The Study of How English Teachers’ Background Characteristics Affect Their Holistic Rating Scores in Speaking Assessment
### Conference Schedule

**Day 1  10/15 (Fri.)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Venue</th>
<th>Topic</th>
<th>Presenters</th>
<th>Moderators</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:20-8:55</td>
<td>A302</td>
<td>Registration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8:55-9:00</td>
<td>A302</td>
<td>Inaugural</td>
<td>President</td>
<td>Chair of Language Center</td>
</tr>
<tr>
<td>9:00-10:00</td>
<td>A302</td>
<td>(Keynote Speech) Dealing with a foreign language speaking anxiety in an EFL classroom</td>
<td>Dr. Svetlana Timina (Shih Chien University Kaohsiung campus)</td>
<td>Dr. Wan-Lih Chang (張婉麗教授) Shih Chien University Kaohsiung campus</td>
</tr>
<tr>
<td>10:00-10:20</td>
<td>A302</td>
<td>Tea Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10:20-11:10</td>
<td>A302</td>
<td>(Invited Speech I) Topic: Modules of EOP Curricula</td>
<td>Dr. Chin-Cheng Huang (黃金誠教授) Fooyin University</td>
<td>Dr. Chuan-Hsiu Hung (洪銓修教授) National Yunlin University of Science &amp; Technology</td>
</tr>
<tr>
<td>11:10-12:00</td>
<td>A302</td>
<td>Exploring Social Interactions of TEFL in the TVE--A Macro and Micro Perspective</td>
<td>Dr. Chuan-Hsiu Hung (洪銓修教授) National Yunlin University of Science &amp; Technology</td>
<td>Dr. Chin-Cheng Huang (黃金誠教授) Fooyin University</td>
</tr>
<tr>
<td>12:00-13:00</td>
<td>A302</td>
<td>Lunch Recess</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13:00-14:00</td>
<td>A302</td>
<td>(Invited Speech III) Topic: Daily EFL writing and its possible impact on EFL performance</td>
<td>Timothy Henry Bullins (Fooyin University)</td>
<td>Dr. Wan-Lih Chang (張婉麗教授) Shih Chien University Kaohsiung campus</td>
</tr>
<tr>
<td>14:00-15:00</td>
<td>A302</td>
<td>(Invited Speech IV) Topic: Teachers’ Perceptions of EAP Teaching at Business Graduate Schools</td>
<td>Dr. Shih-Guey Joe (周碩貴教授) Chang Gung Institute of Technology</td>
<td>Dr. Hsueh-Yu Cheng (鄭雪玉教授) Aletheia University Matou Campus</td>
</tr>
<tr>
<td>15:00-15:20</td>
<td>A302</td>
<td>Tea Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15:20-15:40</td>
<td>A302</td>
<td>Patterns of Peer Interaction and Scaffolding in Adult EFL Classes: An L2 Socialization Perspective</td>
<td>Chi-Lin Wang (王啟琳), Ngaanleng Mak (麥雁鈴) National Taiwan University of Science and Technology</td>
<td>Dr. Hsueh-Yu Cheng (鄭雪玉教授) Aletheia University Matou Campus</td>
</tr>
<tr>
<td>15:40-16:00</td>
<td>A302</td>
<td>A Study of English Learning Motivation of University Administrative Staff</td>
<td>Ching-I Lin (林靜憶) Tajen University</td>
<td></td>
</tr>
<tr>
<td>16:00-16:20</td>
<td>A302</td>
<td>The Study of How English Teachers’ Background Characteristics Affect Their Holistic Rating Scores in Speaking Assessment.</td>
<td>Ying-Ying Chuang (莊嬿纓) Cheng Shiu University</td>
<td></td>
</tr>
<tr>
<td>16:20-16:40</td>
<td>A302</td>
<td>The Earlier, The Better? The Effects of the Test-Intensive Instructional Program on English Majors’ Test Performances</td>
<td>Hsing-Cheng Tung (董幸正) Chia Nan University of Pharmacy and Science</td>
<td></td>
</tr>
<tr>
<td>16:40-17:00</td>
<td>A302</td>
<td>Learners’ Perceptions on the Use of Blogs for EFL Learning.</td>
<td>Hui-Ju Wu (吳蕙如) University of South Florida</td>
<td></td>
</tr>
<tr>
<td>17:00~</td>
<td></td>
<td>Home~</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## 2010 International Conference on English/Foreign Languages Education

### Conference Schedule  
**Day 2  10/16 (Sat.)**

<table>
<thead>
<tr>
<th>Time</th>
<th>Venue</th>
<th>Topic</th>
<th>Presenters</th>
<th>Moderators</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:30-8:50</td>
<td>A302</td>
<td>Registration</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 8:50-10:20 | A302  | (Keynote Speech)  
Topic: Facing SLA in Taiwan                                             | Dr. Raung-Fu Chung  
(Southern Taiwan University)                                                                 | Dr. Pi-Ching Chen  
(Chang Jung Christian University)                                                                 |
| 10:20-10:40| A302  | Tea Time                                                              |                                                                                                 |                                                                                                       |
| 10:40-12:10| A302  | (Invited Speech I)  
Topic: EGP+ESP: Curriculum Design, Material Development and Learning Assessment for the ESP Instruction | Dr. Pi-Ching Chen  
(Chang Jung Christian University)                                                                 | Dr. Raung-Fu Chung  
(Southern Taiwan University)                                                                 |
| 12:10-13:00| A302  | Lunch Recess                                                         |                                                                                                 |                                                                                                       |
| 13:00-14:00| A302  | (Invited Speech II)  
Topic: Foreign language speaking anxiety: Theoretical and empirical perspectives | Dr. Po-Chi Kao  
(Shih Chien University Kaohsiung campus)                                                | Dr. Wan-Lih Chang  
(Shih Chien University Kaohsiung campus)                                                     |
| 14:00-15:00| A302  | (Invited Speech III)  
Topic: Computer Mediated Communication                                    | Paul Blewchamp  
(Shih Chien University Kaohsiung campus)                                                 | Dr. Po-Chi Kao  
(Shih Chien University Kaohsiung campus)                                                      |
| 15:00-15:20| A302  | Tea Time                                                              |                                                                                                 |                                                                                                       |
| 15:20-15:40|       | An Investigation of the Impacts of Reader's Theater on English for Specific Purpose (ESP) Learning in Taiwan | Li-Li Kuo  
(Tzu Chi College of Technology)                                                    | Dr. Yu-Su Lan  
(Chang Jung Christian University)                                                            |
| 15:40-16:00|       | “A Big Different World”—Implementing a Project-Based Learning Instruction in the EFL Classrooms | Pei-Chi Chen  
(Tzu-Hui Institute of Technology)                                            |                                                                                                       |
| 16:00-16:20|       | Investigation of the Motivation dynamics in a pre-ESP classroom         | Yin-Hsi Chen  
(Chihlee Institute of Technology)                                              |                                                                                                       |
| 16:20-16:40|       | Challenges and Strategies of English-Only Pedagogy: Teacher-Student Perspectives | Shih-Chi Yu  
(National Yunlin University of Science & Technology)    |                                                                                                       |
| 16:40-17:00|       | The Study of Cultural Perception of Taiwanese Junior High School Participants in TEFL | Su-Ching Wang  
(Dayeh University)                                               |                                                                                                       |
| 17:00~     |       | Home~                                                                |                                                                                                 |                                                                                                       |
The Study of How English Teachers’ Background Characteristics Affect Their Holistic Rating Scores in Speaking Assessment

Ying-Ying Chuang
Department of Applied Foreign Languages, Cheng Shiu University, Taiwan

ABSTRACT

Language speaking assessment usually employs human raters to judge learners’ proficiency by scoring their performance. Therefore, the role of the raters is important in the assessments process: not only their subjective judgment impacts the scores obtained, but their reliability also influences the meaning and quality of the scores. Some relevant previous research studies have examined the relationship between raters and test scores in ELL oral proficiency tests, and in which most of their findings indicated that raters’ background differences did affect their rating behaviors in certain aspects. This paper focuses on four specific sections as factors of rater effect categories: the gender factor, the rater’s native language factor, the rater’s academic background factor, and the rater training factor. The results of the study show that the raters’ features somehow cause a certain degree of impact on test score differences, especially that the raters’ academic background factor reaches statistically significant differences: the overall holistic scores rated by the raters with linguistics or literature major backgrounds were significantly much stricter than the raters with TESOL backgrounds and other major backgrounds. In other words, the results and findings did provide the evidence to support that the rater effects are the influencing factors in learners’ scores of oral proficiency assessment. Some implications for English teachers regarding speaking tests are included.

Keywords: Foreign Language Speaking Test; Background Factors, Language Assessment, Rater Effects.
I. INTRODUCTION

- In language testing, speaking is challenging to measure due to its complicated process.
- Learners’ scores are affected by the potential variables in testing process: the rater, test types, scale criteria, the learner’s preparation and skills, test conditions, and so on.
- This paper intends to explore the issue of rater features which may cause some impacts towards test results, particularly the rater effects—the factors such as gender, native language, academic and professional background or experience, and training.
- The study aims to examine how the college English teachers’ characteristics and background differences affect their rating scores on the student speech samples.

II. LITERATURE REVIEW

- The role of the raters is influential in the process of speaking test: not only their judgment impact decision-making in scoring, but the reliability also affects the meaning and quality of the scores.
- Some studies have examined the relationship between raters and scores in EFL speaking tests (Bachman, 1995; Brown, 1995; Lumley & McNamara, 1995; Caban, 2003; Kim, 2005; Chuang, 2007), and their findings have proven that raters’ background differences do affect their rating behaviors in certain aspects.
- Recently, Liu & Wen’s study (2007) indicated that the rater errors, in terms of severity, leniency, and unexpected ratings, did exist.
- In this study, The author focuses on 4 specific factors of the rater effect: (1) gender, (2) native language, (3) academic background, and (4) the rater training.
The Gender Factor

O’Sullivan (2000) examined if any distinctive features of the male and female interviewers associated with gender effect. The results showed:

- All the examinees received higher scores rated by female interviewers with a high inter-rater reliability.
- The examinees produced more grammatical accuracy with the female interviewers.
- The male raters significantly employed more fillers (well, uh, okay, um) and minimal responses (short utterances only such as yeah, mmm, uh-huh) and less supportive responses during the interview.
- The female raters tended to provide more supportive ways to expand questions and express interest on what the examinees responded.

This result matched Brown and Levinson’s (1978) observation that “women tend to express politeness and support by acknowledging and building on the utterances of other speakers in an interaction” (p. 10).

The Gender Factor (continued)

- Porter and Shen (1991) investigated the rater effects of gender, status, and interaction styles factors; the findings showed significance only in gender factors, but no differences in the others.
- Buckingham’s (1997) study showed that examinees who were interviewed by the same gender (female interviewee vs. female interviewer; male interviewee vs. male interviewer) achieved higher scores.
- Furthermore, Sunderland (1995) points out that the different styles of male and female interviewers are viewed as a type of gender effect which may affect learners’ outcomes, either strengthen or weaken their performance.
- O’Sullivan (2000) believes that learners’ affective reaction may impact their performance relating to the characteristics of their interviewer.

The Native Language Factor

- Whether or not the native speakers are more suitable to judge English learners’ speaking than the nonnative speakers?
- Whether or not the raters’ L1 backgrounds influence their judgments?
- Generally, in large-scale language tests, such as the International English Language Testing System (IELTS), nonnative speakers are not allowed to be trained as raters. However, Brown (1995) argues that there is still little empirical evidence showing that native speakers are better or more reliable raters than nonnative speakers.
- Smith and Bisazza (1982) tested the comprehensibility of 3 different varieties of English with native and nonnative English users in 7 countries to see if any significant differences exist. They found out that great exposure to varieties of English can enhance comprehensibility.
III. THE METHOD

- This study aims to look at how the college English teachers’ background differences affect their rating behaviors in speaking assessment. 75 teachers served as the raters to score students’ speech samples.
- Each rater was asked to rate the same 5 speech samples by holistic rating scale. The scale range was from 1 (very low proficiency) to 7 (very high proficiency).
- Finally, they had to complete a Rater Survey, including raters’ personal information: age, native language, academic major, teaching and rating experiences and so on.

IV. THE RESULTS

To explore if the raters’ gender made any difference in their given scores, the teachers were divided into two groups: male or female.

Table 1 Descriptive Statistics of Holistic Scores Based on Raters’ Gender (N=75)

<table>
<thead>
<tr>
<th>Rater’s Gender</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>4.39</td>
<td>27</td>
<td>.32</td>
</tr>
<tr>
<td>Female</td>
<td>4.45</td>
<td>48</td>
<td>.71</td>
</tr>
</tbody>
</table>

The descriptive statistics showed that the scores rated by male raters were lower than the scores rated by the female raters.

Table 2 Independent Samples t-Test for Raters’ Holistic Scores Based on Rater’s Gender

<table>
<thead>
<tr>
<th>Equality of Variance</th>
<th>t df Sig. (2-tailed)</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal Varniances Assumed</td>
<td>3.91 .06 39 73 -.70 -.06</td>
<td></td>
</tr>
<tr>
<td>Equal Varniances Not Assumed</td>
<td>-.43 68 .67 -.06</td>
<td></td>
</tr>
</tbody>
</table>

Note. * p < .05

The result showed that even though the scores rated by the male raters were stricter than the female raters, NO statistically significant difference was found between their mean scores (p = .70 > .05).
V. DISCUSSIONS AND CONCLUSION

Research Question: "How college English teachers’ background characteristics—(1) gender, (2) native language, (3) academic background, and (4) training factors—affect their rating scores?"

The collected data indicates that teachers' background differences in gender, native language, and training did somehow make their rating score different: male teachers tended to give lower scores than female teachers, native speakers tended to give lower scores than nonnative speakers, and trained teachers gave lower scores than non-trained teachers. However, these 3 factors reached NO statistically significant differences.

In terms of academic-background factor, the result do show strong statistically significant differences—the raters with linguistics/literature background were the strictest among the 3 groups.

Implications:
The diverse rating behaviors can be controlled more or less by well-designed rater trainings which include clear criteria and standards, and samples of different levels of oral performance—thus to enhance raters' broader perspectives (Brown, 1995). Also, the author believes that such training experience definitely draw benefits for teachers in their classroom teaching and testing practices.