Modern English Education, Vol. 17, No. 1, Spring 2016
http://dx.doi.org/10.18095/meeso.2016.17.1.06

Examining the Effects of Gender in Oral Interviews

Mi-jin Joo*
Kangwon National University
Min-hee Bang
Sangmyung University

Joo, Mi-jin & Bang, Min-hee. (2016). Examining the effects of gender in an oral interview. *Modern English Education, 17*(1), 117-133.

In response to growing interest in identifying the factors which are irrelevant to the ability being measured in oral interviews, this study aimed to explore the gender effect of the people involved in the interviews including raters, test-takers, and interviewers. Gender has been seen as one of the factors which affect test performances in spoken interactions during the interviews. 19 Korean university students (seven females and 12 males) were interviewed, once by a male and once by a female. All interviews were videotaped and scored by 17 raters (10 females and seven males). The scores were analyzed using the multi-faceted Rash measurement. The findings from the analysis were as follows. First, the male raters were more lenient in scoring than the female raters. Secondly, there was no significant variation in test-taker performances although the male test-takers performed better than the female test-takers. Thirdly, the male raters were significantly biased against the female test-takers. They consistently scored the female test-takers more leniently. Lastly, test performances were not meaningfully affected by whether their interviewer was of the same or opposite gender.

**[gender effect/oral interview/rater/bias/
성별효과/구술인터뷰/채점자/편견]**

I. INTRODUCTION

The performance in an oral interview can be influenced by diverse factors which are related to the characteristics of an interviewer, a rater, and a test-taker. Whereupon, the

* First author: Mi-jin Joo, Corresponding author: Min-hee Bang

same test-takers may give different performances or attain different scores depending on the set of characteristics the interviewer has. When the factors which are irrelevant to the test-takers' ability are found, attempts should be made to eliminate such factors from the test for the sake of the test fairness and accuracy. They may threaten and deform test validity. A number of researchers have been studying the potential factors (e.g., Locke, 1984; O'Sullivan, 1995; O'Sullivan & Porter, 1995; Porter, 1991a, 1991b; Porter & Shen, 1991).

One of the main factors which affects test performances and test results has been a gender effect in oral interviews. One possible reason why gender may be a potential source of measurement error is that males and females have quite distinct conversational styles (Coates, 1993; J. Lim, 2004; Maltz & Borker, 1982; Tannen, 1990; Thwaite, 1993). M. Song (2014) also assumed gender differences in language use even though no gender-related significant effect was found in her study. The studies on gender and language suggest that females have a collaborative, co-operative, symmetrical and supportive conversational style, while the male gender seems to be more controlling, unco-operative, asymmetrical and unsupportive. Males and females "have different sets of norms for conversational interaction" (Coates, 1993, p. 140). Thus, gender differences may be the cause of the production of different styles of output (Coates, 1993; Maltz & Borker, 1982; Thwaite, 1993). Accordingly, it remains possible that raters may have bias to a particular gender of test-takers giving higher or lower scores with the preference of the particular gender conversational style. Also, the quantity and quality of test-taker output may be affected by the interaction with the interviewer who is of the same or opposite gender.

Another reason is that a particular gender of a rater or an interviewer may favour a particular gender of test-takers. For example, the rater or the interviewer may favour the opposite gender of the test-takers and give unnecessarily higher scores or support during the interview. Or the test-takers may feel more comfortable talking with the same gender of the interviewer or vice versa. In other words, it is quite possible that the ratings of the rater and the interview process by the interviewer may be influenced by their own gender and/or the gender of the test-takers.

If the cases above mentioned were to occur, the test would not be fair and reduce the accuracy of the interpretations drawn from the test scores. For the test fairness and accuracy, it is crucial to discover the test factors which are not relevant to the test-takers' ability being measured.

In spite of some studies on gender issues (Buckingham, 1997; Locke, 1984; O'Loughlin, 2002; O'Sullivan, 2000; Porter, 1991a, 1991b; Porter & Shen, 1991; Young & Milanovic, 1992), as described in Chapter Ⅱ, the role of gender in oral interviews has not received sufficient attention. Moreover, the results of the previous studies were inconsistent. This

may stem from the lack of systematic exploration of the interaction between gender factors in oral interviews. In most of the studies, the researchers did not distinguish the role of the interviewer and that of the rater (i.e., the interviewer also performed as the rater). The interview itself may, however, result in the gender effect. As mentioned earlier, a particular gender of the interviewer may, for instance, give more or less benefits to a particular gender of the test-taker by giving more or less supportive and cooperative feedback during the interviews. Thus, the role of the interviewer was separated from that of the rater in the current study in order to see who or what actually causes the gender effect. Furthermore, the number of raters participating in the previous studies may have been too small to identify the effects of gender: the number of the raters was less than 10. Therefore, this study will attempt to answer the following questions:

1) Are there any gender differences in rater performance?
2) Are there any gender differences in test-taker performance?
3) Is there any gender bias between raters and test-takers?
4) Is there any gender bias between interviewers and test-takers?

II. PREVIOUS STUDIES

There have been several studies on the effects of gender in oral interviews, which measures general language proficiency in the form of face-to-face interaction. Locke (1984) investigated the effects of gender of an interviewer in oral interviews and found Arab test-takers received higher scores (i.e., performed better) when interviewed by a male (interviewer) than by a female (interviewer). Porter's study (1991a) on gender, personality and acquaintanceship of the participants also found a similar result to that of Locke's. Arab students comprising of 11 males and two females found that significantly higher scores were gained with a male interviewer. In the following study, Porter (1991b) explored the effects of gender of the interviewer on performance and test-takers' perception of their interviewer status. The study involved 12 male and four female Algerian test-takers. Again, the findings of the study showed that the test-takers tended to have higher scores with a male interviewer and with a person who had not been perceived to be in a higher status. Porter and Shen (1991) investigated the effects of gender status, and interaction style in oral interviews with the participation of 14 males and 14 females. The results showed that while there was a significant gender impact on test performance, no significant effect of status was found. However, unlike the results of the previous studies above, the test-takers had higher scores when interviewed by a female. Also, it was found that the interviewers using a more female or supportive interaction style awarded significantly higher scores to

the test-takers. Young and Milanovic (1992) found, in their qualitative study, gender differences in language suggesting the gender of both interviewer and test-takers may be one of the factors that accounts for discourse variation in the interviews.

Buckingham (1997) examined the effects of gender and age of an interviewer on test performances, involving 16 male and 16 female Japanese students. Findings showed that there was no age effect, but the interviewer gender did affect test performances. There was interaction between the gender of the interviewer and that of the test-takers. The test-takers gained higher scores when interviewed by the same gender of interviewer. However, O'Sullivan (2000) found contradictory results to Buckingham (1997). O'Sullivan (2000) explored the effects of gender in oral interviews. 12 Japanese test-takers participated in the study and were interviewed once by a male and once by a female. The results indicated that most of the test-takers attained higher scores with the female interviewer irrespective of their own gender. Systematic gender differences were found in the analysis of interviewer language, while there was a tendency to produce more grammatically accurate language with a female interviewer in the analysis of the test-taker responses. O'Loughlin (2002) examined the effects of gender in IELTS interviews. Eight female and eight male test-takers were interviewed once by a male and once by a female. Their performances were scored by four raters (two males and two females) and the scores were analyzed using multi-faceted Rasch bias analyses. The results indicated that gender did not have a significant impact on the IELTS interviews. The results of the studies above were not the same, but most of them showed the impact of the interviewer gender. They would be comparable with the result of the current study.

III. DESIGN OF THE STUDY

1. Participants

In the current study, 19 Korean students (seven females and 12 males) majoring in English were involved. They were all interviewed twice in the researcher's office: once by a male and then once by a female. The students were divided into two groups. Half of the students were interviewed by a female first and the other half by a male first in order to minimize the order effect of test-taking. The interviewers' role was to interview the students and did not involve any participation in awarding scores. After the two interviews, each interview was scored by 17 raters (10 females and seven males). O'Loughlin's (2002) rating scale were modified and used for rating of the students. Their pronunciation, fluency, grammar, vocabulary, cohesion, overall communicative effectiveness were rated on a 6-point scale.

The two interviewers and 17 raters were intensively trained for one day before the actual interviews and ratings. Both interviewers were Korean, in their forties and with more than seven years of teaching experience. Seven of the 17 raters were Korean and the other 10 raters were English native speakers. The age of the raters ranged from in their twenties to fifties. Their years of teaching experience varied from two years to more than 20 years. The raters' inter- and intra- reliabilities were stated in Chapter IV.

2. Data Collection

The oral interviews were individually administered to each test-taker in a university office and were all completed in one semester. All interviews were video-taped for rating. The type of interview tasks and items were constructed referred to M. Joo (2014). The interviews began with short "question and answer" items followed by "picture descriptions," "making up a stories" and "role play" items. Slightly different items were used in each interview to avoid the practice effect. The items were also carefully made to be a similar degree of familiarity and difficulty. The time taken for each interview was approximately between 15-20 minutes.

3. Data Analysis

Multi-faceted Rasch measurement with the program FACETS (version 3.45) (Linacre, 2008) was employed in order to enhance validity and reliability of the analysis. It is known as one of the most encouraging developments in analyzing performance assessments. It can compensate test-takers for differences in severity between raters and other facets such as item or task difficulty in calculating the probable scores (logit scores or measures). Also, it is far easier to examine and compare the differences between the interviews conducted by male and female interviewers because it is not necessary to subject the test-takers to the interview with each interviewer.

The program FACETS allows for the analysis of bias, which identifies unexpected but consistent pattern of significant interaction between facets. The bias analysis provides more useful and precise information about any changes of raters or test-takers on different testing situations. For instance, it is able to detect if one gender of raters score the same gender of test-takers as their own significantly more severely or leniently than they score the opposite gender of test-takers. Thus, bias analysis was employed in this study to investigate the significant interaction between rater gender and test-taker gender and between interviewer gender and test-taker gender.

For the analysis, three FACETS analyses were conducted. One for the interview conducted by a female (see Appendix 1), another for the interview conducted by a male

(see Appendix 2), and the other for the combined data (see Appendix 3). For each interview, five facets included test-taker, rater, test-taker gender, rater gender, and items in the analysis. For the combined data, the six facets of test-taker, rater, interviewer gender, test-taker gender, rater gender, and items were included.

IV. FINDINGS

1. Are There Any Gender Differences in Rater Performance?

Table 1 indicates rater severity showing the logit scores together with their standard error and infit mean square values. In this table, 17 raters are ordered from highest to lowest in terms of their severity. Thus, the most severe rater was Rater 11 (0.5 logits) and the least lenient rater was Rater 15 (-1.77 logits). It also revealed a significant chi-square given the separation index, 12.25 with reliability 0.99, suggesting that the raters differ in their severity. However, the infit mean square values ranging between 0.5 and 1.5 show that neither of the raters showed a misfit, indicating that their ratings were consistent. In short, raters had quite fair intra-reliability even though they significantly differed in their relative severity.

TABLE 1

Rater Measurement Report (Combined)

Rater	Gender	Fair-*M* Average	Logit	*SE*	Infit *MS*
11	F	4.13	-1.77	.10	1.2
8	M	3.99	-1.44	.10	.8
13	M	3.89	-1.21	.10	.8
12	F	3.87	-1.16	.10	1.2
10	M	3.85	-1.11	.10	1.1
16	M	3.77	-.93	.10	.9
14	F	3.60	-.57	.10	1.1
4	F	3.54	-.44	.10	1.1
7	F	3.54	-.44	.10	1.3
2	F	3.43	-.20	.10	.9
6	F	3.05	.60	.10	1.2
1	F	3.04	.64	.10	.7
17	M	3.03	.65	.10	1.2
3	M	2.72	1.31	.10	1.2
5	F	2.56	1.67	.10	.9
9	F	2.40	2.03	.10	.7
15	M	2.26	2.38	.11	.5

Note. Separation: 12.25, Reliability: .99, Fixed (all same) chi-square: 2478.8, *df*=16, *p*=.00, Observed exact agreements: 33.0%, Expected exact agreement: 32.2%

In order to obtain the overall severity of the raters according to gender, the gender of the

rater measurement was examined. Table 2 presents that the male raters had the higher logit value, which means that they were slightly less harsh than the other gender (p=0.03). Namely, the male raters gave more generous scores to the test-takers.

TABLE 2

Gender of Rater Measurement Report (Combined)

Rater Gender	Fair-*M* Average	Logit	*SE*	Infit *MS*
Male	3.36	-.06	.04	.9
Female	3.31	.06	.03	1.0

Note. Separation: 1.21, Reliability: .59, Fixed (all same) chi-square: 4.9, df=1, p=.03

To conclude, the rater ratings were consistent, but their severity was significantly differed. In gender differences in severity, all in all the male raters were less severe in scoring than the female raters.

2. Are There Any Gender Differences in Test-taker Performance?

In Table 3, test-takers are ordered in terms of their ability from highest to lowest. Thus, the higher the logit value was, the better the test-taker: Test-taker 5 was able to perform the best, and Test-taker 15 was the least well. None of the test-takers had infit mean square values outside of a range of 0.5 to 1.5. This means that their performances were consistent with the pattern of performances on the test overall. That is, no overfitting test-takers

TABLE 3

Test-taker Measurement Report (Combined)

Test-taker ID	Gender	Fair-*M* Average	Logit	*SE*	Infit *MS*
5	M	5.67	5.68	.16	1.1
17	F	4.08	1.21	.11	.7
13	M	3.89	.77	.11	1.0
18	M	3.85	.67	.11	1.0
8	F	3.68	.29	.10	1.4
10	F	3.34	-.43	.10	1.1
4	M	3.34	-.44	.10	1.1
16	M	3.25	-.62	.10	1.0
14	M	3.24	-.63	.10	.9
6	M	3.21	-.70	.10	.7
7	M	3.03	-1.08	.10	.8
19	M	3.01	-1.13	.10	.8
2	M	2.94	-1.27	.10	.8
12	F	2.89	-1.39	.11	1.2
9	F	2.84	-1.49	.11	1.0
11	M	2.82	-1.53	.10	.8
3	F	2.74	-1.71	.11	.9
1	F	2.57	-2.07	.11	1.3
15	M	2.41	-2.45	.11	1.3

(below 0.5) whose scores did not vary significantly across all rating items and no misfitting test-takers (above 1.5) whose scores had too much variation in performances.

As shown in Table 4 and Table 5, the higher logit score for the female test-takers indicates that the test was more difficult for them than for the male test-takers in both interviews. That is, the male test-takers performed better than the female test-takers in each interview. The performance difference between the male and female test-takers when interviewed by either gender of interviewers were significant (p=0.00), and the difference showed a slight increase with the male interviewer (Table 5).

TABLE 4

Gender of Test-taker Measurement Report (Interview by a Female)

Test-takers	Fair-*M* Average	Logit	*SE*	Infit *MS*
Male	3.45	-.21	.05	1.0
Female	3.27	.21	.06	1.1

Note. Separation: 3.83, Reliability: .94, Fixed (all same) chi-square: 31.3, df=1, p=.00

TABLE 5

Gender of Test-taker Measurement Report (Interview by a Male)

Test-takers	Fair-*M* Average	Logit	*SE*	Infit *MS*
Male	3.43	-.29	.05	.9
Female	3.19	.29	.06	1.2

Note. Separation: 5.36, Reliability: .97, Fixed (all same) chi-square: 59.5, df=1, p=.00

In brief, there was no real significance in variation or uniformity in test-taker performances. The male test-takers obtained higher scores than the female test-takers in both interviews.

3. Is There Any Gender Bias between Raters and Test-takers?

In order to see whether there is any interaction between rater gender and test-taker gender, bias analysis was carried out. Raters may have given more or less severe scores to the test-takers depending on their gender. Table 6 and Table 7 show the results from the bias analysis between rater gender and test-taker gender. Table 6 reveals that the male raters gave -0.23 higher logit scores to the female test-takers than to the male test-takers. In Table 7, the Z-score value falls below -2.0 (z-score=-2.32) suggesting the male raters scored the female test-takers significantly more leniently than the other gender. In other words, the male raters were significantly biased in favour of the female test-takers.

On the other hand, the female raters judged the performances of the female test-takers more harshly than the other gender. They tended to give more severe scores to the female test-takers than the male test-takers by 0.16 although the difference was not significant.

The infit mean square values, which indicate how similar the rater's scoring is for the test-takers overall, indicate that the pattern of interaction between the raters and the test-takers was consistent: all values are between 0.5 and 1.5. To sum up, the male raters consistently scored the female test-takers more leniently compared to the way that the raters scored the male test-takers.

TABLE 6

Bias Pairwise Report: Test-taker-Rater Interaction (Combined)

Rater	Test-taker	Logit	Error	Test-taker	Logit	Error	Diff	Error	*t*	*df*
Male	Female	-.20	.06	Male	.03	.05	-.23	.08	-2.94	1594
Female	Female	.16	.05	Male	.00	.04	.16	.07	2.44	2278

TABLE 7

Bias Calibration Report: Test-taker-Rater Interaction (Combined)

Test-taker	Rater	Bias Logit	Error	*z*-score	Infit *MS*
Female	Male	-.14	.06	-2.32	.9
Male	Female	-.06	.04	-1.49	1.0
Male	Male	.09	.05	1.81	.9
Female	Female	.10	.05	1.93	1.2
M		.00	.05	-.02	1.0
SD		.10	.01	1.91	.1

Note. Fixed (all=0) chi-square: 14.6, *df*=4, *p*=.01

4. Is There Any Gender Bias between Interviewers and Test-takers?

The bias analysis was conducted again to see whether there was a significant interaction between interviewer gender and test-taker gender. Table 8 and Table 9 present that all the *z*-scores are within the range of -2 to +2 indicating that the interaction was not significant. There was no significant interaction between the interviewer gender and the test-taker gender. Thus, it is reasonable to assume that the interviewer gender had no impact on test performances regardless of the test-taker gender.

They gained -0.18 higher logit score when the female test-takers were interviewed by the female compared to the male interviewer. The male test-takers received higher scores with the male interviewer (logit difference=0.11). However, the difference was insignificant even though the test-takers were awarded slightly higher scores by the same gender of the interviewer. The results of an independent-samples *t* test also showed that there was no significant differences between the scores obtained when interviewed by the male and those when interviewed by the female irrespective of the test-takers' own gender ($p>0.05$). Therefore, it can be concluded that the test-takers' performances were not significantly influenced by the interviewers' gender.

TABLE 8

Bias Pairwise Report: Interviewer-Test-taker Interaction (Combined)

Test-taker	Interviewer	Logit	Error	Interviewer	Logit	Error	Diff	Error	*t*	*df*
Female	Female	.16	.06	Male	.34	.06	-.18	.08	-2.31	1426
Male	Female	-.19	.04	Male	-.30	.04	.11	.06	1.79	2446

TABLE 9

Bias Calibration Report: Interviewer-Test-taker Interaction (Combined)

Interviewer	Test-taker	Bias Logit	Error	*z*-score	Infit *MS*
Female	Female	-.09	.06	-1.63	1.0
Male	Male	-.06	.04	-1.26	.9
Female	Male	.06	.04	1.27	1.0
Male	Female	.09	.06	1.63	1.2
M		.00	.05	.00	1.0
SD		.08	.01	1.46	.1

Note. Fixed (all=0) chi-square: 8.5, *df*=4, *p*=.07

V. DISCUSSION

Each finding of the current study is discussed as follows. First, in gender differences in severity, overall the male raters were less severe in scoring than the female raters. It is difficult to understand why the male raters were more lenient in giving scores. This may have resulted from their personality, different understanding of the rating scale descriptors, different views about the nature of speaking, or some other unfounded variables. The one highly possible reason may, however, be that the male raters had a bias to the female test-takers, which was another important finding, the answer to the third research question in this study. Irrespective of interviewer gender, the male raters consistently scored the female test-takers more leniently compared to the way that the raters scored the other gender of test-takers. This may have resulted in making the average scores awarded to the test-takers higher than those of the female raters. In addition, the gender bias might have occurred because all test performances had been video-taped rather than audio-taped. They may simply have favoured the opposite gender of their own or have preferred the conversational style of the female test-takers. Whatever the reasons may be, the result indicates that rater gender could affect their ratings.

Secondly, there was no significant variation in test-takers' performances, which means that the overall patterns of the performances identified among the test-takers were consistent. The male test-takers obtained higher scores than the female test-takers in both interviews. It seems that the scores awarded reflect the test-takers' genuine ability because the raters had high intra-reliability indicating the consistent ratings of the raters. Furthermore, there was no evidence that the oral interviews itself gave more benefits to the

male test-takers. Rather, as mentioned above, the female test-takers received more generous scores from the male raters.

Lastly, the test-taker performances were not significantly affected by their interviewers' gender. This is not consistent with the results of the previous studies showing the gender effect of the interviewers (e.g., Buckingham, 1997; Locke, 1984; O'Sullivan, 2000; Porter, 1991a; Porter, 1991b; Porter & Shen, 1991), but did show the same results as O'Loughlin's (2002). Why was there little or no gender impact of the interviewers?

One probable reason may be that the role of the interviewer is to try and be gender-neutral like O'Loughlin (2002) assumed. Or the test tasks and the test-takers in this study are so gender-neutral that they are not affected by the interviewer gender. Another possible reason may be the one day interviewer training which was conducted before the actual interviews. It was carried out to familiarize the test items, to adapt the limited test time, and to train them to use the same or similar elicitation techniques to avoid other confounding variables besides gender.

VI. CONCLUSION

The purpose of the study was to investigate the gender effect of raters, test-takers and interviewers in an oral interview. The results suggest that the gender of the raters had a significant impact on the oral interviews in this study. In other words, rater gender can play a role as a construct irrelevant variability which affects the rating process and finally the test scores. Therefore, one suggestion for reducing or removing bias is to train raters before actual ratings although it is not possible to entirely eliminate the differences in severity between raters. It is a repetitive but still essential suggestion. At least, it will help them to hold similar perceptions of speech samples and to reduce random errors, rater-related score variances, wrong interpretation of the rating scales, and finally to increase overall inter-and intra- reliability. Although no significant impact of interviewer gender was found in this study, the training of the interviewer is also necessary because many of the previous studies showed the impact of the interviewer gender. The training should devote attention to the variable aspects of interviewer support using a range of different speaking strategies (Fulcher, 2003). Thus, they should be trained to make the interaction with the test-takers more similar so that they cannot give more or less advantages to a particular group of test-takers. The test accuracy and test fairness might be threatened unless the training for both raters and interviewers are conducted.

The results of this study contribute to the ongoing discussion on the impact of gender and to the literature of gender effect in an oral interview, but limitations should be acknowledged for future research. First, the sample in this study is quite limited. Although

the sample size of the test-takers and the raters was relatively large enough compared with those of the previous studies, only two interviewers participated and caution should be taken in generalizing these results. Second, the participants were all Korean students who were similar or the same age, language proficiency levels, and occupation. This study might lack the sample of students of diverse backgrounds. Thus, it may be difficult to generalize the findings to all male and females.

Given the limitations, there is a need for comprehensive research on the validity of an oral interview investigating a wide range of variables affecting test scores with larger participants of different backgrounds. Further research is also necessary to understand better about the impact of gender by examining gender discourse. This study mainly investigated the gender effect on the basis of the scores given by the raters and not in any differences in actual test performances. Several researchers suggest that the male and female conversational styles have distinct differences (e.g., Coates, 1993; Maltz & Borker, 1982; Tannen, 1990; Thwaite, 1993). Thus, the language of test-takers and interviewers needs to be looked closely at to detect the differences in gender speech and its impact on test scores.

REFERENCES

Buckingham, A. (1997). *Oral language testing: Do the age, status and gender of the interviewer make a difference?* Unpublished MA dissertation, University of Reading, UK.

Coates, J. (1993). *Women, men and language* (2nd ed.). London: Longman.

Fulcher, G. (2003). *Testing second language speaking*. Malaysia: Pearson Longman.

Joo, Mi-jin. (2014). Examining rater reliability and bias in measuring English speaking performance: Through a comparison of scores on an oral interview, a computerized oral test and a Versant. *English Teaching, 69*(2), 247-270.

Lim, Jeong-wan. (2004). Gender and politeness: A comparative study of Korean and American students. *Modern English Education*, *5*(2), 54-67.

Linacre, J. M. (2008). *A user's guide to FACETS: Rasch-model computer programs.* Chicago: MESA Press.

Locke, C. (1984). *The influence of the interviewer on student performance in tests of foreign language oral/aural skills.* Unpublished MA dissertation, University of Reading, UK.

Maltz, D., & Borker, R. (1982). A cultural approach to male-female miscommunication. In J. Gumperz (Ed.), *Language and social identity* (pp. 196-216). Cambridge: Cambridge University Press.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing, 19*(2)*,* 169-192.

O'Sullivan, B. (1995). *Oral language testing: Does the age of the interlocutor make a difference?* Unpublished MA dissertation, University of Reading, UK.

O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System, 28*, 373-386.

O'Sullivan, B., & Porter, D. (1995). *The importance of audience age for learner-speakers and learner-writers from different cultural backgrounds*. Paper presented at the RELC Conference, Singapore.

Porter, D. (1991a). Affective factors in language testing. In J. C. Alderson & B. North (Eds.), *Language testing in the 1990s* (pp. 32-40). London: Modern English Publications.

Porter, D. (1991b). Affective factors in the assessment of oral interaction: Gender and status. In S. Anivan (Ed.), *Current developments in language testing* (Anthology Series 25, pp. 92-102). Singapore: SEAMEO Regional Language Centre.

Porter, D., & Shen, S. H. (1991). Sex, status and style in the interview. *The Dolphin, 21*, 117-128.

Song, Min-jong. (2014). Gender effect on the use of communication strategies in case of Korean college students. *Modern English Education*, *15*(3), 43-63.

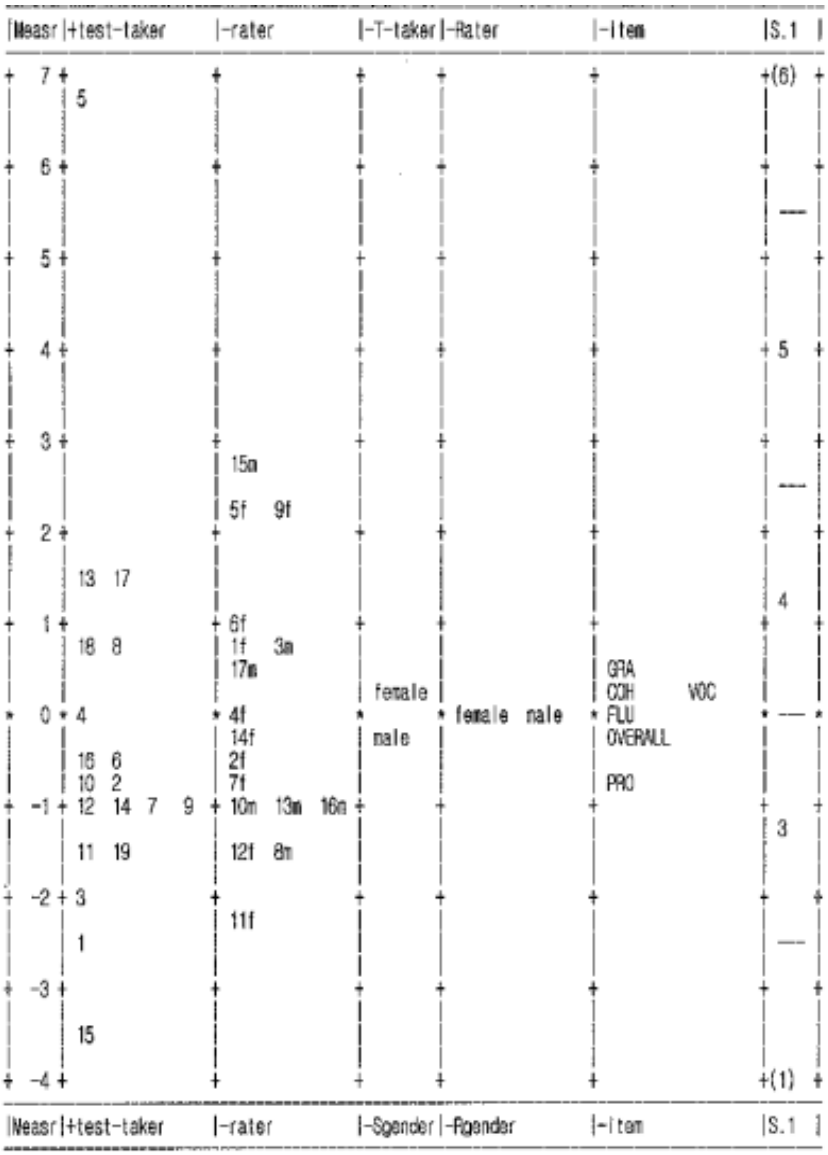
Tannen, D. (1990). *You just don't understand: Women and men in conversation.* New York: William Morrow.

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, pp. 149-179). Canberra: Applied Linguistics Association of Australia.

Young, R., & Milanovic, M. (1992). Discourse validation on oral proficiency interviews. *Studies in Second Language Acquisition, 14*, 403-424.

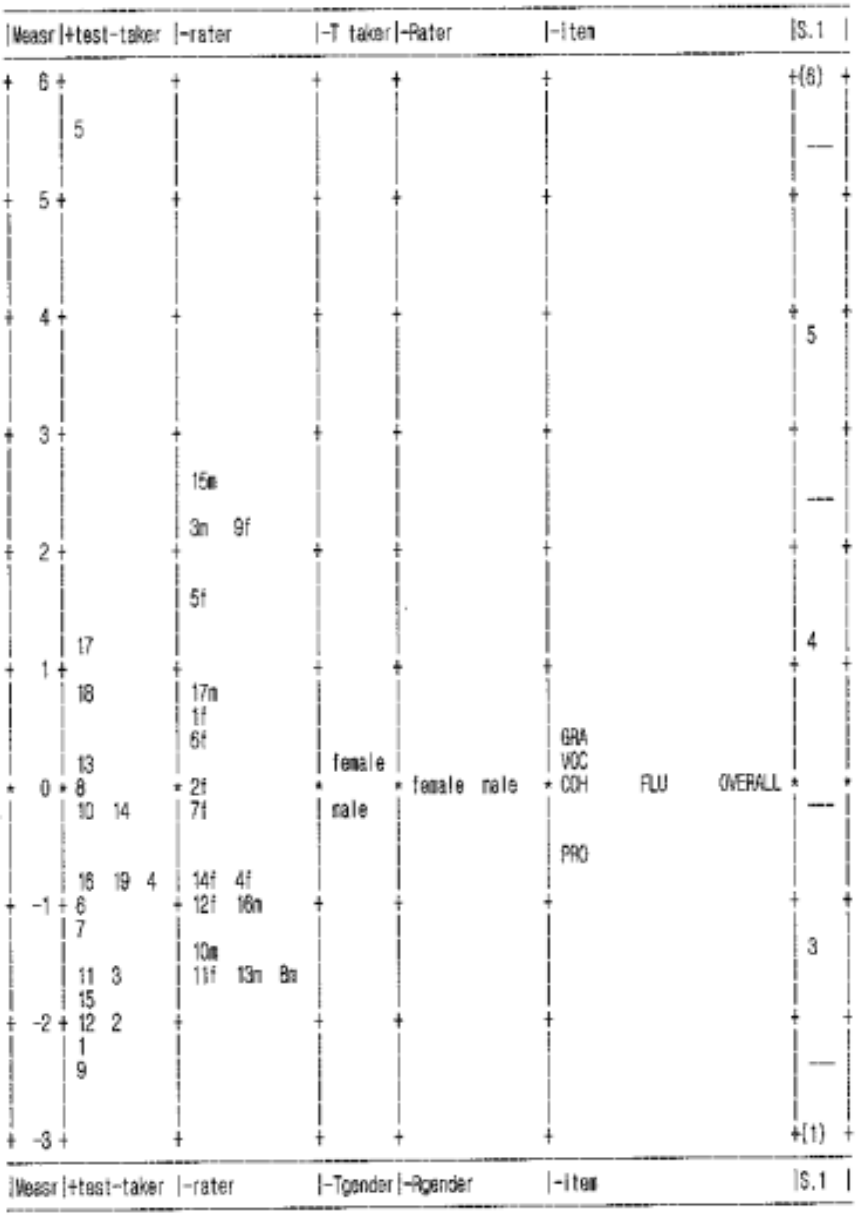
APPENDIX 1

All Facet Vertical Rulers (Interview by a Female)



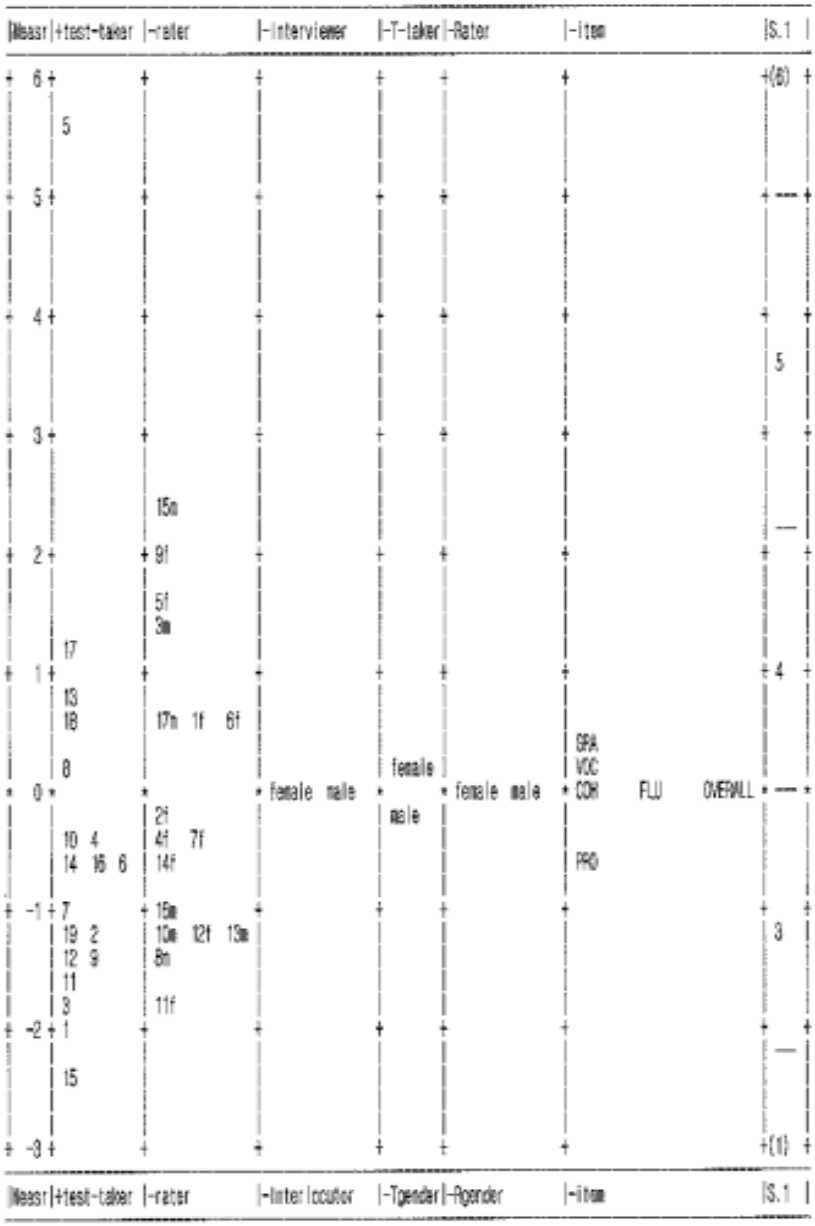
APPENDIX 2

All Facet Vertical Rulers (Interview by a Male)



APPENDIX 3

All Facet Vertical Rulers (Combined)



Examples in: English
Applicable Language: English
Applicable Levels: Tertiary

Mi-jin Joo
Department of English
Kangwon National University
200-701, Joongang-ro, Samcheok-si, Gangwon-do, Korea
Tel: 033-570-6657
E-mail: ing1115@hanmail.net

Min-hee Bang
College of General Studies
Sangmyung University
31066, Sangmyungdaegil 31, Dongnamgu, Cheonanshi, Chungnam, Korea
Tel: 010-5797-3799
Email: minheeb@gmail.com

Received 6 December 2015
Revised 27 January 2016
Accepted 12 February 2016