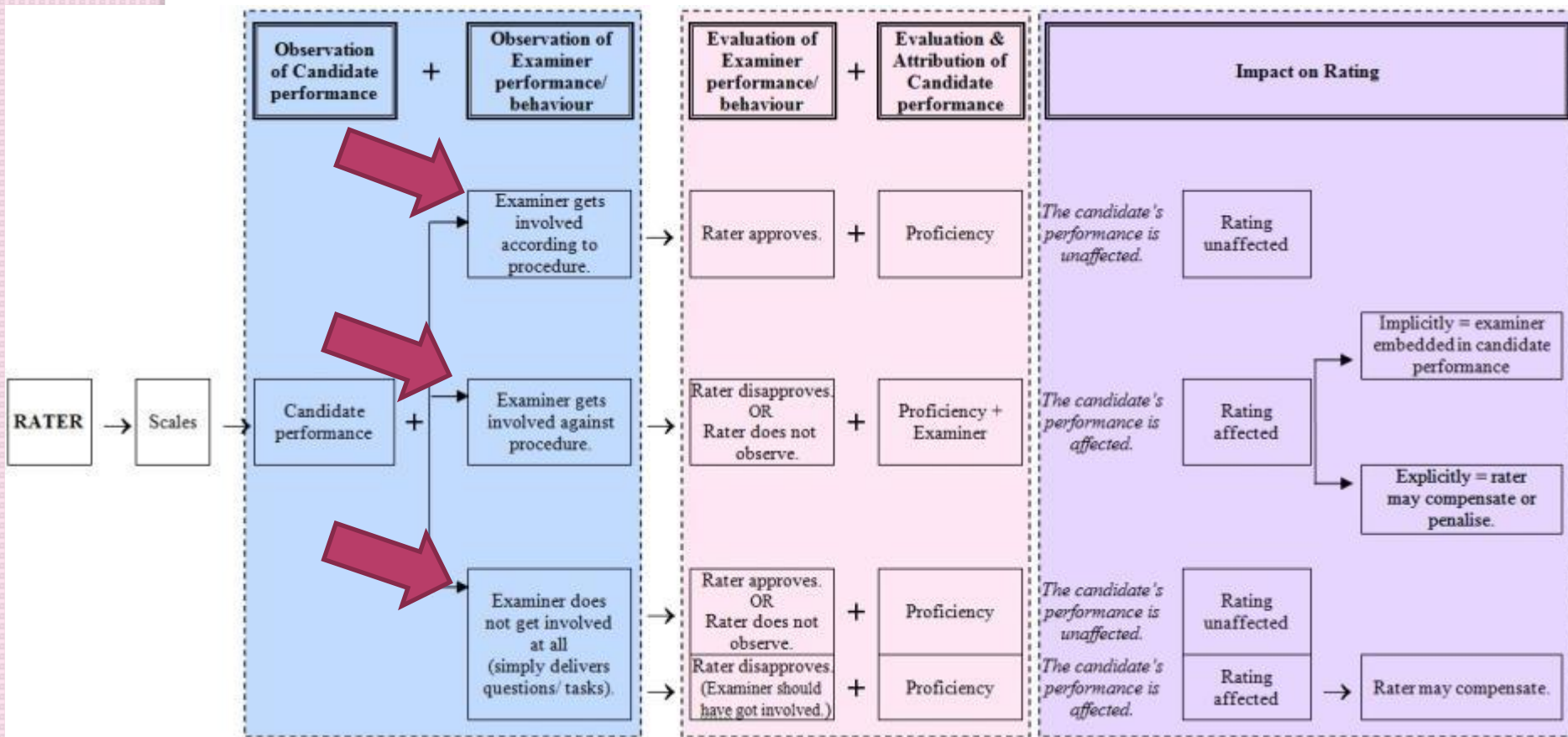
A model of the examiner's impact on the assessment of oral performance in the KPG oral exam in English

- This model offers a detailed view of the role the examiner-as-interlocutor plays in the candidate's performance – provides a pictorial representation of the various steps of assessment and the way in which the examiner may be embedded in the process as a factor which cannot be distinguished and evaluated separately

- According to the model, raters consult the scales in order to assess a candidate - they observe candidate performance - they also observe the examiner's involvement in this performance – involvement categorised into three basic types of courses of actions

In the 1st type, the examiner uses the Interlocutor Frame as given and

In the 2nd type, the examiner gets involved in ways which have been

In the 3rd type, the examiner uses the Interlocutor Frame, delivers tasks and questions verbatim, but does not get involved in any way whatsoever, even if/ when there are trouble situations. When this is approved of or not observed by the rater , the final mark =unaffected. In cases where the rater feels that the examiner should have got involved, s/he may compensate - the final mark = affected.

Observation of Candidate performance	+	Observation of Examiner performance/ behaviour		Evaluation of Examiner performance/ behaviour	+	Evaluation & Attribution of Candidate performance		Impact on Rating		
		Examiner gets involved according to procedure.	→	Rater approves.	+	Proficiency	The candidate's performance is unaffected.	Rating unaffected		
Candidate performance	+	Examiner gets involved against procedure.	→	Rater disapproves. OR Rater does not observe.	+	Proficiency + Examiner	The candidate's performance is affected.	Rating affected	Implicitly = examiner embedded in candidate performance
									Explicitly = rater may compensate or penalise.
		Examiner does not get involved at all (simply delivers questions/ tasks).	→	Rater approves. OR Rater does not observe.	+	Proficiency	The candidate's performance is unaffected.	Rating unaffected		
			→	Rater disapproves. (Examiner should have got involved.)	+	Proficiency	The candidate's performance is affected.	Rating affected	→	Rater may compensate.

RATER → Scales →

A model of the examiner's impact on the assessment of oral performance in the KPG oral exam in English

Thank you for your attention.

Xenia Delieza,

xdelieza@enl.uoa.gr or delieza@otenet.gr