

# How much does test-takers' listening proficiency matter in oral interview tests?

---

Fumiyo Nakatsuhara  
CRELLA  
University of Bedfordshire

The 8th Annual EALTA Conference  
5 – 8 May 2011 Siena, Italy

# Acknowledgement

*•The presenter acknowledges the role of the IELTS Partners in making this study possible: The British Council provided the research grant which enabled me to conduct the study as part of the 2010-11 IELTS Funded Research Programme. The University of Cambridge ESOL Examinations provided access to two sets of examiner prompts from the DVD, IELTS Scores Explained for the purpose of this project.*

---

*•Any opinions, findings, conclusions or recommendations expressed in this material are those of the presenter and do not necessarily reflect the views of the British Council, its related bodies or its partners.*

# Research Background

## Listening in paired/group oral tests

- **Ducasse & Brown (2009): 2 demonstrations of listening comprehension that contribute to successful interaction**
  - showing evidence of comprehension by the listener (e.g., filling in with a missing word to help the partner)
  - showing supportive listening by providing audible support (e.g., back-channeling).
- **Nakatsuhara (2009): listening-related problems in group oral tests**
  - Leading to negotiation of meaning

# Possible effects of Listening on oral interview test performance (IELTS studies)

- **Seedhouse and Egbert (2006):**
  - Interactional problems can be caused by test-takers' *misunderstanding* of what the examiner has said
- **O'Sullivan and Lu (2006):**
  - A far greater number of examiner deviations from the interlocutor frame (particularly *paraphrasing questions*) in Part 3 (discussion) than Parts 1 & 2 (interview on familiar topics & Individual long turn). → pointing to difficulty with *listening comprehension*.

**Are oral interview tests at least to some extent tapping the construct of listening-into-speaking?**

# Research Questions

**RQ 1:** Is there evidence of any **difference in difficulty** between Part 2 (individual long turn) and Part 3 (discussion) of the IELTS Speaking test identified by **overall** scores and scores given to **each analytical category**?

**RQ2:** What are the relationships between test-takers' **listening proficiency** and **overall and analytical scores** awarded on Part 2 and Part 3 respectively?

**RQ 3:** How do **communication problems** in Part 3 that seem to be related to test-takers' difficulties in understanding the examiner occur and how are they dealt with?

# Method of Data Collection & Data Analysis

# Data collection

## Participants

- 36 pre-sessional course students at a UK University (IELTS 3.0 - 8.0)
- 4 trained examiners

## Listening test

- Test of *general* listening ability → a Cambridge Main Suite test
- Both FCE (B2) and CAE (C1) items (27 items: Cronbach's Alpha .918)

## Speaking test

<b>Warming-up</b>	a very brief warm-up conversation (30 sec - 1 min)	
<b>Part 2</b>	<b>Individual long turn</b> (3-4 mins)	2 prompts:
<b>Part 3</b>	<b>Discussion</b> (4-5 mins)	1) <i>Interest &amp;</i> 2) <i>Parties</i>



## **Audio-rating of the speaking performance**

- Non-live marking
- A mixture of separate Part 2 and Part 3 audio-recordings from different test-takers were given to the examiners.
- *1) Fluency and Coherence, 2) Lexical Resource, 3) Grammatical Range and Accuracy, and 4) Pronunciation*

## **Examiners' notes**

- Reasons for awarding the scores that they did on each analytical category

## **A short semi-structured interview immediately after each speaking test**

- The participating students' perceptions of any communication problems encountered with the examiner.

# Data Analysis

- **Quantitative analysis**

- Comparing overall and analytical speaking scores awarded on Parts 2 and 3 (**RQ1**).
- Comparing the strength of the correlations between the listening scores and the overall and analytical speaking scores awarded on Parts 2 and 3 (**RQ2**).

- **Qualitative analysis**

- CA: to illustrate how listening-related communication problems in Part 3 occurred (**RQ3**)
  - Short interview data about the students' speaking test experience
  - Examiners' notes on scoring

# Results

# Comparing Parts 2 and 3 overall/analytical scores (RQ1): Rasch analysis

- **Overall scores:**

- **Analysis of 5 facets** (*examinee ability, examiner harshness, prompt difficulty, part difficulty and rating category difficulty*)
- The *part* facet did **not** show a significant difference between Parts 2 and 3 overall scores

- **Analytical scores:**

- The *part* facet had a **statistically significant impact** only on the ***Fluency and Coherence*** category ( $\chi^2=7.4$ ,  $p=.01$ ).
- **Part3 was significantly more difficult** than Part 2, although the actual difference was rather small (Fair average scores: Part 3=4.88, Part2=4.99).

- In some cases, there were **noticeable differences** in the band scores obtained by individual test-takers on these two parts of the test (greater than 1.0 band).

# Relationship between listening & speaking scores (RQ2)

Speaking in <b>Part 2</b>	Overall	Flu	Lex	Gram	Pron
Spearman's rho	<b>.418</b>	<b>.471</b>	<b>.490</b>	<b>.481</b>	.294
Sig	<b>.011</b>	<b>.004</b>	<b>.002</b>	<b>.003</b>	.082
Speaking in <b>Part 3</b>	Overall	Flu	Lex	Gram	Pron
Spearman's rho	<b>.597</b>	<b>.522</b>	<b>.643</b>	<b>.643</b>	<b>.411</b>
Sig	<b>.000</b>	<b>.001</b>	<b>.000</b>	<b>.000</b>	<b>.013</b>

- **Stronger correlations** between listening scores and speaking scores in **Part 3** (large strength) than those in Part 2 (medium strength).
- The correlational differences for *Overall* and *Lexical Resource* **approached significance** (*Overall*:  $t(33)=-1.604, p=.059$  ; *Lexical Resource*:  $t(33)=-1.543, p=.066$ )

## Listening-related communication problems (RQ3)

- **Type a)** asking a question and then responding relevantly
- **Type b)** asking a question and then responding irrelevantly
- **Type c)** misunderstanding a question and
  - **c1)** responding very irrelevantly or
  - **c2)** giving somewhat related but mostly irrelevant response
- **Type d)** echoing uncomprehended parts
- **Type e)** answering “no” to an uncomprehended question

**17 instances in total**

# Type *a*) asking a question and then responding relevantly

## Excerpt 1. Student *S30*

Listening	Speaking	Flu	Lex	Gram	Pron
5	Part 2	4.02	3.62	4.02	4.62
	Part 3	4.03	3.85	3.85	4.49

Fair average scores

1 E: What makes a good family party?

2 → S: Sorry? ((moving forward))

3 E: What makes a good family party?

4 S: Uh:: uh Maybe just ah food. (1.0) uh:: (.5) uh in China, family party ...

- (*S30*) “Sometime I don't understand questions. About parties, family party, friends party, I don't understand questions”.
- No effect on the examiners' impression of the candidate's performance, as examiners' comments between Parts 2 and 3 were highly consistent

- Students who scored over 5.0 hardly encountered listening difficulties.

## Excerpt 2. Student S24

Listening	Speaking	Flu	Lex	Gram	Pron
21	Part 2	4.77	4.86	4.67	4.22
	Part 3	4.81	5.83	5.83	5.00

1 E: So what what needs to be planned when you are organising a formal party?

2 → S: **What do you mean, to be [planned]?**

3 E: [What needs to be planned?]

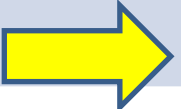
4 S: =Ah:: firstly uh time, venue, venue the(h)re ((scratching his head)) ah...

- (S24) “I didn’t find understanding the examiner difficult at all”.  
→ No decoding difficulty, but perhaps taking some time in processing the meaning in the specific context.



## Type b) asking a question and then responding irrelevantly

### Excerpt 3. Student S07

Listening	Speaking	Flu	Lex	Gram	Pron
10	Part 2	5.05	4.87	4.92	5.36
	<b>Part 3</b>	<b>3.79</b>	<b>4.01</b>	<b>4.01</b>	<b>4.04</b>

1 E: What about, ah how do family parties compare to more formal parties?

2→S: **Uh::: (.5) uh:: ((smiling)) ca(h)n you repeat uh ( ) your question?**

3 E: Family parties, how do they compare with more formal parties, such as

4 parties in schools or universities or in a work place?

5→S: **Uh: So compare uh the uh party from home and another ah: place. uh (.5) uh:**

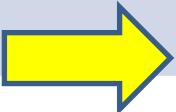
- Long filled pauses → Band 4 of *Fluency and Coherence*: “cannot respond without noticeable pauses” = Examiners’ comments in Part 3



Related to the fluency construct in Part 3, which involves a role for listening proficiency.

**Type c) misunderstanding a question, and c1) responding very irrelevantly or c2) giving a somewhat related but mostly irrelevant response**


**c1) Excerpt 4. Student S17**

Listening	Speaking	Flu	Lex	Gram	Pron
3	Part 2	4.82	4.87	4.92	4.92
	<b>Part 3</b>	<b>2.00</b>	<b>2.85</b>	<b>2.85</b>	<b>3.49</b>

- 1 E: How do families celebrate birthdays in your country?
- 2 → S: **uh birthday is the ah first (.5) ah October.**
- 3 E: Uh huh
- 4 S: yeah.
- 5 E: How how do they celebrate birthdays?
- 6 → S: (2.5) Ah (.5) I don't know how to call this (.5) ah (1.0)
- 7 E: Do families have parties for a birthday?

- Considerably lower scores in Part 3

## c1) Excerpt 5. Student S18

Listening	Speaking	Flu	Lex	Gram	Pron
5	Part 2	3.66	4.03	4.07	4.07
	<b>Part 3</b>	<b>3.00</b>	<b>3.06</b>	<b>3.06</b>	<b>3.06</b>

- 1 E: Do you think having a hobby is good for people's social life?
- 2 → S: %to socia% Yeah, I think it's good to people. Ah: (.5) uh
- 3 example, for example uh **I like ah to (eat) apple everyday.**
- 4 I my ( ) tell me uh ah:: one day one people, people have.
- 5 → E: Is it a good way to meet new people?
- 6 S: Yeah
- 7 (1.0)

- When the response was very irrelevant (c1) → Considerably lower scores in Part 3
- If irrelevant but somewhat related (c2) → The scores did not differ much

## Type d) echoing uncomprehended parts

### Excerpt 6. Student S11

Listening	Speaking	Flu	Lex	Gram	Pron
10	Part 2	4.04	4.04	4.12	3.88
	Part 3	3.79	4.01	4.01	3.66

1 E: Have ah celebrations changed in your life time?

2→S: ah:: **Changed ((pinching his cheek))**

3 E: Have have they changed? Are they different?

4→S: (.5) Uh: (1.0) Can you repeat it again?

- *Examiner C*: “frequently fails to understand the question” as her reason for awarding the Fluency and Coherence score

## Type e) answering “no” to an uncomprehended questions

### Excerpt 7. Student S09

Listening	Speaking	Flu	Lex	Gram	Pron
11	Part 2	3.82	3.35	4.12	4.06
	Part 3	3.79	3.03	3.03	4.04

1 E: Ah, Have family celebrations changed in your lifetime?

2 (2.5)

**3→ S: No no uh**

4 E: Have they, they have always been the same?

5 (2.0)

**6→ S: N(h)o**

7 E: Uh, can you tell me more about this?

**8→S: (1.0) Uh .hh can repeat the question, sorry?**

- A test-taking strategy: “Examiners are likely to move on to the next question if I give a negative answer”

# Conclusion

# Main Findings

1. **No sig** difference in ***Overall Scores*** between Parts 2 and 3
2. Part 3 was sig more difficult than Part 2 for the ***Fluency and Coherence*** category (but actual difference: small)
3. Some test-takers got **considerably different scores** between Parts 2 and 3.
4. **Stronger correlations** between listening and Part 3 scores than between listening and Part 2 scores

**5. A certain type of listening-related problems could be contributing more to lowering test-takers' Part 3 scores.**

- **Type b)** asking a question and then responding irrelevantly
- **Type c1)** misunderstanding a question and responding very irrelevantly

**6. A possible boundary in bands where the degree of impact of test-takers' listening proficiency changes might be Band 5.0.**



- Brown (2006): **The *Fluency and Coherence* scale was the most difficult for examiners to interpret:** interpreting hesitation (a search for ideas or for language?)



Another level of complexity: hesitating because of their speaking-related limitations or listening-related limitations.

- ✧ Part 3: Tapping listening-into-speaking construct to some extent → **Formulating different *Fluency and Coherence* descriptors for Parts 2 and 3**, explicitly making the Part 3 descriptors reflect the construct of listening-into-speaking → Scoring each part separately.

- ✧ Some students get very different scores between Part 2 and Part 3 → worth considering to **score each part separately for all categories** to provide a clearer picture of test-taker's differential abilities across these different modes of speaking (e.g., O'Sullivan, 2007).
- ✧ Too difficult to understand almost all questions → worth considering **to prepare easier question scripts for low-level test-takers**, to obtain ratable speech samples from them.

**Thank you!** 😊

---

Fumiyo Nakatsuhara  
[Fumiyo.Nakatsuhara@beds.ac.uk](mailto:Fumiyo.Nakatsuhara@beds.ac.uk)  
[fumettina@gmail.com](mailto:fumettina@gmail.com)

The 8th Annual EALTA Conference  
5 – 8 May 2011 Siena, Italy