

1. The Impact of Gender in the IELTS Oral Interview

Kieran O'Loughlin
Department of Language, Literacy & Arts Education
The University of Melbourne

Abstract

This paper reports on a study into the impact of gender on the IELTS oral interview. This is a relatively underinvestigated issue in the assessment of oral proficiency. The study examines the issue of gender on two levels: firstly, its impact on the discourse of the interview and secondly, its effect on the rating process.

There is a large body of research which suggests that male and female speakers have distinctive communicative styles. Therefore, it might be anticipated that such differences would be reflected in the discourse of interviewers in the oral test interview, possibly affecting the quantity and quality of the candidate's output. Furthermore, candidates' output may vary in relation to their own gender and whether their interviewer is of the same or opposite sex. It is also possible that the gender of the rater and/or candidate may significantly influence assessment of the oral interview. In the case of tests like the IELTS interview where the interlocutor also acts as the rater this poses the question of whether gender bias, where it exists, stems from the interview itself, the rating decision or a combination of both these 'events'.

The study is based on interviews undertaken with sixteen candidates (eight female and eight male) who were each interviewed by a female and male interviewer. This yielded a total of 32 interviews. Each interview was rated by the interviewer and audiotaped. Four other raters (two females and two males), drawn from a pool of eight females and eight males, subsequently assessed each of the interviews using the audio-recordings. The audio-recordings were then transcribed and several features of language use which have been identified in previous research as key markers of gendered communication were examined, specifically the use of overlaps, interruptions and Minimal responses by both interviewers and candidates. The test score data was analysed using a facility of the multi-faceted Rasch computer program FACETS (Linacre, 1989-1995) known as bias analysis.

In the discourse analysis of the interviews it was found that there were some gender differences between female and male interviewers and candidates, but these did not form a consistent gender pattern. In general, most interviewers and candidates adopted a supportive and collaborative speech style irrespective of their own gender or the gender of their interlocutor. Furthermore, the analysis of test scores indicated there was no evidence of significant bias in the rating process in relation to the gender of raters or candidates. Both sets of findings therefore suggest that gender does not have a significant impact on the IELTS interview.



Publishing details

International English Language Testing System (IELTS) Research Reports 2000 Volume 3

Editor: Robyn Tulloh

IELTS Australia Pty Limited
ABN 84 008 664 766
Incorporated in the Australian Capital Territory
Web: www.ielts.org

© 2000 IELTS Australia.

This publication is copyright. Apart from any fair dealing for the purposes of private study, research or criticism or review, as permitted under the Copyright Act, no part may be reproduced by any process without written permission. Enquiries should be made to the publisher.

National Library of Australia
Cataloguing-in-Publication Data
2000 ed
IELTS Research Reports 2000 Volume 3
ISBN 0 86403 036

1.0 Introduction

Recent research into oral language interviews has indicated that interviewers vary considerably from each other in relation to their test behaviour. Such variability includes the amount of support interviewers give to candidates, the amount of rapport they establish with candidates and the extent to which they follow the instructions in terms of the type of discourse elicited from candidates. A previous study of the IELTS oral interview (Brown and Hill, 1996) indicates that different interviewers present different 'hurdles' for candidates and that there are particular interviewing styles which characterise 'easy' and 'difficult' interviewers. Brown and Hill found that the easier interviewers tended to shift topic more frequently and asked simpler questions. The more difficult interviewers tended to use a wider range of interactional behaviours, such as interruption and disagreement as well as asking more challenging questions.

A key issue arising from such findings is *why* interviewers vary from each other. One possibility is that such variability stems, at least partly, from gendered differences in communicative style. There is a large body of research in the field of language and gender which suggests that male and female conversational styles are quite distinct (see, for example, Maltz and Borker, 1982; Tannen, 1990; Coates, 1993; Thwaite, 1993). These studies characterise the female conversational style as collaborative, co-operative, symmetrical and supportive whereas its male equivalent is portrayed as controlling, unco-operative, asymmetrical and unsupportive.

More recently, such research has been strongly criticised for its tendency to overgeneralise its findings to all men and all women irrespective of other social identity factors such as their age, ethnicity and class and, most relevantly for this project, situational factors such as the communicative context and the gender of their interlocutors (see, for example, Freed, 1995; Freed and Greenwood, 1996; Freeman and McElhinny, 1996; Stubbe, 1998). In the testing context, whether such clearly distinct styles are consistently evident for male and female interviewers in oral test interviews needs to be investigated together with what effects such gendered differences (if they exist) have on candidate performance.

Sunderland (1995) suggests that differences in male and female interviewer styles can be viewed as one potential source of test bias. Another possibility she identifies is that the behaviour of interviewers of either sex may vary according to whether they are paired with a male or female candidate. In both cases, it is feasible that the gendered behaviour of the interviewer will influence the outcome of the test by either strengthening or undermining the candidate's performance.

Another obvious gender consideration in oral test interviews is that candidates' output may vary according to their own gender. Equally, as suggested above, the quantity and quality of their output may be affected positively or negatively by the gender of their interviewer.

Finally, it is also worth considering whether there is a gender effect in the rating of oral interviews. It is possible in oral interviews that male and female raters may assess differently. It is also possible that their assessments are influenced by the sex of the candidate. In the case of tests like the IELTS interview where the interlocutor also acts as the rater this poses the question of whether gender bias, where it exists, stems from the interview itself, the rating decision or a combination of both these 'events'.

If there is a gender effect in oral interviews then it needs to be seriously addressed in the interests of minimising test bias.

The study aimed therefore to address the following questions:

- 1 What impact does the gender of participants have on the discourse produced in the IELTS oral interview?
- 2 What impact does the gender have on the rating of the IELTS oral interview?
- 3 If gender bias is found to exist in the course of interviewing and/or rating, how might its effects on test scores be minimised?

2.0 Methodology

2.1 The IELTS oral interview

The Speaking Module takes between 10 and 15 minutes. It consists of "an oral interview, a conversation, between the candidate and an examiner" (IELTS, 1997:14). It should be noted that 'an interview' is not always 'a conversation'. However, the IELTS interview includes a variety of task types, at least some of which do appear to tap the candidates' conversational ability (see particularly phases 2 and 4 below).

There are five sections:¹

- Phase 1. Introduction**
The examiner and candidate introduce themselves. The candidate is made to feel comfortable and encouraged to talk briefly about their life, home, work and interests.
- Phase 2. Extended Discourse**
The candidate is encouraged to speak at length about some very familiar topic either of general interest or of relevance to their culture, place of living, or country of origin. This will involve explanation, description or narration.
- Phase 3. Elicitation**
The candidate is given a task card with some information on it and is encouraged to take the initiative and ask questions either to elicit information or to solve a problem. Tasks are based on 'information gap' type activities.
- Phase 4. Speculation and Attitudes**
The candidate is encouraged to talk about their future plans and proposed course of study.
Alternatively the examiner may choose to return to a topic raised earlier.
- Phase 5. Conclusion**
The interview is concluded.

¹ Editor: The format of the Speaking test will change from 1 July 2001. See Appendix 6.1 for details

The Speaking Module assesses whether candidates have the necessary knowledge and skills to communicate effectively with native speakers of English. Interviewers are given an outline which includes suggested topics for Phases 2 and 4 and a prescribed task for Phase 3. The interviewers also carry out the assessment of the candidate's proficiency using a global band scale with 9 increments. Assessment takes into account evidence of communicative strategies, and appropriate and flexible use of grammar and vocabulary. The interviewer is a qualified teacher and certificated examiner appointed by the test centre and approved by University of Cambridge Local Examinations Syndicate (UCLES). All interviews are recorded (UCLES 1997:14).

2.2 Interview Design

Sixteen different students (eight male and eight female) and eight accredited IELTS interviewers (four male and four female) participated in this stage of the study. Each of the candidates were interviewed on two different occasions by a male and a female interviewer yielding a total of 32 interviews. Each of the interviews were audiotaped, as they are in the official IELTS Speaking component.

The candidates were ELICOS students engaged in an IELTS preparation course with the aim of undertaking further study here in Australia. Consequently, they volunteered for this project on the basis of experiencing the interview under exam-like conditions, gauging their readiness and receiving feedback from the interviewers about possible areas to develop in their preparation for the official test. The candidates came from a range of language and cultural backgrounds. The women came from China, Indonesia, Japan and Thailand and ranged in ages from 19 - 31. The men came from China, Indonesia, Japan, Korea and Thailand and were in the age range of 20 - 30.

The interviewers were all fully trained, current IELTS examiners ranging in age, work place and length of time as an examiner. They were all paid \$60 for their participation but, like the candidates, were not given any indication of the focus of the project beyond it being a study of the discourse produced in the oral interview. As indicated above, the interviewers were each asked to provide feedback to the candidates about their strengths and weaknesses in preparation for the official test. This was done immediately after each interview.

All the students were interviewed twice, once by a female interviewer and once by a male interviewer. Half of the students were interviewed by a male interviewer first and the other half by a female interviewer first. The interviews were done at the same site on two different days. Candidates were not exposed to the same topics in the two interviews so as to minimise any potential practice effect. For this purpose two different test versions (Test A and Test B) were used as outlined below.

Test A

Phase 1: Introduction.

Phase 2: Extended discourse.

Topics such as Ceremonies, Education, Tourism, Leisure interests and Building and architecture were suggested to the interviewers.

- Phase 3: Elicitation.
'Living Cheaply in the City' was the prescribed task.
- Phase 4: Speculation and attitudes.
Topics such as Work plans/vocational field - intentions, benefits to home country, family and friends attitudes, career stages were suggested to the interviewers.
- Phase 5: Conclusion.

Test B

- Phase 1: Introduction.
- Phase 2: Extended discourse.
Topics such as Travel and transport, Work, Pollution, Family life and Wealth/poverty were suggested to the interviewers.
- Phase 3: Elicitation.
'Football Match' was the prescribed task.
- Phase 4: Speculation and attitudes.
Topics such as Study plans/academic field - importance, benefits, family and friends' attitudes were suggested to the interviewers.
- Phase 5: Conclusion.

2.3 Rating Design

Each candidate was subsequently assessed by two female and two male accredited IELTS interviewers, again from a range of ages, workplaces and experience as IELTS examiners, using the audio-recordings of the interviews. Each were paid \$80 for their participation and, like the interviewers were not given any indication of the focus of the project beyond it being a study of the discourse produced in the oral interview. A mixed design was used for these additional ratings whereby each interview was assessed by different combinations of male and female raters drawn from a pool of eight females and eight males with each rater carrying out a total of eight assessments. This design enabled the raters to be calibrated against each other in relation to their potential gender bias in the statistical analyses which followed. For the purposes of clarity note that in the rest of this report 'interviewer score' refers to the rating assigned by the original interviewer and 'rater score' refers to an assessment subsequently made by one of the additional raters based on the audio-recordings of the interview.

Table 1 below outlines the design for both the interviews and ratings.

Candidate No.	Candidate Gender	Interviewer No.	Interviewer Gender	Female Raters Nos.		Male Raters Nos.	
1	Male	1	Male	5	6	9	10
1	Male	2	Female	1	2	11	12
2	Male	1	Male	3	4	13	14
2	Male	2	Female	5	6	15	16
3	Female	1	Male	7	8	10	12
3	Female	2	Female	2	4	9	11
4	Female	1	Male	1	3	14	16
4	Female	2	Female	6	8	13	15
5	Male	3	Male	5	7	9	12
5	Male	4	Female	1	4	10	11
6	Male	3	Male	2	3	13	16
6	Male	4	Female	5	8	14	15
7	Female	3	Male	6	7	12	16
7	Female	4	Female	4	8	9	13
8	Female	3	Male	1	5	10	14
8	Female	4	Female	2	6	11	15
9	Male	5	Male	3	7	9	14
9	Male	6	Female	1	6	10	13
10	Male	5	Male	2	5	11	16
10	Male	6	Female	3	8	12	15
11	Female	5	Male	4	7	10	16
11	Female	6	Female	2	8	9	15
12	Female	5	Male	1	7	12	14
12	Female	6	Female	4	6	11	13
13	Male	7	Male	3	5	9	16
13	Male	8	Female	1	8	10	15
14	Male	7	Male	2	7	11	14
14	Male	8	Female	3	8	12	13
15	Female	7	Male	4	5	11	16
15	Female	8	Female	3	4	9	10
16	Female	7	Male	1	2	15	16
16	Female	8	Female	7	8	13	14

Table 1: Interview and rating design

3.0 Findings

3.1 Discourse analysis

The interviews were transcribed, using transcription notation (see Appendix 1.1), and then analysed in detail in relation to effect of the different gender pairings as follows:

<i>Interviewer</i>	<i>Candidate</i>
Female	Female
Female	Male
Male	Male
Male	Female

Coates (1993) takes up Hymes (1972) notion of communicative competence as a sense of knowing how language is used in a given society, ie. it is essential to understand socio-cultural factors as well as grammar and phonology when learning a language. She acknowledges that our understanding of when to speak, when to remain silent, what to talk about and how to talk about it in different circumstances is learned. She argues that women and men seem to differ in their communicative competence: they differ in their sense of what is *appropriate* speech behaviour.

The analysis which follows focuses on three of Coates' (1993) key markers of gendered communication as used in the IELTS interviews examined in this study: overlaps, interruptions and Minimal responses. These were considered to be potentially the most salient categories for comparison between females and males in this context.

3.1.1 Overlaps

Coates (1993:109) defines overlaps as "instances of slight over-anticipation by the next speaker: instead of beginning to speak immediately following current speaker's turn, next speaker begins to speak at the very end of current speaker's turn, overlapping the last word (or part of it)".

Coates (1993) cites studies that observed more overlaps in same-sex pairs. These studies also showed that all overlaps in mixed-sex pairs were caused by males, and that women used no overlaps with men but did in same-sex talk. She suggests women are concerned not to violate the man's turn but to wait until he has finished speaking.

The overlaps observed in the IELTS interview data collected for this study appeared to express different meanings. They have therefore been classified in two ways. There were 'positive overlaps' that seemed to be offering support for the person whose turn it was, both by confirming information and continuing the topic. Also some 'negative overlaps' were observed, particularly attempts to control the topic. Examples of both kinds of overlaps follow.

3.1.1 A Interviewer Overlaps

Examples of interviewer overlaps are included below.²

Positive

i) Giving Confirmation

Below an example is provided of the interviewer confirming the candidate's idea that 'many people want to see this game'. As the candidate reformulated this (line 221), the interviewer perhaps recognised the candidates need for support in this idea and thus joined in to confirm it:

C: So many people want to see this game.

I: Sure.

C: Looking forward to [see this game.]

* *I: [lots of students] want to go and see this game.*

Interviewer 3 (male)/Candidate 6 (male): lines 219-222

ii) Continuation of topic/Supporting

In the instance below the interviewer overlaps to support the candidate by continuing the topic and supplying a word that matched what the candidate appeared to be looking for:

C: so (.) that case Tanzanian people are very kind but not very friendly, I can say that in Ken Kenya the peoples is very friendly but ah: how can I say they try to sell many things (.) [for tourists,]

* *I: [they are more commercialised,]*

C: Commercialised yeah.

I: I see.

C: Yeah it looks they are just interested in our money.

Interviewer 2 (female)/Candidate 3 (female): 107-114

Negative

i) Attempt to control topic

The interviewer in this example below attempts to introduce the idea of unemployment. Firstly she refers to people losing jobs and then when the student continues by talking about government protection of industry the interviewer overlaps with the question of the unemployment rate. Following the overlap she then reiterates the question thus reinforcing the direction she wants the discussion to go:

² In each of the examples * indicates the focus of analysis, I = Interviewer and C = Candidate

C: Um (0.5) because I want to I think it's OK to trade with (.) to trade with another country. Because I think labour in Thailand have trend have trend to ah expensive in the future, yeah. so if we use (0.5) if we use ah: (.) not no (.) we can ah we can import something from the other countries?

I: Mm,

C: which will cheaper than in my country in the future=

I: =but will that help your country develop? If if people lose jobs? Because you traded from outside?

C: Um I think it doesn't matter because my government will have a policy to protect (.) um (.) some industrial in Thailand. Yes. Same [Australia (.) in here,]

* I: [what's the unemployment] what's the unemployment rate in Thailand.

C: Unemployed?

Interviewer 4 (female)/Candidate 5 (male):269-281

These examples illustrate how overlapping was used both 'positively' and 'negatively' by interviewers in the course of the interviews.

A quantitative analysis of interviewer overlaps is reported below.

The total number of interviewer overlaps across all 32 interviews was only 79. This indicates a fairly sparing use of overlaps by individual interviewers.

Table 2 below shows female interviewer overlaps in the data. Column 1 indicates whether the overlaps were positive or negative. Column 2 shows the number of interviews where overlaps were found and column 3 the instances of overlaps within those interviews. Both these figures are for the female interviewer and male candidate pairing. Column 4 indicates the number of interviews with female candidates where overlaps were employed and column 5 the number of instances of use found therein. (The total number of interviews for each gender pairing is eight)

	Interviewer Female Candidate Male		Interviewer Female Candidate Female	
	No. of Interviews	No. of Overlaps	No. of Interviews	No. of Overlaps
Positive Overlap	7	24	5	22
Negative Overlap	1	1	1	1
Total	8	25	6	23

Table 2: Instances of female interviewer overlaps

Table 3 summarises the incidence of male interviewer overlaps in the data. Again, column 1 indicates whether the overlaps were positive or negative. Columns 2 and 3 show the male interviewer and male candidate pairing: the number of interviews in which overlaps were observed and the number of overlaps across the number of interviews indicated. Similarly, columns 4 and 5 show the number of interviews in which overlaps were observed in the pairing of male interviewers and female candidates, and also the number of overlaps found within these.

	Interviewer Male Candidate Male		Interviewer Male Candidate Female	
	No. of Interviews	No. of Overlaps	No. of Interviews	No. of Overlaps
Positive Overlap	5	7	7	24
Negative Overlap	0	0	0	0
Total	5	7	7	24

Table 3 Instances of male interviewer overlaps

A Chi-square analysis, using a 2x2 contingency table, was used to test the null hypothesis, (H_0), that there was no relation between interviewer gender and candidate gender.

Result:

Interviewer overlaps ('positive' & 'negative' combined): $\chi^2=5.72$, $df=1$, $p<0.05$

Interviewer 'positive overlaps': $\chi^2=5.58$, $df=1$, $p<0.05$

The reason for this result seems to be that the total number of 'positive overlaps' by male interviewers with male candidates is clearly lower than the other three pairings (refer to Tables 2 and 3). However, this result should be viewed cautiously in light of the very limited use of interviewer overlaps in the interviews overall.

3.1.1 B Candidate overlaps

Examples of candidate overlaps are included below.

Positive:

i) Giving confirmation

In this example of giving confirmation the candidate is affirming the interviewer's response that her home town Nagoya is well known:

I: But everybody knows the name. [Nagoya]
* *C: [Oh yeah:] Mm hm I hope so.*
Interviewer 1 (male)/Candidate 3 (female):21-22

ii) Continuation of topic/Supporting

Here the candidate (female) is supporting the interviewer's idea of shopping at the local supermarket and continuing this by offering examples of names of supermarkets:

I: And and the supermarket um (0.5) the local supermarket a good idea to buy [buy food?]
* *C: [Mm: like Coles?] Target,*
I: Coles Target Safeway,
Interviewer 4 (female)/Candidate 8 (female): 162-165

In this example a male candidate is developing the topic of moving from a homestay to a shared house:

- I: Are you in a homestay?*
C: No I'm living share house.
I: Oh [yeah well that's very cheap,]
 * *C: [Yesterday] I moving.*
I: Oh you moved yesterday.
C: From from homestay.

Interviewer 4 (female)/Candidate 7 (male): 131-135

These examples demonstrate the ways in which both female and male candidates used overlaps for 'positive' ends i.e., to confirm information and for topic development with their interviewers.

A quantitative analysis of candidate overlaps was also undertaken and the results reported below.

The total number of candidate overlaps across all 32 interviews was only 77. As with the interviewers, this indicates a fairly sparing use of overlaps by individual candidates.

Table 4 shows the type of overlap, the number of interviews and the number of overlaps in those interviews for the gender pairings female candidate/male interviewer and female candidate/female interviewer.

	Candidate Female Interviewer Male		Candidate Female Interviewer Female	
	No. of Interviews	No. of Overlaps	No. of Interviews	No. of Overlaps
Positive Overlap	5	33	8	22
Negative Overlap	0	0	0	0
Total	5	33	8	22

Table 4 Instances of female candidate overlaps

Table 5 provides the same information for the gender pairings with male candidates.

	Candidate Male Interviewer Male		Candidate Male Interviewer Female	
	No. of Interviews	No. of Overlaps	No. of Interviews	No. of Overlaps
Positive Overlap	5	13	6	11
Negative Overlap	0	0	0	0
Total	5	13	6	11

Table 5 Instances of male candidate overlaps

A Chi-square analysis, using a 2x2 contingency table, was used to test the null hypothesis, (H₀), that there was no relation between candidate gender and interviewer gender.

As shown in both Tables 4 and 5 male candidates used fewer overlaps with both male and female interviewers. Again however, this result should be viewed with caution in light of the limited use of candidate overlaps in the interviews overall.

3.1.2 Interruptions

"Interruptions on the other hand," Coates (1993:109) states, "are violations of the turn-taking rules of conversation. Next speaker begins to speak while current speaker is still speaking, at a point in current speaker's turn which could not be defined as the last word. Interruptions break the symmetry of the conversational model; the interruptor prevents the speaker from finishing their turn, at the same time gaining a turn for themselves."

Again Coates (1993) cites studies that observed few interruptions in same-sex pairs: where men rarely interrupt one another. She explains that 46 out of 48 observed interruptions in one study were performed by males in mixed-sex pairs.

Interruptions were employed in these 32 IELTS interviews seven times by interviewers and 17 times by candidates. It was also observed that these interruptions were actually functioning in a positive way by assisting in topic development and providing confirmation to support the interlocutor's understanding.

3.1.2 A Interviewer interruptions

Examples of interviewer interruptions are shown below.

i) Developing Topic

In this example we see the interviewer interrupting to take up and develop the first response given by the candidate:

- C: Yeah. Firstly I would like to improve my English because I think it's important for me to (.) to study English [and also,]*
- * *I: [Why?] Why do you think you need English?*
- Interviewer 1 (male)/Candidate 2 (male):lines 245-248**

ii) Attempt to Control Topic

In this instance the interviewer attempts to control the topic by re-directing the candidate away from discussing her husband's training in environmental protection to the broader area of environmental issues in south-east Asia and the whole world:

C: Ah but after he graduate he plan he has a plan to have a small company like a consulting company,

I: Right,

C: Yeah. Because will popular I think we hope because the government will launch a new policy new law for the environment,

I: Right,

C: To protect the our environment,

I: Yes,

C: [So I think it very good next to train as,]

* *I: [Yes that that's um a very (.) significant area isn't it.]*

C: Mm. Mm.

I: Well for all south-east,

C: Mm.:

I: Well for the whole world

C: Mm.

Interviewer 4 (female)/Candidate 7 (female):194-208

In these examples both female and male interviewers encouraged topic development in the candidates.

The quantitative analysis of the number of interruptions used by interviewers is reported below.

Table 6 below shows female interviewer interruptions. Column 1 introduces the interruptions. Columns 2 and 3 show the number of interviews in which they were observed (out of the possible eight for each gender pairing), and how many interruptions were found in those interviews between female interviewers and male candidates. Columns 4 and 5 provide the same information for the gender pairings of female interviewer and female candidate.

	Interviewer Female Candidate Male		Interviewer Female Candidate Female	
	No. of Interviews	No of Interruptions	No. of Interviews	No of Interruptions
Total	2	2	1	1

Table 6: Instances of female interviewer interruptions

Table 7 gives the same information of the number of interviews in which interruptions were observed, the number of interruptions found therein for each of the pairings of male interviewer with male candidate and male interviewer with female candidate.

	Interviewer Male Candidate Male		Interviewer Male Candidate Female	
	No. of Interviews	No. of Interruptions	No. of Interviews	No. of Interruptions
Total	1	4	0	0

Table 7: Instances of male interviewer interruptions

A 2x2 contingency table was established based on this data. However, the expected frequencies were too small to carry out a Chi-square analysis.

3.1.2 B Candidate Interruptions

Examples of candidate interruptions are included below.

i) Continuing Topic

In this example the interruption continues the topic of Melbourne's very unpredictable weather referred to in the interviewers' previous turn:

I: Because you know what Melbourne's like? Huh? =

C: =Yeah. (laughs)

I: Always very unpredictable? Um so just listen carefully to the forecast, =

C: =Mm hm, =

I: =And then take the right stuff for this [kind of weather.]

* *C: [Because sometimes] we can't believe them.*

I: That's right.

Interviewer 3 (male)/Candidate 8 (female):lines 232-240

ii) Confirming

In this instance the candidate is confirming for the interviewer that he has understood correctly what the candidate had previously explained:

I: How about the river. Do they use the river much for (.) local transportation for travel?

C: Yeah,(.) they always use ah (1.0) my ah my they always use ah boat but (.) it doesn't have much way to go to (.) it doesn't have much way to connect with another part,

I: Mm hm,

C: so if sometimes they use a boat (.) and (.) and then they use bus,

I: Mm hm,

C: to continue um (2.0) their (.) to continue to go to work,

I: Mm. So part of the journey [by boat]

* *C: [yes]*

I: then catch the OK. So do they have river taxis? Can you (.) catch a small boat just to go quickly across the river? Or

C: Mm: doesn't have private but have a (?) you go together with another person.

I: Mm.

C: Yeah.

Interviewer 3 (male)/Candidate 5 (male):137-153

These examples show how both female and male candidates used interruptions to express positive responses to their interviewer' turns.

The quantitative analysis of candidate interruptions is given below.

Table 8 shows the number of interviews in which interruptions occurred and the number of interruptions within those interviews between female candidates and male interviewers, and female candidates and female interviewers.

	Candidate Female Interviewer Male		Candidate Female Interviewer Female	
	No. of Interviews	No of Interruptions	No. of Interviews	No of Interruptions
Total	3	7	2	3

Table 8: Instances of female candidate interruptions

Table 9 provides the same information for the gender pairings of male candidate with male interviewer and male candidate with female interviewer.

	Candidate Male Interviewer Male		Candidate Male Interviewer Female	
	No. of Interviews	No of Interruptions	No. of Interviews	No of Interruptions
Total	1	2	3	5

Table 9: Instances of male candidate interruptions

Again, a 2x2 contingency table was established based on this data. However, the expected frequencies were too small to carry out a Chi-square analysis.

3.1.3 Minimal Responses

Coates (1993:109) describes Minimal responses such as *yeah* and *mhm* as not constituting a turn. "...they are a way of indicating the listener's positive attention to the speaker, and thus a way of supporting the speaker in their choice of topic (Coates 1993:109)".

For Coates then, Minimal responses (MRs) are a way of indicating the listener's positive attention; a listener, therefore, has an active not a passive role. She also found research to be unanimous in showing that women use Minimal responses more than men and at more appropriate moments.

In the IELTS interview data analysed here, Minimal responses appeared to be used for encouraging the interlocutor to continue and supporting them by providing a signal to show active listening. There were many more instances of Minimal responses throughout the 32 interviews than there were of either overlaps or interruptions. Interviewers' Minimal responses totalled 805 and candidate's totalled 291. No delays in Minimal responses were detected.

3.1.3 A Interviewer Minimal responses

Examples of interviewer Minimal responses are given below.

i) Encouraging Continuation

Here a male interviewer is encouraging the candidate to continue by displaying his positive attention through the use of Minimal responses:

- I: Mm. And is it also the case that it's important that the other people in the family help each other ah when there is a problem with not earning enough money?*
- C: Ah, (.) I I think this is the important from this time.*
- * *I: Mm hm,*
- C: Because ah in ah in ah in ah Bangkok big city in Thailand,*
- * *I: Mm,*
- C: The capital of Thailand,*
- * *I: Mm,*
- C: And now (?) not have a lot of job.*
- * *I: Mm,*
- C: So when the people when they come back to their city so they will be help another people,*
- * *I: Mm hm,*
- C: Around them.*
- I: Mm, mm, And ah I mean I've been in Bangkok and I wasn't aware that there were many people begging or having to ask for money and so on, ah is this becoming more of a problem now because of unemployment? And the problems with the economy?*
- C: Ah I think this is the problem of economic.*

Interviewer 3 (male)/Candidate 6 (male):lines 100-119

The next example shows more a female interviewer using Minimal responses to encourage the candidate to continue the idea she is trying to express:

- C: Ah I'm marketing supervisor,*
- * *I: Mm hm,*
- C: Also I still work hard. Everyday busy,*
- * *I: Mm hm,*
- C: Because I have a analyst analyst team? And ah (.) investigate (?) marketing information and I should do I should start I should do project and ah supervise the project how the progress,*
- * *I: Mm.,*
- C: And how affect in this project,*
- * *I: Mm.,*
- C: And I feel stress and too busy and ah no too much time for holiday yeah so I cracked,*
- I: Not too much free time.*

Interviewer 6 (female)/Candidate 11 (female):49-61

ii) Supporting

In this example the interviewer is supporting the candidate's development of the topic by employing the MR *right*:

- * C: *Mm I hope ah: (0.5) become teacher,*
I: *Right,*
C: *In high school,*
- * I: *Right,*
C: *High school or college,*

Interviewer 5 (male)/Candidate 11 (female):278-282

Alternatively, in the example below the interviewer uses the MR *mm* to support the candidate's development of the topic:

- * C: *Um yeah actually I really wanted to study about film,*
I: *Mm.:*
C: *But ah now my parents support me,*
- * I: *Mm.:*
C: *So if I insis insisted on studying movie,*
- * I: *Mm.:*
C: *Maybe they said absolutely no.*

Interviewer 8 (female)/Candidate 13 (male):205-211

These examples indicate that neither female nor male interviewers seemed to use Minimal responses differently.

The quantitative analysis of interviewer Minimal responses is reported below.

Table 10 presents the figures for the use of Minimal responses by female interviewers. Columns 2 and 3 show the number of interviews in which Minimal responses were found (out of the possible 8 for each gender pairing) and the number of Minimal responses within those. Columns 4 and 5 provide this information for the gender pairing of female interviewer and female candidate.

	Interviewer Female Candidate Male		Interviewer Female Candidate Female	
	No. of Interviews	No. of MRs	No. of Interviews	No. of MRs
Total	8	199	8	169

Table 10: Instances of female interviewer Minimal responses

Table 11 below provides the parallel information for the male interviewers.

	Interviewer Male Candidate Male		Interviewer Male Candidate Female	
	No. of Interviews	No. of MRs	No. of Interviews	No. of MRs
Total	81	204	8	233

Table 11: Instances of male interviewer Minimal responses

A Chi-square analysis, using a 2x2 contingency table, was used to test the null hypothesis, (H_0), that there was no relation between interviewer gender and candidate gender.

Result:

Interviewer Minimal responses: $\chi^2 = 4.09$, $df = 1$, $p < 0.05$

The reason for this result seems to be that the total of Minimal responses used by female interviewers with female candidates is clearly lower than the other three pairs particularly, male interviewers with female candidates.

3.1.3 B Candidate Minimal Responses

Examples of Candidates Minimal Responses are given below.

i) Encouraging Continuation

In this example the candidate signals that he is listening to the interviewer and is encouraging him to continue with the question:

I: Ah ha oh I see. And um (0.5) when you go back to to Bangkok and you you work for a few years,

* *C: yeah,*

I: In this sort of area, (0.5) what do you think ah will be the result of your work. Do you think that ah Bangkok needs a lot of construction? A lot of industrial (.) work?

Interviewer 1 (male)/Candidate 1 (male):lines 206-211

The next example shows a female candidate using Minimal responses for the same purpose:

I: OK if you ah if you ah find that with your Australian qualification it's rather difficult to get a job in Japan,

* *C: Mm hm,*

I: because as your mother said it might be easier with a Japanese qualification,

* *C: Mm,*

I: Ah have you thought about other possible careers? As well as a career in hospitality? Are there other things you could do with your qualifications.

Interviewer 5 (male)/Candidate 12 (female):184-192

ii) Showing Attention

Here the candidate is using Minimal responses to indicate that he is paying attention to the information being provided by the interviewer:

I: =Ok alright. Ah (0.5) I think motel will be cheaper than a hotel.(0.5) so if you're looking for the cheapest form then choose the motel accommodation, but if you want something even cheaper than motel, then you should look for hostel accommodation.

* *C: Mm.*

I: In the city (.) you'll find several hostels for example the YWCA? And they offer you rooms as well as dormitories.

* *C: Mm,*

I: So you have a choice there, or you have the Miami hostel (.) where a lot of students stay but tourists can stay there too. You will have to look at the Yellow Pages, =

C: =Yeah=

Interviewer 2 (female)/Candidate 2 (male):170-182

iii) Supporting

In this example the candidate uses Minimal responses to show that she is listening and to provide support for the information being given by the interviewer:

- I: Well Japanese is usually expensive.*
C: Yeah I think so,
I: Um there is a nice Japanese restaurant in the city,
 * *C: Yeah,*
I: At the top of mm: (1.0) off the main at the top of Bourke Street you might know it.
C: No I don't know it,
I: Anyway that's about,
 * *C: Yeah,*
I: That's one of the you know for value,
 * *C: Oh:?*
I: That's probably the cheapest Japanese restaurant?
 * *C: Yeah,*
I: But if you want Japanese you have to pay in Australia.

Interviewer 8 (female)/Candidate 15 (female):134-147

These examples show that both female and male candidates seemed to use Minimal responses to support, encourage and express interest in their interviewer in similar ways.

The quantitative analysis of candidate Minimal responses is given below.

All but one male candidate employed Minimal responses, although not to the extent used by the interviewers. This is to be expected given their respective roles. Table 12 indicates the female candidates use of Minimal responses. Columns 2 and 3 show the number of interviews and the number of Minimal responses within those for the gender pairing female candidate and male interviewer. (There were eight of each gender pairing). Columns 4 and 5 show this information for the interviews in which there were female candidates and female interviewers.

	Candidate Female Interviewer Male		Candidate Female Interviewer Female	
	No. of Interviews	No. of MRs	No. of Interviews	No. of MRs
Total	81	48	8	115

Table 12: Instances of female candidate Minimal responses

Table 13 expresses the same information for the male candidates and their use of Minimal responses.

	Candidate Male Interviewer Male		Candidate Male Interviewer Female	
	No. of Interviews	No. of MR	No. of Interviews	No. of MRs
Total	8	71	7	57

Table 13: Instances of male candidate Minimal responses

A Chi-square analysis, using a 2x2 contingency table, was used to test the null hypothesis, (H_0), that there was no relation between candidate gender and interviewer gender.

Result:

Candidate Minimal responses: $\chi^2=19.03$, $df=1$, $p<0.001$

The reason for this result seems to be that female candidates used a lot more Minimal responses with female interviewers than any of the other three pairings, especially female candidates with male interviewers.

Overall, the results of the Chi-square analysis for interviewer and candidate Minimal responses do not reveal any clear gender pattern.

3.1.4 Summary

In this section the results for the three features are summarised and briefly discussed.

The results indicated that male and female interviewers used about the same number of overlaps except for when male interviewers were paired with male candidates in which case the number was smaller. On the other hand, female candidates produced a larger number of overlaps than male candidates irrespective of the gender of their interlocutor. Therefore there is no consistent gender pattern across interviewers and candidates. However, these results should be viewed cautiously because of the very limited use of overlaps in the interviews overall.

Given the low incidence of interruptions used across 32 interviews it was not possible to draw any clear conclusions about the impact of gender on the IELTS oral interview from this perspective.

Female and male interviewers employed Minimal responses more than either the female or male candidates. This is probably because of the role of the interviewer in facilitating the candidates' discussions. The Chi-square analysis for both interviewers' and candidates' use of Minimal responses suggested there was a significant relation between interviewer and candidate gender. However, post-hoc inspection of the data showed that this relationship was not the same in the two analyses. Thus, there was no consistent gender pattern in the use of Minimal responses by interviewers and candidates.

While the frequency of use showed no clear gender patterns, these discourse features were used by all participants in similar ways irrespective of gender. Through their use of overlaps, interruptions and Minimal responses they sought to provide confirmation of ideas, to encourage continuation or development of a topic or question, to express support for their interlocutors contributions and to indicate active attention to the interlocutor. The few instances where interviewers attempted to control the topic were still within the bounds of encouraging the candidate to develop the topic in another way; they were not trying to seize their turn.

On the basis of these findings, it would seem that interviewers and candidates generally adopted a more collaborative, co-operative and supportive communicative style irrespective of their gender or the gender of their interlocutor. Both participants appeared to understand that a co-operative dialogue would provide the best situation for the candidate to achieve the best possible result.

Having explored the impact of the gender of participants on the discourse produced in the IELTS oral interview, the second question the study aimed to address was the impact of gender on the rating of the interview.

3.2 Test Scores

The primary focus of the analysis of test scores which follows is on the scores of the raters who assessed the audio-recordings of the interviews rather than the original interviewers' scores for several reasons. Firstly, this analysis provides a more controlled investigation into possible gender bias in scoring since it is based on comparisons of four different ratings (two female and two male) of every interview (N =32). Each interview, however, was only scored by one interviewer and it would therefore be extremely difficult to make meaningful comparisons of interviewers in respect to their scoring and possible gender bias. Furthermore, as outlined in Section 2.2 above, by using a mixed design whereby each interview was assessed by different combinations of male and female raters drawn from a pool of eight females and eight males, raters could be calibrated against each other in relation to their potential gender bias. Due to practical problems in conducting the interviews such a design could not be employed for the interviews thus disallowing this kind of intra-group comparison of their ratings. In any case, each interviewer only rated the four interviews they conducted which would provide insufficient evidence on which to make claims about any possible gender bias.

The band scores assigned to candidates by their interviewers and the other raters in each of the 32 interviews together with information about the gender of both candidate and interviewer are given below in Table 14.

It is interesting to note in this table that the interviewers used a more restricted range of band scores (ie. 5 - 7) than the other raters (ie. 4 - 8). In addition, in 24 out of the total of 32 cases the interviewer was more lenient than the average rater score. This may be due to a method difference in the way the assessments were carried out ie. in the live face-to-face context by the interviewers on the basis of audio-recordings by the raters. It may also be the case that interviewers are less harsh in their scoring because of their personal engagement with the candidate in the course of the interview.

The scores of the raters who assessed the audio-recordings of the interviews were examined using a facility of the multi-faceted Rasch computer program FACETS (Linacre, 1989-1995), known as *bias analysis*.

Bias analysis in multi-faceted Rasch measurement identifies unexpected but consistent patterns of behaviour which may occur from an interaction of a particular rater or group of raters with respect to some component or 'facet' of the rating situation such as, in this study, candidate gender or interviewer gender. The output of these analyses shows firstly, whether individual raters are scoring say, candidates of one gender significantly more harshly or leniently than candidates of the other gender, and secondly, whether they are behaving consistently towards candidates of each gender. These analyses therefore assist in identifying potential important sources of measurement error in the rating process. Multi-faceted Rasch-based bias analysis has been used in a number of recent studies for this purpose (see, for example, Wigglesworth (1993), McNamara (1996) and O'Loughlin (1997)). Bias analysis is used in the current study to investigate the impact of candidate and rater gender on the reliability of test scores.

Candidate		Interviewer			Raters' Score			
Number	Gender	Number	Gender	Score	Female		Male	
1	Male	1	Male	5	5	6	6	5
1	Male	2	Female	6	6	5	5	4
2	Male	1	Male	6	6	7	6	6
2	Male	2	Female	7	6	6	6	7
3	Female	1	Male	7	7	7	8	7
3	Female	2	Female	7	6	6	7	7
4	Female	1	Male	5	6	5	5	5
4	Female	2	Female	7	6	5	5	6
5	Male	3	Male	6	5	4	5	5
5	Male	4	Female	5	6	5	5	4
6	Male	3	Male	6	5	6	5	5
6	Male	4	Female	6	5	4	5	5
7	Female	3	Male	7	6	6	6	6
7	Female	4	Female	6	6	5	6	5
8	Female	3	Male	6	6	5	7	5
8	Female	4	Female	6	5	5	5	6
9	Male	5	Male	6	4	4	5	4
9	Male	6	Female	5	4	5	5	4
10	Male	5	Male	6	5	6	5	6
10	Male	6	Female	6	6	5	6	5
11	Female	5	Male	5	5	5	5	3
11	Female	6	Female	6	5	4	6	5
12	Female	5	Male	5	5	4	5	4
12	Female	6	Female	6	5	4	4	4
13	Male	7	Male	6	5	6	5	6
13	Male	8	Female	6	7	5	5	5
14	Male	7	Male	6	5	5	5	6
14	Male	8	Female	7	6	6	7	5
15	Female	7	Male	6	5	6	5	7
15	Female	8	Female	5	7	5	5	6
16	Female	7	Male	6	5	6	6	5
16	Female	8	Female	7	5	6	6	6

Table 14: Summary of results

Question 1: Is there a significant interaction between raters' scoring and candidate gender?

The first issue to be examined here involves the interaction between raters' scores and candidate gender. In other words, do raters score candidates of either gender significantly more harshly than the other? The output from the bias analysis conducted here provides detailed information about individual raters in relation to this question.

Table 15 (page 25) shows the output from the bias analysis. In this table Column 1 provides the rater identity number and Column 2 candidate gender. Column 3 provides the total observed score of each rater for female and male candidates respectively, while Column 4 shows each rater's total expected score for the two versions. Column 5 shows the number of ratings given by the rater to candidates of the specified gender. Column 6 then provides the average difference between the expected and observed score. A bias logit, based on this difference, is then calculated together with its standard error (columns 7 and 8). The bias score is then converted into a standardised Z-score by dividing it by its standard error (column 9). The Z-score values are the most revealing figures in this analysis. Where the Z-score values fall between -2.0 and +2.0, the rater may be considered to be scoring candidates from the specified gender without significant bias. Where the value falls below -2.0 the rater is marking candidates from the specified gender significantly more leniently than the other gender. On the other hand, where the value is greater than +2.0 the rater is scoring candidates of the specified gender significantly more harshly compared to the way that rater treats the other gender. Furthermore, in this analysis the infit mean square value (column 10) indicates how *similar* the rater's scoring is for the specified gender overall. Where the value is less than 0.7 the rater's scoring for candidates of that gender lacks variation, ie. it is too similar. Conversely, where this value is greater than 1.3 the rater's scoring tends to be inconsistent for the given gender.

Since all of the Z-scores are within the range of -2 to +2 it can be concluded that none of the raters are significantly biased in favour of candidates of either gender. The infit mean square values, however, suggest that there are a number of raters whose scoring is too similar for female candidates overall ie. raters 1, 2, 4, 5, 6, 9 and 15 and for male candidates overall ie. raters 2, 5, 6, 7, 8, 9, 10, 11, and 16. This is perhaps not surprising given that the whole group of raters only assigned band scores between 4 and 8 and mostly, 5, 6 or 7. Conversely, raters 7, 8, 10, 11 and 16 show a significant tendency to be inconsistent in their scoring of female candidates and raters 1 and 12 in their scoring of male candidates. However, given the relatively small number of assessments carried out by each rater for candidates of either gender (N=2-6) this trend towards inconsistency should be regarded with a degree of caution.

Rater ID	Cand gender	Observed score	Expected score	Observed Count	Obs-Exp score	Bias (logit)	Error	Z score	Infit mn sq
1	Female	22	23.0	4	-0.25	0.34	0.59	0.6	0.4
1	Male	23	22.0	4	0.25	-0.34	0.58	-0.6	1.6
2	Female	22	21.5	4	0.13	-0.19	0.59	-0.3	0.4
2	Male	20	20.6	4	-0.14	0.23	0.64	0.4	0.0
3	Female	12	11.6	2	0.19	-0.25	0.81	-0.3	1.3
3	Male	33	33.4	6	-0.06	0.09	0.49	0.2	0.8
4	Female	32	33.4	6	-0.23	0.33	0.50	0.7	0.3
4	Male	12	10.7	2	0.67	-0.92	0.81	-1.1	1.3
5	Female	11	11.4	2	-0.18	0.25	0.84	0.3	0.4
5	Male	33	32.6	6	0.06	-0.09	0.49	-0.2	0.4
6	Female	21	22.5	4	-0.37	0.54	0.62	0.9	1.0
6	Male	23	21.5	4	0.31	0.51	0.58	-0.9	0.3
7	Female	27	25.4	5	-0.54	0.48	0.54	-0.9	1.5
7	Male	13	14.6	3	0.19	0.93	0.76	1.2	0.4
8	Female	27	26.1	5	-0.33	0.28	0.54	-0.5	1.5
8	Male	14	15.0	3	0.25	0.55	0.75	0.7	0.4
9	Female	24	23.0	4	-0.25	0.33	0.57	-0.6	0.6
9	Male	21	22.0	4	-0.25	0.37	0.62	0.6	0.3
10	Female	26	23.5	4	0.62	0.82	0.58	-1.4	1.7
10	Male	20	22.5	4	-0.62	0.93	0.64	1.5	0.0
11	Female	21	20.4	4	0.14	0.21	0.62	-0.3	1.8
11	Male	19	19.6	4	-0.15	0.25	0.65	0.4	0.3
12	Female	25	23.5	4	0.37	0.49	0.57	-0.9	0.9
12	Male	21	22.5	4	-0.37	0.54	0.62	0.9	1.8
13	Female	20	20.4	4	-0.11	0.18	0.64	0.3	0.8
13	Male	20	19.6	4	0.10	0.16	0.64	-0.2	0.8
14	Female	20	21.0	4	-0.24	0.38	0.64	0.6	0.8
14	Male	21	20.1	4	0.23	0.36	0.62	-0.6	1.0
15	Female	23	23.5	4	-0.13	0.17	0.58	0.3	0.3
15	Male	23	22.5	4	0.13	0.17	0.58	-0.3	0.9
16	Female	19	22.0	4	-0.74	0.17	0.65	1.8	2.0
16	Male	24	21.1	4	0.74	0.02	0.57	-1.8	0.6

Table 15: Bias calibration report, rater - candidate gender interaction

Question 2: Is there a significant interaction between candidates' scores and rater gender?

The second bias analysis examined whether there was a significant interaction between candidate scores and rater gender. In other words, the issue here is whether raters of one gender scored candidates significantly more harshly than raters of the other gender. The results of this bias analysis are shown in Table 16.

Cand ID	Rater gender	Observed score	Expected score	Observed count	Obs-Exp score	Bias (logit)	Error	Z score	Infit mn sq
1	Female	22	20.9	4	0.26	-0.71	0.79	-0.9	0.6
2	Female	25	24.9	4	0.02	-0.05	0.74	-0.1	0.4
3	Female	26	27.4	4	-0.35	0.91	0.77	1.2	0.6
4	Female	22	21.4	4	0.14	-0.37	0.79	-0.5	0.6
5	Female	20	19.5	4	0.13	-0.43	0.90	-0.5	1.6
6	Female	20	20.0	4	0.01	-0.03	0.90	0.0	1.6
7	Female	23	22.9	4	0.02	-0.05	0.75	-0.1	0.4
8	Female	21	21.9	4	-0.23	0.62	0.85	0.7	0.5
9	Female	17	17.5	4	-0.12	0.32	0.82	0.4	0.5
10	Female	22	22.4	4	-0.10	0.25	0.79	0.3	0.6
11	Female	19	19.0	4	0.01	-0.02	0.87	0.0	0.6
12	Female	18	17.5	4	0.13	-0.36	0.83	-0.4	0.7
13	Female	23	21.9	4	0.27	-0.63	0.75	-0.8	1.5
14	Female	22	22.4	4	-0.10	0.25	0.79	0.3	0.6
15	Female	23	22.9	4	0.02	-0.05	0.75	-0.1	1.5
16	Female	22	22.4	4	-0.10	0.25	0.79	0.3	0.6
1	Male	20	21.1	4	-0.26	0.81	0.90	0.9	1.6
2	Male	25	25.0	4	-0.01	0.03	0.75	0.0	0.4
3	Male	29	27.5	4	0.37	-1.10	0.88	-1.3	0.6
4	Male	21	21.6	4	-0.14	0.38	0.85	0.4	0.5
5	Male	19	19.6	4	-0.14	0.45	0.87	0.5	0.6
6	Male	20	20.1	4	-0.02	0.05	0.90	0.1	0.0
7	Male	23	23.1	4	-0.01	0.03	0.75	0.0	0.4
8	Male	23	22.1	4	0.24	-0.55	0.75	0.7	1.5
9	Male	18	17.6	4	0.10	-0.28	0.83	0.3	0.7
10	Male	23	22.6	4	0.11	-0.25	0.75	0.3	1.5
11	Male	19	19.1	4	-0.02	0.06	0.87	0.1	3.6
12	Male	17	17.6	4	-0.15	0.40	0.82	0.5	0.5
13	Male	21	22.1	4	-0.26	0.70	0.85	0.8	0.5
14	Male	23	22.6	4	0.11	-0.25	0.75	0.3	1.5
15	Male	23	23.1	4	-0.01	0.03	0.75	0.0	1.5
16	Male	23	22.6	4	0.11	-0.25	0.75	-0.3	0.4

Table 16: Bias calibration report, candidate - rater gender interaction

The results indicate that none of the candidates were treated significantly more harshly by raters of either gender since all Z-scores fell within the range of -2 and +2. Once again, however, the infit mean square values indicate a tendency for raters to assess candidates either too similarly or too inconsistently. On the one hand, the scoring of female raters as a group for candidates 1, 2, 3, 4, 7, 8, 9, 10, 11, 14 and 16 and male raters as a group for candidates 2, 3, 4, 5, 6, 7, 12, 13 and 16 was too similar. As in the previous analysis, this is not unexpected given that the whole group of raters only assigned band scores between 4 and 8 and mostly, 5, 6 or 7. On the other hand, the scoring of female raters as a group for candidates 5, 6, 13 and 15 and male raters as a group for candidates 1, 8, 10, 11, 14, and 15 showed too much variability. Again, given the relatively small number of assessments carried out for each candidate by raters of either gender group (N=4) this trend towards inconsistency should also be regarded with a degree of caution.

Question 3: Is there a significant interaction between candidate gender and rater gender?

The third bias analysis examined whether there is a significant interaction between candidate gender and rater gender. Table 17 below summarises the output from this analysis:

Cand gender	Rater gender	Observed score	Expected score	Observed count	Obs- Exp Score	Bias (logit)	Error	Z score	Infit mn sq
Female	Female	174	175.4	32	-0.04	0.06	0.20	0.3	0.8
Male	Female	171	169.7	32	0.04	-0.06	0.20	-0.3	0.9
Female	Male	178	176.4	32	0.05	-0.06	0.20	-0.3	1.4
Male	Male	169	170.6	32	-0.05	0.07	0.21	0.3	0.9

Table 17: *Bias calibration report, candidate gender-rater gender interaction*

The Z score values indicate that the interaction between candidate gender and rater gender is not significant ie. candidate scores are not significantly affected by whether their rater is of the same or opposite sex. Furthermore, in only one gender combination ie. male raters with female candidates, is the infit mean square value outside the acceptable range: in this case a value of 1.4 suggests that male raters tended to score female candidates less consistently than the three other gender pairings overall.

From the above analyses it appears that the impact of both candidate and rater gender on test scores in the IELTS oral interview is not significant. However, this conclusion should be regarded with some caution given the relatively small data set available for analysis. Furthermore, the findings here do not imply that the measurement process can be considered flawless: it could still be true that certain candidates are rated significantly more harshly or leniently by individual raters compared to the way that rater treats other candidates irrespective of candidate or rater gender.

Question 4: Is there a significant interaction between individual candidates and raters?

The final bias analysis therefore examined whether there were any candidates who were treated by a particular rater significantly more or less harshly than that rater would treat other candidates. The output from this analysis revealed there were only two such occurrences. Table 18 below provides the output from the analysis for these cases.

Cand ID	Rater ID	Observed Score	Expected score	Observed count	Obs - Exp score	Bias (logit)	Error	Z score	Infit mn sq
1	12	4	5.3	1	-1.3	5.07	1.92	2.6	0.0
11	16	3	4.7	1	-1.71	4.47	1.93	2.3	0.7

Table 18: *Bias calibration report, significant interactions between candidates and raters.*

In both instances, since the Z score values are greater than +2, the raters are marking the specified candidate significantly more harshly than they would other candidates. Considering there were only two such occurrences out of a total of 128 ratings it can be concluded that there was a high degree of intra-rater reliability in this study.

4.0 Conclusions

The study set out to investigate whether there was an effect for gender on either the discourse of the oral interview and/or the rating process. While there were some gender differences between female and male interviewers and candidates these did not form a consistent gender pattern. The majority of candidates and, more particularly, interviewers generally adopted a collaborative, co-operative and supportive communicative style irrespective of their own gender or the gender of their interlocutor. Such a style is therefore clearly not exclusively the province of female speakers, at least in the testing context. It would seem that, in most cases, both participants understand that a co-operative dialogue will produce the best possible speech sample and thus best possible outcome for the candidate. In this sense, it can be concluded that the IELTS oral interview is not a strongly gender differentiated event.

Perhaps even more importantly, the analysis of test scores indicated there was no evidence of significant bias in the rating process in relation to the gender of raters or candidates.

Both sets of findings therefore suggest that gender does not have a significant impact on the IELTS oral interview and that it is therefore unnecessary to allow the gender of the candidate to determine the gender of the interviewer/rater.

Acknowledgments

I wish to thank IELTS Australia for the research grant which enabled me to undertake the study, the language centre in Melbourne which provided the facilities for the data collection, as well as the candidates, interviewers and raters who participated in the study. I am also extremely grateful to Jeanette Carter for her assistance in many aspects of this research project.

References

- Brown, A. and Hill, K. (1996) *Interviewer style and candidate performance in the IELTS oral interview. Report submitted to IELTS Australia.*
- Coates, J. (1993) *Women, men and language.* London: Longman (second edition).
- Freed, A. F. (1995) Language and gender. *Annual review of Applied Linguistics* 15, 3-22.
- Freed, A. F. & Greenwood, A. (1996) Women, men, and type of talk: what makes the difference? *Language in society* 25, 1-26.
- Freeman, R. and McElhinny, B. (1996) Language and gender. In S.L. McKay & N.H. Hornberger. *Sociolinguistics and language teaching.* Cambridge, UK: Cambridge University Press, 218- 280.
- Hymes, D (1972) *On Communicative Competence.* Philadelphia, PA. University of Pennsylvania Press.
- Linacre, J.M. (1989-1995) *FACETS: a computer program for many faceted Rasch measurement.* Chicago, IL: Mesa Press.

- Maltz, D. and Borker, R. (1982) A cultural approach to male-female miscommunication. In J. Gumperz (ed.) *Language and social identity*. Cambridge, UK: Cambridge University Press, 196-216.
- McNamara, T.F. (1996) *Measuring second language performance*. London and New York, Longman.
- O'Loughlin, K.J. (1997) The comparability of direct and semi-direct speaking tests: a case study. Unpublished PhD thesis, University of Melbourne.
- Stubbe, M. (1998) Are you listening? Cultural influences on the use of supportive verbal feedback in conversation. *Journal of Pragmatics* 29:257-289
- Sunderland, J. (1995) Gender and language testing. *Language Testing Update* 17:24-35.
- Tannen, D. (1990) *You just don't understand: women and men in conversation*. New York.: William Morrow.
- Thwaite, A. (1993) Gender differences in spoken interaction in same dyadic conversations in Australian English. In J. Winter & G Wigglesworth (eds) *Language and gender in the Australian context. Australian Review of Applied Linguistics Series S No.10*, 149-179.
- UCLES. (2000) *The IELTS Handbook* . University of Cambridge Local Examinations Syndicate, The British Council and IDP Education Australia Ltd.
- Wigglesworth, G. (1993) Exploring bias analysis as a tool for improving rater consistency in assessing oral interaction. *Language Testing* 10, 3: 303-335.