

# Influence of Perceived Teacher Personality on Student Evaluations of Teaching: An Examination of the Variables of Students' English Proficiency and Gender

TANABE, Yoshitaka  
MORI, Setsuko  
Kindai University

## Abstract

A total of 245 students (115 male and 130 female) at a career college (*senmon gakkō*) of foreign languages in Japan completed a questionnaire regarding instruction and instructor personality. The purpose of the current study was to examine whether the relations between instructional and personality ratings and the overall evaluation of course varied by the student background characteristics, namely proficiency in English and gender. This study also investigated whether the students' instructional and/or non-instructional ratings predicted their general evaluation of course. The results of the study revealed that male students rated significantly higher than female students on one instructional factor, interest in class, and one personality factor, neuroticism. This study also observed that instructional factors significantly predicted the overall evaluation while none of the non-instructional factors, namely proficiency, gender and perceived teacher personality, did.

**Keywords:** student evaluations of teaching, perceived teacher personality, instructional ratings, non-instructional ratings, student gender

## Introduction

Over recent decades, it appears that student evaluations of teaching (hereinafter SET) have taken root at most colleges and universities throughout Japan. In order to assure educational quality, and thus provide evidence for institutional accountability, higher education institutions have come under close scrutiny from both governments and consumers alike (Grace, Weaven, Bodey, Ross, & Weaven, 2012). It seems unlikely that SET will fade away from higher education institutions anytime soon.

Along with the increasing demands for SET, there has always been concern about their reliability and validity, and numerous researchers have investigated this issue. Although such studies have seemingly provided a general consensus that “properly designed student ratings can be a valuable source of information for evaluating certain aspects of faculty teaching performance” (Chen & Hoshower, 2003, p. 73), it is not uncommon for researchers (e.g., Beren & Violato, 2005; Griffin, 2004; Olivares, 2001; Remedios & Lieberman, 2008) to report potential biases that can affect SET. In order to properly design and utilize SET, it is, therefore, essential to study possible sources of bias in the SET process to seek a better

understanding of its nature. The current study is an investigation of the effects of students' perception of the teacher's personality, their English proficiency, and gender.

### **Potential Sources of Bias in Student Evaluations of Teaching**

Among a variety of factors that can affect the validity of SET, student characteristics such as motivation, prior subject interest, and their reasons for taking a course have frequently been claimed to have significant influence on student ratings (e.g., Griffin, 2004; Chen & Hoshower, 2003; Spooen, 2010). The possible impact of students' grades, which in this paper are regarded as essentially equal to language students' proficiency in English, has also gained the attention of researchers in this field. Cashin (1990) in his review of the research, for instance, argued that students who learn more will earn higher grades and give higher ratings, and also that high motivation leads to greater learning, hence higher grades and student ratings. Likewise, Langbein (2008) and Spooen (2010) also indicated that students' course grades are one statistically significant predictor of SET. However, based on a summary of student evaluation research, Patrick (2011) pointed out that, of the 61 studies she reviewed, 24 did not find a statistically significant relationship between student grades and SET and the remainder reported a moderate average correlation of .18. Radmacher and Martin (2001) attributed the complexity of the issue to the interrelatedness of the variables, illustrating possible circumstances as follows: "the positive relationship between student grades and student evaluation of faculty may be a matter of reciprocity, but it may also be that effective teaching results in higher grades" (p. 261).

A further factor that may bias SET ratings is students' expected grades. Although the distinction between this variable and actual student grades still remains controversial (Feldman, 2007), expected grades have been a frequently investigated source of potential bias. Some researchers (e.g., Marsh, 1987; Centra, 2003; Stodnick & Rogers, 2008) assert that possible effects of students' expected grades and instructors' grading practices on SET are unlikely to be influential in the use of student ratings. Other researchers (e.g., Beran & Violato, 2005; Griffin, 2004; Olivares, 2001; Remedios & Lieberman, 2008) indicate a significant relationship between students' expected grades and SET ratings: the higher the anticipated grade, the higher the SET. Mason, Edwards, and Roach (2002), whose findings are consistent with the latter, suggested that SET may be "more a measure of their satisfaction with the rigor of grading than a valid measure of teaching effectiveness" (p. 6).

Another frequently researched issue of potential bias on SET is gender difference, the influence of which has been found to differ from study to study. Aleamoni (1999), who conducted an extensive review of previous studies on 16 common aspects of SET, cited 17 studies that reported no difference in overall ratings of instructors made by male and female students. He also cited 10 other studies that reported some difference between faculty ratings made by male and female students. He then concluded that the impacts of gender differences appear to be only partial, and, thus, skepticism on the issue is by and large a 'myth.' Similarly, Feldman (1993) reviewed past studies that claim the possible existence of a same-gender preference on overall ratings, and criticized those studies for having failed to control for such variables as subject and discipline. Richardson (2005) noted that neither the age nor gender of students related to their ratings in his study. He then asserted the result implied that the

students in his study were capable of evaluating their courses, independent of their personal circumstances.

In contrast, there have been findings in the literature indicating that gender difference has a sizeable impact on SET. For instance, Griffin, Coates, McInnis and James (2003) observed gender differences, with the tendency of female students to rate higher in response to three scales (graduate qualities, appropriate assessment, and intellectual motivation) though there was not found to be a great deal of difference between gender groups on the other eight scales they researched. This finding indicates the possibility that male and female students might perceive teaching effectiveness differently. Basow, Phelan and Capotosto (2006) found that male students tend to rate female instructors lower than female students do. Similarly, Carson (2001) qualitatively analyzed the views of female teaching staff and concluded that students generally rate male instructors higher than female ones. Tatro (1995), in contrast to Carson (2001), reported that female instructors were rated higher than male instructors, and that female students gave significantly higher ratings to female instructors than male students did.

Some researchers also focused on factors other than students' overall ratings of a course. Anderson and Siegfried (1997) examined the role of gender from the perspective of both students and instructors. They reported that female students, compared to male students, generally gave lower ratings on interest in the course content as well as other points: amount learned and importance of the course content. Darby (2006) took stereotypic gender differences into account and investigated favorable and unfavorable responses in evaluating behavior between the sexes. She found that female students evaluated courses more favorably than males when provided open-ended evaluations, and then argued that gender differences might exist in how males and females respond to their instructors. Thus, the relationship between SET and gender remains unclear.

Students' perception of instructor personality has also been identified as a possible source of bias influencing the validity of SET (e.g., Clayson, 2013; Murray, Rushton, & Paunonen, 1990; Patrick, 2011; Radmacher & Martin, 2001). The authors paid particular attention to this issue because relatively little research has been conducted in this area in the field of second/foreign language learning. Mori and Tanabe (2012) conducted a survey with 280 Japanese university students who were learning English as a general education subject and found teacher personality to have a considerable degree of influence on class evaluations. Tanabe and Mori (2013) investigated whether students' attitudes would differ depending on teachers' language background, Japanese or native English speakers, and found that the teacher personality factor held more weight for the Japanese instructors than native English speaking instructors in terms of general course evaluation ratings. In a subsequent study, Mori and Tanabe (2015) compared English major and non-English major groups to explore whether the relations between instructional and personality ratings and general course evaluation ratings varied by major. The results showed a significant correlation between the instructional scale and the overall evaluation of the course regardless of students' majors, but also indicated that while three perceived teacher personality traits, namely teacher's extroversion, thoughtfulness and neuroticism, mattered to the non-English major group, none of them was found to influence the English major group.

## Research Purposes

Based on the findings of the previous research, the primary purpose of this research is to investigate whether students' proficiency in English and gender, and their perception of teacher personality influence one another in any way, or the ratings of the instructional effectiveness. Specifically, the following research questions were formulated for this study:

1. To what degree does students' proficiency affect their ratings of classes and teachers?
2. To what degree does students' gender affect their ratings of classes and teachers?
3. To what degree do proficiency and gender predict the overall class evaluation?
4. To what degree do instructional ratings predict the overall class evaluation?
5. To what degree does students' perceived teacher personality predict the overall class evaluation?

## Method

### Participants

The participants in this study were a total of 245 students (115 male and 130 female, 123 first-year and 122 second-year students) at a career college (*senmon gakkō*) of foreign languages in Japan. Convenience sampling was used. Ages of the students were from 18 to 25. All participants were majoring in English and were in 17 different intact English classes that were randomly chosen. Focuses of those classes included vocabulary, TOEIC test preparation, writing and reading. Class sizes varied from 8 to 18 students ( $M = 13$ ). Students' English proficiency greatly varied from a low score of 150 to a high score of 980 on the TOEIC. It must be noted that the students in this study did not fit the standard model for university students in Japan. Purposes for attending the college ranged from a gateway to regular four-year colleges, preparation for full-time employment, to preparation for study abroad.

Seven male teachers and one female teacher participated in the study. Their ages ranged from 25 to 50. They were all part-time instructors at this institution and teach elsewhere either part-time or full-time. All of the instructors have master's degrees or higher, and, except for one instructor, have more than five years of experience teaching in higher education. Both students and instructors agreed to participate in this study.

### Measures

The participants completed both the instructional rating and teacher personality rating sections of the questionnaire. The instructional rating section is comprised of 24 items including one item concerning the overall evaluation of the class. These items were written in Japanese based on the Instructional Rating Form (Tomasco, 1980), and European Portfolio for Student Teachers of Languages (Newby et al., 2007) (See the Appendix A for an English translation). The teacher personality rating section consists of 28 items. All of the items were derived from Murray et al. (1990) and translated into Japanese. Although Murray's measures of personality included 29 items, one item concerned with aesthetical sensitivity was omitted as it was not relevant to the context (See the Appendix A for an English translation). Except for the item asking about students' overall evaluation of the class on a 10 point Likert scale, all

the items were on a six point Likert scale with one being strongly disagree and six being strongly agree.

### **Procedure**

Both sections of the questionnaire were administered in Week 12 in the 15-week semester. The participants were instructed to rate the class and their instructor's personality and provide information regarding their gender and TOEIC scores. In order to create an environment where the participants would be able to answer the questionnaire more freely and comfortably, the instructors were requested to leave the room. Prior to administration, the participants were told that the questionnaire was anonymous and the results would never be exposed to the instructors or used for any other purposes but for research. The questionnaire was collected by one of the students assigned by the instructor.

### **Data Analysis**

The data from the completed questionnaires was entered into a Microsoft Excel spreadsheet and checked for accuracy. First, to reduce the number of questionnaire items and assess its validity, a principal components analysis was performed. Second, a principal components analysis of Rasch residuals (Linacre, 1998) using Winsteps was conducted to ensure the unidimensionality of the subscales of the instructional and teacher personality scales as well as the model-data fit of the scales. Third, to determine the effects of proficiency and gender on the instructional and teacher personality scales, multivariate analysis of variance (MANOVA) was performed. Self-reported TOEIC scores were used as an indicator of proficiency. TOEIC scores of the low proficiency group were from 150 to 375 ( $M = 282$ ,  $SD = 52.54$ ) whereas the scores of the high proficiency group were from 380 to 980 ( $M = 545$ ,  $SD = 122.59$ ). The result of the two-sample t-test indicated that the two groups were statistically significantly different ( $p = 0.00$ ). Finally, a multiple regression analysis was run to examine the degree to which English proficiency, gender, and teacher personality factors predict overall class evaluation.

## **Results**

### **Principal Components Analyses of the Instructional and Teacher Personality Ratings**

After omitting 52 cases with missing data, 193 SET responses (96 male and 97 female students) were analyzed (See Appendix B and C for means and standard deviations of the instructional and teacher personality rating sections of the questionnaire).

In order to reduce the instructional items, first principal components analysis was performed. Four criteria were used to determine the number of factors to rotate: a minimum eigenvalue of 1.0, the scree test, a minimum loading of .45, and the interpretability of the factor solution. Based on these criteria, two factors were rotated using a Varimax rotation procedure. Two interpretable factors, *interest in class* (12 items) and *class management* (11 items), were found. Principal components analysis was performed with the personality ratings section as well. The same criteria were used to determine the number of factors to rotate for this section. Based on these criteria, four factors were rotated using a Varimax rotation procedure. Four interpretable factors, *neuroticism* (10 items), *agreeableness* (10 items),

*achievement* (4 items), and *diffidence* (4 items), were found.

To ensure the unidimensionality of the subscales of the instructional and teacher personality scales, a principal components analysis of Rasch residuals (Linacre, 1998) using Winsteps was also conducted on each subscale. Table 1 shows the eigenvalues of Rasch dimension and the first contrast for each subscale. The Eigenvalues of the first contrast for all the subscales were all 2.0 or less, which means that the items in those subscales measure a single construct (Linacre, 2006).

Table 1  
*The Eigenvalues of Rasch Dimension and the First Contrast for Subscales (n = 193)*

Subscale	Eigenvalue	
	Rasch Dimension	First Contrast
Class Interest	18.3	2.0
Class Management	8.4	2.0
Neuroticism	9.9	2.0
Agreeableness	8.0	1.7
Achievement	4.3	1.4
Diffidence	3.4	1.7

In order to determine whether all participants' responses fit the model's expectations, a Rasch person fit analysis was conducted. The Rasch reliability of person responses was estimated at .90 with a Rasch person separation of 3.08, which indicates that the instrument is sufficiently sensitive to distinguish participants who score high and those who score low. There was only one misfitting person response (a male student of low proficiency) where misfit was defined as less than 0.50 and greater than 1.50 mean squared. After deleting the misfitting person response, Rasch item fit analysis was conducted. The Rasch item reliability of the instructional and teacher personality sections was .94 and .99, and the Rasch item separation was 4.14 and 11.04, respectively, demonstrating that both sections of the questionnaire are highly reliable.

Additionally, standard fit statistics (i.e., Infit and Outfit MNSQ) were computed for each item to inspect the model-data fit of the scale. Linacre (2006) suggested that MNSQ falling in the range of 0.5–1.5 indicates a productive measurement. In the first round of analysis, the values of Infit and Outfit MNSQ for all items were greater than 0.5 and less than 1.5 except item 14 in the instructional scale (The examinations were fair), and item 1 in the personality scale (The teacher is unassertive). After deleting those two misfitting items, the data were input into the Rasch model again. The remaining 49 items exhibited good fit. The Wright maps for the six subscales were also observed to assess targeting, but not included in this paper due to space limitations. The average measures for the persons for the six subscales were all close to zero, suggesting that the items were well matched to the respondents. Based on the Wright maps, the items were reasonably spread, and performing as expected.

A principal components analysis was performed with the new set of data to obtain factor scores for further analyses. For the instructional scale, the result found two interpretable

factors, *interest in class* and *class management*, which accounted for 64.24% of the variance (See Tables 2 and 3). For the teacher personality scale, the results found four interpretable factors, *neuroticism*, *agreeableness*, *achievement*, and *diffidence*, which accounted for 54.24% of the variance (See Tables 4 and 5). Please note that personality item 2, ambitious, received strong close loadings for agreeableness and achievement. Although item 2 is labeled as ambitious, it actually states that the teacher is ambitious and has high expectations for students, and is not directly concerned with teacher's achievement. Thus, it was analyzed as an item measuring agreeableness.

Table 2  
*Principal Components Analysis Summary for the Instructional Rating Section:  
Eigenvalues and Percent of Variance Explained (N = 192)*

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	12.71	57.80	56.45
2	1.42	6.44	64.24

Table 3  
*Principal Components Analysis Results for the Instructional Rating Section (N = 192)*

	Factor			$h^2$
	Interest in Class	Class Management		
1. Arouses interest	<b>0.77</b>	0.36		0.72
2. Expands viewpoints	<b>0.80</b>	0.27		0.71
3. Informative lectures	<b>0.65</b>	0.45		0.62
4. Interprets clearly	<b>0.64</b>	0.49		0.65
5. Useful examples	<b>0.69</b>	0.39		0.63
6. Inspires confidence	<b>0.80</b>	0.27		0.71
7. Encourages initiative	<b>0.73</b>	0.32		0.63
8. Provides new tools	<b>0.78</b>	0.23		0.65
9. Stimulates thinking	<b>0.76</b>	0.43		0.76
17. Challenges students	<b>0.69</b>	0.51		0.73
18. Motivates students	<b>0.74</b>	0.42		0.72
23. Challengeable assignments	<b>0.62</b>	0.38		0.53
10. Organized presentation	0.43	<b>0.70</b>		0.68
11. Uses time effectively	0.43	<b>0.70</b>		0.67
12. Respects opinions	0.19	<b>0.71</b>		0.54
13. Sensitivity	0.26	<b>0.73</b>		0.61
15. Progress report	0.42	<b>0.74</b>		0.72
16. Class preparation	0.30	<b>0.80</b>		0.72
19. Good atmosphere	0.50	<b>0.61</b>		0.62
20. Clear rules	0.37	<b>0.55</b>		0.44
21. Effective materials	0.37	<b>0.66</b>		0.57
22. Clear evaluation	0.29	<b>0.66</b>		0.51

Table 4

*Principal Components Analysis Summary for the Personality Rating Section:  
Eigenvalues and Percent of Variance Explained (N = 192)*

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	6.24	23.09	23.09
2	4.89	18.10	41.20
3	2.17	8.04	49.24
4	1.35	5.00	54.24

Table 5

*Principal Components Analysis Results for the Teacher Personality Rating Section (N = 192)*

	Neuroticism	Agreeableness	Achievement	Diffidence	$h^2$
4. Aggressive	<b>0.76</b>	-0.14	0.03	-0.06	0.61
7. Seeks definiteness	<b>0.48</b>	0.16	0.41	-0.06	0.42
8. Defensive	<b>0.76</b>	-0.04	0.07	0.10	0.60
9. Dominant	<b>0.84</b>	-0.02	-0.02	0.12	0.72
11. Attention-seeking	<b>0.54</b>	0.19	-0.16	0.42	0.53
13. Impulsive	<b>0.63</b>	0.12	-0.32	0.31	0.61
20. Anxious	<b>0.66</b>	-0.29	-0.02	0.27	0.58
25. Compulsive	<b>0.52</b>	-0.07	0.46	-0.03	0.48
26. Authoritarian	<b>0.72</b>	-0.10	0.04	0.07	0.54
28. Neurotic	<b>0.64</b>	-0.13	-0.16	0.30	0.54
2. Ambitious	-0.07	<b>0.46</b>	<b>0.45</b>	0.03	0.42
3. Sociable	-0.37	<b>0.65</b>	0.07	0.04	0.57
5. Independent	0.29	<b>0.55</b>	0.03	0.05	0.39
6. Changeable	-0.11	<b>0.67</b>	0.28	-0.08	0.55
14. Supporting	-0.35	<b>0.57</b>	0.22	0.09	0.50
16. Fun-loving	-0.16	<b>0.77</b>	-0.11	0.19	0.66
19. Intellectually curious	-0.02	<b>0.63</b>	0.33	-0.08	0.51
22. Liberal	0.07	<b>0.73</b>	0.26	-0.07	0.60
24. Objective	-0.29	<b>0.51</b>	0.38	-0.02	0.49
27. Extraverted	0.01	<b>0.72</b>	-0.06	0.23	0.57
10. Enduring	-0.01	0.24	<b>0.62</b>	0.21	0.48
15. Orderly	-0.14	0.04	<b>0.69</b>	0.09	0.50
21. Intelligent	0.02	0.45	<b>0.47</b>	-0.28	0.50
23. Shows leadership	0.19	0.38	<b>0.54</b>	-0.11	0.47
12. Harm-avoiding	0.25	-0.11	0.13	<b>0.61</b>	0.47
17. Approval-seeking	0.02	0.24	0.13	<b>0.77</b>	0.67
18. Seeks help and advice	0.26	0.04	-0.09	<b>0.78</b>	0.68

## Research Questions 1 and 2: The Effect of Proficiency and Gender Differences on Instructional and Teacher Personality Factors

A one-way MANOVA was performed to examine the effect of proficiency (low and high) and gender (male and female) on the two instructional factors and four teacher personality factors. As Table 6 shows, the results for the MANOVA indicated a significant main effect for gender on the dependent variables, Wilks's  $\Lambda = .89.5$ ,  $F(6, 183) = 3.595$ ,  $p < .00$ ,  $\eta^2 = .105$ . Neither a significant main effect for proficiency nor a significant interaction between proficiency and gender was seen.

Table 6

*The Summary of MANOVA Results with Proficiency and Gender as Independent Variables, and Instructional and Personality Factors as Dependent Variables (N = 192)*

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	1	.009b	6	183	1	0
Proficiency	0.954	1.463b	6	183	0.193	0.046
Gender	0.895	3.595b	6	183	0.002	0.105
Proficiency * Gender	0.97	.929b	6	183	0.475	0.03

Given the significance of the overall test, univariate main effects were examined. Significant main effects for gender were obtained for interest in class,  $F(1, 188) = 11.29$ ,  $p < .001$  and neuroticism,  $F(1, 188) = 5.09$ ,  $p < .025$ . The comparisons of means indicated that male students rated significantly higher than female students on interest in class and neuroticism.

## Research Questions 3, 4 and 5: Significant Predictors of Overall Evaluation

To predict the general impression of the course, standard multiple regression analysis was employed between the overall rating as a dependent variable, and TOEIC scores, gender, the factor scores of the instructional and teacher personality components as independent variables. These variables significantly predicted the general impression of the course,  $F(8, 314) = 26.88$ ,  $p < .00$ ,  $R^2 = .54$ . Standard multiple regression was chosen over sequential and statistical multiple regression as it is more suitable when there is no preferable order for entering each independent variable.

As shown in Table 7, two of the eight variables, namely interest in class and class management, significantly predicted the overall class evaluation at  $p < .001$ . The instructional factors together accounted for 52.6% of the variance of the general impression of the course. The result suggests that when students find the class interesting and feel that the class is handled appropriately, they tend to give a higher score on overall impression of the course. The result also shows that proficiency, gender and teacher personality did not contribute to the overall evaluation of the class.

Table 7

*Summary of Multiple Regression Analysis*

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
(Constant)	8.214	0.371		22.146	0	7.482	8.945
TOEIC	-0.207	0.18	-0.06	-1.155	0.249	-0.562	0.147
Gender	-0.043	0.185	-0.012	-0.233	0.816	-0.409	0.323
Interest in class	0.987	0.103	0.566	9.593	0	0.784	1.19
Class management	0.829	0.104	0.476	8.004	0	0.624	1.033
Neuroticism	0.148	0.096	0.085	1.535	0.127	-0.042	0.337
Agreeableness	0.061	0.099	0.035	0.612	0.541	-0.135	0.256
Achievement	0.023	0.093	0.013	0.25	0.803	-0.16	0.207
Diffidence	0.088	0.093	0.05	0.937	0.35	-0.097	0.272

### Conclusion

The present study attempted to answer five questions. With regards to students' proficiency and gender differences in rating classes and teachers, which were Research Questions 1 and 2, the results highlighted two features concerning gender differences: male students rated significantly higher than female students on interest in class and neuroticism. In terms of interest in class, this seems to be somewhat relevant to the results reported by Anderson and Siegfried (1997) that claimed female students, compared to male students, generally gave lower ratings on interest in the course content and other dimensions: amount learned and importance of the course content. As for the result of male students scoring neuroticism higher than females, Darby (2006), who argued that female students evaluated certain aspects of the courses more 'favorably,' can be referred to. If neuroticism in the present study is interpreted as an 'unfavorable' side of perceived teacher personality, the current result may provide some insights into her argument; the impact of negative, or unfavorable, aspects could be stronger in male students' perceptions of teacher personality, and this could consequently affect the evaluating behavior of students. Therefore, if the findings above are valid, then students' gender differences in rating classes and teachers may be regarded as a predictable bias in course evaluation responses. As pointed out earlier in this paper, however, it should also be noted that a number of studies drew inconsistent findings, and gender differences can have an impact on evaluating certain, or only limited, aspects of courses as Feldman (1993) and Aleamoni (1999) asserted in their comprehensive reviews.

The results of Research Questions 3, 4 and 5 revealed that both instructional factors, namely interest in class and class management, significantly predicted the overall evaluation while none of the non-instructional factors, i.e., proficiency, gender and teacher personality, contributed to the overall rating of the class. In a way, the current study resulted in an ideal outcome if its primary motivation was to demonstrate the validity of SET though a number of previous studies including the authors' reported potential influences of non-instructional factors. Particularly, since the possible effects of teacher personality were consistently found

to be significant in all the prior studies conducted by the authors (Mori & Tanabe, 2012, 2015; Tanabe & Mori, 2013), it was hypothesized that the same or similar results would confirm the trend. Given that the same questionnaire items were adopted in their related studies in order to allow them to compare the results in their long-term research plan, the current results indicate that more careful attention needs to be directed to factors relevant to the participants in their future research.

From another point of view, the results might imply a need for further investigation in terms of gender symmetry. According to Feldman's (2007) overview of related research, there is some evidence across studies to suggest that "students may rate same-gendered teachers a little more highly than they do opposite-gendered teachers" (p. 97). This tendency is also reported in Basow and Silberg (1987) and Martin (1984). Given that the gender ratio of the instructors rated in the present study was seven male to one female, a different result could have been seen if the tendency had also applied to this study. Since it was impossible in practice to make the gender ratio of instructors even, the average overall rating given by male students could have turned out to be higher than that of the female students due to the uneven gender ratio of instructors; there being more male instructors. Alternatively, therefore, the results might suggest that students' possible preferences for same-gendered teachers observed in the earlier studies, if any, have only limited impact on SET and that the impact was so little as to be statistically insignificant, which can accordingly support the above-mentioned argument of Feldman (1993) and Aleamoni (1999). In order to give a clearer explanation for the findings, further investigation needs to be conducted using the same research method as the present study and a different ratio of instructors to be rated.

In addition to the findings on the research questions discussed above, it might also be worth comparing the results of the principal components analyses of the present survey done at a career college (*senmon gakkō*) and the authors' prior studies whose participants were university students (Mori & Tanabe, 2012, 2015; Tanabe & Mori, 2013). Similar, though not exactly the same, factors were found in both of the studies despite the fact that the official purpose of English education is slightly different at the two types of institution; the former for career development and the latter for academic purposes. This might imply that the findings of the current study can be applied to university settings as well.

Finally, given that the data was collected from only one institution, and the results revealed some limitations derived from the exploratory nature of the present study, more cases should be analyzed, and qualitative data analysis also needs to be integrated in the future to generalize and verify the results. While some researchers such as Marsh (2007) assert SET to be "relatively valid against a variety of indicators of effective teaching, relatively unaffected by a variety of potential biases" (p. 374), and place more emphasis on the appropriate integration of SET into faculty development, the authors believe that the question of possible sources of bias on SET still needs continuing exploration to grasp their complexity and to utilize them as a useful source of information about a teacher's performance.

### **Acknowledgements**

The authors are thankful to three anonymous reviewers from JACET Journal for their constructive comments that helped improve the manuscript.

## References

- Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education* 13(2), 153–166.
- Anderson, H. A., & Siegfried, J. J. (1997). Gender differences in rating the teaching of economics. *Eastern Economic Journal*, 23(3), 347–357.
- Basow, S. A., & Silberg, N. T. (1987). Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology*, 79, 308–314.
- Basow, S. A., Phelan, J., & Capotosto, L. (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly*, 30, 25–35.
- Beran, T., & Violato, C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education*, 30, 593–601.
- Carson, L. (2001). Gender relations in higher education: exploring lecturers' perceptions of student evaluations of teaching. *Research Papers in Education* 16(4), 337–358.
- Cashin, W. E. (1990). Student ratings of teaching: A summary of the research. *Management Newsletter*, 4(1), 2–7.
- Centra, J. A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*, 44, 495–518.
- Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment and Evaluation in Higher Education*, 28(1), 71–88.
- Clayson, D. E. (2013). Initial impressions and the student evaluations of teaching. *Journal of Education for Business*, 88, 26–35.
- Darby, J.A. (2006). Evaluating courses: An examination of the impact of student gender. *Educational Studies*, 32(2), 187–199.
- Feldman, K. A. (1993). College students' views of male and female college teachers: Part II: Evidence from students' evaluations of their classroom teachers. *Research in Higher Education*, 34, 151–211.
- Feldman, K. A. (2007). Identifying exemplary teachers and teaching: Evidence from student ratings. In R.P. Perry & J.C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 93–143). New York, NY: Springer.
- Grace, D., Weaven, S., Bodey, K., Ross, M., & Weaven, K. (2012). Putting student evaluations into perspective: The course experience quality and satisfaction model (CEQS). *Studies in Educational Evaluation*, 38(2), 35–43.
- Griffin, P., Coates, H., McInnis, H., & James, R. (2003). The development of an extended course experience questionnaire. *Quality in Higher Education*, 9(3), 259–266.
- Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*, 29, 410–425.
- Langbein, L. (2008). Management by results: Student evaluation of faculty teaching and the mis-measurement of performance. *Economics of Education Review*, 27, 417–428.
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, 2(3), 266–283.

- Linacre, J. M. (2006). A user's guide to WINSTEPS/MINISTEP: Rasch-model computer programs. Chicago, IL: Winsteps.com.
- Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, 11, 253–388.
- Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases and usefulness. In R.P. Perry & J.C. Smart (Eds.), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective* (pp. 319–383). New York, NY: Springer.
- Martin, E. (1984). Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Signs: Journal of Women in Culture and Society*, 9(3), 482–492.
- Mason, K. H., Edwards, R. R., & Roach, D. W. (2002). Student evaluation of instructors: A measure of teaching effectiveness or of something else? *Journal of Business Administration Online*, 1(2), 1–11.
- Murray, H., Rushton, P., & Paunonen, S. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82(2), 250–261.
- Mori, S., & Tanabe, Y. (2012). Effects of perceived teacher personality on student class evaluations. *Kinki University Center for Liberal Arts and Foreign Language Education Journal (Foreign Language Edition)*, 3(1), 163–181.
- Mori, S., & Tanabe, Y. (2015). Influence of instructor personality on student evaluation of teaching: A Comparison between English majors and non-English majors. *English Language Teaching*, 8(1), 1–10.
- Newby, D., Allan, R., Fenner, A., Jones, B., Komorowska, H., & Soghikyan, K. (2007). European portfolio for student teachers of languages: A reflection tool for language teacher education. Retrieved from [http://archive.ecml.at/mtp2/publications/C3\\_Epostl\\_E\\_internet.pdf](http://archive.ecml.at/mtp2/publications/C3_Epostl_E_internet.pdf).
- Olivares, O. J. (2001). Student interest, grading leniency, and teacher ratings: A conceptual analysis. *Contemporary Educational Psychology*, 26, 382–399.
- Patrick, C. L. (2011). Student evaluations of teaching: Effects of the Big Five personality traits, grades and the validity hypothesis. *Assessment & Evaluation in Higher Education*, 36, 239–249.
- Radmacher, S. A., & Martin, D. J. (2001). Identifying significant predictors of student evaluations of faculty through hierarchical regression analysis. *The Journal of Psychology*, 135(3), 259–268.
- Remedios, R., & Lieberman, D. A. (2008). I liked your course because you taught me well: The influence of grades, workload, expectations and goals on students' evaluations of teaching. *British Educational Research Journal*, 34, 91–115.
- Richardson, J. (2005). Students' perceptions of academic quality and approaches to studying in distance education. *British Educational Research Journal* 31(1), 7–27.
- Spooren, P. (2010). On the credibility of the judge. A cross-classified multilevel analysis on student evaluations of teaching. *Studies in Educational Evaluation*, 36, 121–131.
- Stodnick, M., & Rogers, P. (2008). Using SERVQUAL to measure the quality of the classroom

- experience. *Decisions Sciences Journal of Innovative Education*, 6, 115–133.
- Tanabe, Y., & Mori, S. (2013). Effects of perceived teacher personality on student class evaluations: A comparison between Japanese instructors and native English speaking instructors. *International Journal of English Linguistics*, 3(3), 53–65.
- Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *Journal of Research and Development in Education*, 28(3), 169–173.
- Tomasco, A. (1980). Student perceptions of instructional and personality characteristics of faculty: A canonical analysis. *Teaching of Psychology*, 7(2), 79–83.

## Appendices

### Appendix A

Translation of Instructional Rating and Teacher Personality Sections of the Questionnaire

In this questionnaire, you are asked about this class and your impression on the instructor.

When answering the questions, please keep in mind the following two points:

\*The results of the questionnaire will never be exposed to the instructor. Your personal information will not be provided to the instructor.

\*Although the results will be statistically analyzed and reported at a conference and/or in a journal, the responses to individual items will not be disclosed.

Read each statement and circle the number to indicate the extent to which you agree or disagree with each statement.

- |                     |                     |                  |
|---------------------|---------------------|------------------|
| 1 Strongly disagree | 2 Somewhat disagree | 3 Disagree       |
| 4 Agree             | 5 Somewhat agree    | 6 Strongly agree |

#### A. About this class

---

1. This class aroused my interest.
  2. This class expanded my viewpoints.
  3. This class provided me with useful knowledge.
  4. This class was clear and easy to understand.
  5. This class offered useful examples.
  6. This class inspired my confidence.
  7. This class encouraged me to take an initiative in learning.
  8. This class provided me with new learning tools.
  9. The content of this class stimulated my intellectual curiosity.
  10. This class was clearly organized.
  11. Time was effectively used in this class.
  12. The instructor respected students' opinions.
  13. The instructor tried to meet students' needs.
  - \*14. The examinations were fair.
  15. The objects of the class were clear and reasonable.
  16. The teacher was well prepared for the class.
  17. The content of this class was challenging.
  18. This class motivated me to study English.
  19. The class atmosphere was appropriate.
  20. This class was disciplined.
  21. Materials and instruments used for this class were effective.
  22. Evaluation criteria for presentations and assignments were clearly explained.
  23. Presentation tasks and assignments were challenging.
  24. Give your overall evaluation to this class on a scale from 1 to 10.
- 

\*The item was dropped from the final analyses.

## B. About the instructor

---

- \*1. unassertive and conformable
  2. ambitious and have high expectations for the students
  3. friendly and sociable
  4. argumentative and gets angry easily
  5. avoids restraints and enjoys being free
  6. flexible, and likes new and different ideas
  7. does not like ambiguity
  8. suspicious and takes offense easily
  9. forceful and attempts to control environment
  10. patient and enduring
  11. attention-seeking
  12. careful and avoids risks
  13. impulsive and reckless
  14. supporting and gives sympathy
  15. neat and organized
  16. fun-loving
  17. works for approval of others
  18. desires support and sympathy from others
  19. intellectually curious and insightful
  20. nervous and unstable
  21. bright and intelligent
  22. liberal
  23. shows leadership
  24. fair and free of bias
  25. meticulous and perfectionistic
  26. authoritarian and opinionated
  27. extraverted and optimistic
  28. moody and constantly worried things will go wrong
- 

\*The item was dropped from the final analyses.

Appendix B

*Descriptive Statistics and Mean Differences of the Instructional Rating Sections*

	Male (n = 96)				Female (n = 97)			
	Low (n = 51)		High (n = 45)		Low (n = 43)		High (n = 54)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1. Arouses interest	4.88	1.11	4.78	1.22	4.56	1.05	4.24	1.10
2. Expands viewpoints	4.88	1.07	4.58	1.32	4.58	0.98	4.20	1.09
3. Informative lectures	5.04	1.11	5.16	0.88	4.95	0.84	4.80	0.83
4. Interprets clearly	5.06	1.05	4.82	1.21	5.02	0.91	4.52	1.06
5. Useful examples	4.96	1.02	4.71	1.29	4.88	0.91	4.31	1.19
6. Inspires confidence	4.35	1.31	4.31	1.28	3.95	0.95	3.94	1.16
7. Encourages initiative	4.59	1.06	4.40	1.12	4.00	1.15	4.06	1.12
8. Provides new tools	4.35	1.28	4.29	1.34	3.93	1.06	3.67	1.44
9. Stimulates thinking	4.65	1.15	4.49	1.24	4.51	1.03	3.94	1.16
10. Organized presentation	4.86	1.08	4.78	1.15	4.65	0.90	4.52	1.04
11. Uses time effectively	4.76	1.16	4.69	1.14	4.72	0.91	4.37	1.12
12. Respects opinions	4.84	1.05	5.02	1.03	4.77	0.84	4.76	1.08
13. Sensitivity	4.86	1.08	4.93	1.05	5.05	0.84	4.80	1.22
14. Fair examinations	4.71	1.19	4.76	1.26	4.58	0.85	4.80	1.05
15. Progress report	4.73	1.06	4.87	1.01	4.79	0.91	4.46	0.97
16. Class preparation	4.73	1.04	5.13	1.04	5.05	0.75	4.78	0.92
17. Challenges students	4.63	1.22	4.71	1.20	4.74	0.95	4.17	1.34
18. Motivates students	4.71	1.22	4.60	1.21	4.56	0.93	4.20	1.28
19. Good atmosphere	4.55	1.19	4.93	1.14	4.72	0.85	4.41	1.21
20. Clear rules	4.61	1.18	5.04	1.00	4.67	0.97	4.70	0.94
21. Effective materials	4.80	1.15	4.87	1.10	5.00	0.69	4.76	1.03
22. Clear evaluation	4.71	1.15	4.96	1.02	4.63	0.76	4.52	1.14
23. Challengeable assignments	4.59	1.24	4.44	1.31	4.09	0.92	4.13	1.32
24. Overall evaluation	8.16	1.69	8.04	1.87	7.79	1.60	7.41	1.74

*Note.* Low = Low proficiency group with 150–375 on the TOEIC. High = High proficiency group with 380–980 on the TOEIC.

Appendix C

*Descriptive Statistics and Mean Differences of the Teacher Personality Rating Sections*

	Male (n = 96)				Female (n = 97)			
	Low (n = 51)		High (n = 45)		Low (n = 43)		High (n = 54)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1. Unassertive	4.08	1.32	3.89	1.64	3.72	1.80	4.62	1.18
2. Ambitious	4.75	1.02	4.96	1.00	4.60	1.07	4.55	1.05
3. Sociable	5.14	1.04	5.04	1.07	5.14	0.89	5.04	1.06
4. Aggressive	2.37	1.37	1.89	1.13	2.05	1.33	1.83	1.07
5. Independent	3.78	1.33	3.16	1.36	3.60	1.26	3.15	1.31
6. Changeable	4.80	1.13	4.76	1.03	4.60	0.93	4.21	1.13
7. Seeks definiteness	3.78	1.27	3.56	1.34	3.79	1.26	3.09	1.23
8. Defensive	2.80	1.52	2.33	1.35	2.30	1.35	2.04	1.13
9. Dominant	2.53	1.33	2.00	1.24	2.40	1.40	1.77	1.05
10. Enduring	4.14	1.10	4.13	1.25	4.07	1.14	4.08	1.21
11. Attention-seeking	2.78	1.29	2.18	1.17	2.44	1.16	2.30	1.10
12. Harm-avoiding	3.63	1.28	2.69	1.18	2.91	1.06	2.92	1.19
13. Impulsive	2.90	1.45	2.22	1.31	2.58	1.35	1.92	0.92
14. Supporting	4.84	0.99	4.67	1.30	4.70	1.01	4.75	1.09
15. Orderly	4.82	0.95	4.67	1.45	4.81	1.05	4.49	1.03
16. Fun-loving	4.94	1.16	4.22	1.20	4.72	0.88	4.42	1.35
17. Approval-seeking	3.86	1.17	3.71	1.41	3.30	1.12	3.45	1.29
18. Seeks help and advice	3.18	1.23	2.82	1.28	2.79	1.06	2.74	1.29
19. Intellectually curious	4.73	1.08	4.29	1.31	4.58	0.98	4.26	1.06
20. Anxious	2.57	1.35	2.09	1.20	2.21	1.17	1.81	1.04
21. Intelligent	4.51	1.24	4.20	1.50	4.44	1.24	4.45	1.03
22. Liberal	4.71	1.10	4.20	1.10	4.21	1.08	4.13	1.23
23. Shows leadership	4.25	1.18	3.60	1.39	4.14	1.15	3.49	1.37
24. Objective	4.78	1.12	4.71	1.32	4.77	0.92	4.98	0.93
25. Compulsive	3.67	1.21	3.72	1.18	3.60	1.34	2.75	1.25
26. Authoritarian	2.82	1.21	2.33	1.46	2.58	1.16	1.91	1.06
27. Extraverted	4.37	1.22	3.91	1.38	4.37	1.16	4.19	1.24
28. Neurotic	2.16	1.35	1.93	1.05	1.93	1.18	1.64	0.90