English Language Teaching; Vol. 8, No. 1; 2015
ISSN 1916-4742 E-ISSN 1916-4750
Published by Canadian Center of Science and Education

Influence of Instructor Personality on Student Evaluation of Teaching: A Comparison between English Majors and Non-English Majors

S. Mori[1] & Y. Tanabe[1]

[1] Faculty of Law, Kinki University, Japan

Correspondence: S. Mori, Professor, Faculty of Law, Kinki University, 3-4-1 Kowakae, Higashiosaka City, Osaka 577-8502, Japan. Tel: 81-6-4307-3081. E-mail: setsukomori@jus.kindai.ac.jp

Received: September 15, 2014 Accepted: October 16, 2014 Online Published: December 17, 2014

doi:10.5539/elt.v8n1p1 URL: http://dx.doi.org/10.5539/elt.v8n1p1

Abstract

160 non-English major students studying at a four-year university and 193 English major students studying at a career college of foreign language in Japan completed a questionnaire regarding instruction and instructor personality. The purpose of the current study was to examine whether the students' instructional and personality ratings predicted their general evaluation of course. This study also investigated whether the relations between instructional and personality ratings and the general course evaluation varied by major. A significant correlation was found between the instructional scale and the overall evaluation of the course regardless of students' majors: The more the students found the class interesting and was appropriately managed, the higher the overall evaluation. However, the findings indicate that while teacher's extroversion, thoughtfulness and neuroticism mattered to the non-English major group when evaluating the overall effectiveness of the class, teacher personality did not influence the English major group. The authors believe the findings of the present study could contribute to a better understanding of the nature of student evaluations that have always been a source of controversy, and sometimes discontent during their history.

Keywords: student evaluation of teaching, bias on SET, instructor personality, instructional scale, instructor personality scale

1. Introduction

1.1 Possibility of Biases in Student Evaluation of Teaching

At the majority of universities, student evaluation of teaching (SET) is regarded as important and widely used for the purpose of improving the quality of instruction and providing information for instructor appraisal as well as providing evidence for institutional accountability (Spoorenm, Brockx, & Mortelmans, 2013). However, despite the increasing popularity of, and demands for SET, much concern has been expressed regarding potential biases that may affect SET. Considering the importance attached to SET, it is imperative to study possible sources of biases in the SET process.

1.2 Grading Practices and SET

Of a variety of factors that can affect the validity of SETs, most research has investigated possible effects of students' expected grades and instructors' grading practices on students' ratings of instruction (e.g., Beran & Violato, 2005; Greenwald & Gillmore, 1997; Griffin, 2004; McPherson, Todd, Jewell, & Kim, 2009; Olivares, 2001; Remedios & Lieberman, 2008). Many researchers (e.g., Beren & Violato, 2005; Griffin, 2004; Maurer, 2006; McPherson, Todd, Jewell, & Kim, 2009; Olivares, 2001; Remedios & Lieberman, 2008) found a significant relationship between student's expected grades and SET ratings: The higher the expected grade, the higher the SET. Other researchers (Greenwald & Gillmore, 1997; Griffin, 2004; Olivares, 2001) contend that instructors can get higher SET ratings by following a more lenient grading policy.

1.3 Teacher Personality and SET

Another source of bias influencing the validity of SETs has been suggested to be students' perception of instructor personality (e.g., Clayson, 2013; Clayson & Sheffet, 2006; Hart & Driver, 1978; Murray, Rushton, & Paunonen, 1990; Patrick, 2011; Radmacher & Martin, 2001; Shelvin, Banyard, Davies, & Griffiths, 2000). Clayson and Sheffet (2006), for instance, conducted survey research using the Five Factor Model of Personality often referred to as the Big Five (Digman, 1990). The Big Five represents five dimensions of personality, namely

agreeableness, conscientiousness, stability, extroversion, and creativity (openness). The study found a consistent and positive relationship between those personality traits and course evaluations, and the association was formed within fewer than five minutes of initial contact and grew stronger over the term. Clayson (2013) later conducted a similar study and confirmed the previous finding that showed students' first impressions of the instructor influenced the final evaluation given in a class. This finding is in congruence with some evidence presented by Ortinau and Bush (1987), and Sauber and Ludlow (1988) indicating that subsequent class experience may do little to change students' initial impressions of the instructor and teaching effectiveness.

Patrick (2011) also examined whether personality traits measured by the perception of students could predict student evaluations of teachers and courses using a similar Big Five Inventory (John, Donahue, & Kentle, 1991, cited in Patrick, 2011). He found agreeableness, conscientiousness, extroversion, and openness correlated positively and neuroticism correlated negatively with student evaluations of the teachers and the courses.

In the field of second/foreign language learning, very little research has been done to investigate possible relationships between students' perception of teacher personality and evaluation of the related class. Mori and Tanabe (2012) conducted a survey with 280 Japanese university students learning English and found a considerable degree of influence of teacher personality on class evaluations. In the subsequent study, Tanabe and Mori (2013) attempted to see whether students' attitudes could differ depending on teacher nationality, 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 overall evaluations of class.

1.4 Course Characteristics and SET

Other researchers (e.g., Basow & Montgomery, 2005; Beran & Violato, 2005; Marsh & Roche, 1997; Remedios & Lieberman, 2008; Santhanam & Hicks, 2001; Ting, 2000) focus on uncovering possible effects of course characteristics such as course type, course discipline, course workload and course level on SETs. For example, the research consistently shows that natural science courses tend to receive lower SET than courses in the social sciences and humanities (Basow & Montgomery, 2005; Beran & Violato, 2005). Other studies (e.g., Petchers & Chow, 1988; Ting, 2000) show that instructors who teach elective courses are rated higher on overall evaluation than those who teach compulsory courses. Ting (2000) also investigated whether course type affects student satisfaction in courses, and found that courses with specific content matters received higher SET ratings.

1.5 Research Purposes

As mentioned above, a number of factors have been suggested as possible sources of bias in student evaluations. However, the majority of previous research was conducted with either business or psychology majors, and almost no relevant studies can be found in the field of second/foreign language learning with a few exceptions (Mori & Tanabe, 2012; Tanabe & Mori, 2013). Providing data from multidimensional perspectives should give meaningful insight into and a better understanding of the fundamental nature of SETs. Therefore, the purpose of the current study was to examine whether language learners also depend on non-instructional factors when rating the overall effectiveness of instructions. Specifically, the relation of students' assessments of instructor personality traits to course ratings was measured. This study also investigated whether such relation differs depending on whether the students are English majors or non-English majors. Since both of the authors' previous studies were conducted with groups of non-English major students in the same university, in other words, only in a limited environment, the current study included a dissimilar group of participants as well so that the results could be generalized to a broader context. In addition, the authors attempted to discuss the consistency of these findings with their previous studies since, as stated above, there are insufficient studies relevant to the field, and therefore, the integrity of their studies needs to be constantly confirmed.

1.6 Research Questions

In this study, the following research questions were investigated:

1) How is instructional scale correlated with general course evaluation?

2) What is the effect of perceived instructor personality on general course evaluation?

3) Is there a strong relationship, based on major, between instructional scale and general course evaluation?

4) Is there a strong relationship, based on major, between perceived instructor personality and general course evaluation?

2. Method

2.1 Respondents

The respondents for this study were two groups of Japanese students. The one group was comprised of 160 non-English major students studying at a four-year university (hereinafter referred to as non-major group) and the other of 193 English major students studying at a career college of foreign language (hereinafter referred to as major group).

The students in the non-major group were all majoring in law, and in seven separate English classes required for all first and second year students in this institution. The focuses of these classes were reading and listening, and the number of students in each class varied from 13 to 26. Their English proficiency ranged from 110 to 156 on the TOEIC Bridge, and from 180 to 420 on the TOEIC. The students in the major group were in a variety of English courses including vocabulary, TOEIC test preparation, writing and reading. The number of students in each class was from 8 to 18. Students' English proficiency greatly varied from 150 to 980 on the TOEIC. Purposes for attending the college ranged from a gateway to regular four-year colleges, preparation for full-time employment, to preparation for study abroad. Ages of the students were from 18 to 25.

2.2 Materials

As they completed the questionnaire comprised of two sections, the students in both groups were asked to consider their experiences with the class they were currently attending and its instructor. The first section of the questionnaire (hereinafter referred to as instructional scale) was concerned with instruction and has 24 closed-ended questions containing one item asking about their general impression of the course. The items of this section draw on the Instructional Rating Form (Tomasco, 1980), and European Portfolio for Student Teachers of Languages (Newby, Allan, Fenner, Jones, Komorowska, & Soghikyan, 2007). The second section of the questionnaire (hereinafter referred to as personality scale) contained 28 closed-ended questions about their instructor personality. All of the items of this section were based on Murray, Rushton, and Paunonen (1990). However, one item associated with aesthetical sensitivity in Murray et al's measures of personality was removed since it was irrelevant to the current context. Except for the item asking about students' general impression of the course which was on a 10 point Likert scale, questions were rated on a six-point Likert scale ranging from strongly disagree (1) to strongly agree (6) (See Tanabe & Mori for details of the questionnaire). The item asking about students' overall evaluation of the class was based on a 10-point scale in accordance with the official student ratings conducted at the institutions in which the respondents are enrolled. Internal reliability (Cronbach's alpha) of the instructional scale was .97 and of the personality scale was .81.

2.3 Procedure

The survey was conducted within the last 20 minutes of each class in the 13th week of a 15-week semester. After detailed instructions and explanations were provided either by the researcher or instructor, the questionnaire was distributed. The study was described as an investigation of students' perception of instruction and instructor personality. Before administration, the instructor was asked to leave the classroom so that the students would be able to fill out the questionnaire without feeling unnecessary pressure from the instructor. In addition, anonymity of responses was emphasized. The questionnaire was then collected either by the researcher or a student representative.

2.4 Statistical Analyses

Data was collected on a large number of instructional and personality items, 24 and 28, respectively. In order to interpret the data in a meaningful way, first, a principal components analysis was performed to identify underlying communalities among the items and, as a result, reduce the number of items. Secondly, multiple regression analysis was conducted with the factor scores obtained from principal components analysis as the independent variables and overall score of evaluation as the dependent variable to investigate which instructional and personality factors may predict the overall impression of class.

3. Results

3.1 Principal Components Analyses of the Instructional and Personality Scales

First, a principal components analysis was conducted to identify and compute composite scores for the factors underlying the instructional scale. Initial eigenvalues indicated that the first two factors explained 56.78% and 7.47%, respectively. The remaining factors had eigenvalues under one, and each explained less than four percent of the variance. Therefore, two-factor solution was examined using a Varimax rotation of the factor loading

matrix. The two-factor solution explained 64.24% of the variance (See Table 1). All of the items were kept because they met a minimum criterion of having a primary factor loading of .40 or above (See Table 2).

After a close examination of the items that loaded on each factor, Factor 1 was labeled as *interest in class* (e.g., arousal interest, informative lectures, stimulates thinking), and Factor 2 as *class management* (e.g., organized presentation, uses time effectively, clear rules). Factor scores were created for each of the two factors for further analysis.

Table 1. Principal components analysis eigenvalue summary for the instructional scale

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	13.06	56.78	56.78
2	1.717	7.47	64.24

Table 2. Factor loadings and communalities based on a principal components analysis with varimax rotation for 23 items from the instructional scale (n = 353)

	Interest in Class	Class Management	Communality
1. Arousal interest	**0.75**	0.42	0.73
2. Expands viewpoints	**0.80**	0.29	0.72
3. Informative lectures	**0.67**	0.40	0.60
4. Interprets clearly	**0.56**	0.56	0.63
5. Useful examples	**0.62**	0.48	0.61
6. Inspire confidence	**0.81**	0.26	0.72
7. Encourage initiative	**0.80**	0.26	0.71
8. Provides new tools	**0.77**	0.19	0.63
9. Stimulates thinking	**0.78**	0.38	0.75
17. Challenges students	**0.70**	0.50	0.73
18. Motivates students	**0.81**	0.34	0.78
23. Challengeable assignments	**0.56**	0.51	0.57
19. Good atmosphere	0.53	**0.58**	0.62
10. Organized presentation	0.33	**0.75**	0.66
11. Uses time effectively	0.39	**0.69**	0.63
12. Respects opinions	0.34	**0.66**	0.55
13. Sensitivity	0.35	**0.69**	0.60
14. Fair examinations	0.11	**0.68**	0.48
15. Progress report	0.35	**0.75**	0.68
16. Class preparation	0.25	**0.80**	0.70
20. Clear rules	0.39	**0.59**	0.50
21. Effective materials	0.35	**0.70**	0.61
22. Clear evaluation	0.28	**0.70**	0.57

A principal components analysis with 28 items from the personality scale was also conducted. The initial eigenvalues suggested that the first three factors explained 26.66%, 16.76% and 6.84% of the variance respectively. The fourth and fifth factors had eigenvalues of just over one, and each explained 4.49% and 3.91% of the variance. Thus, solutions for three, four and five factors were examined with a Varimax rotation. The five-factor solution, which explained 58.68% of the variance, was preferred because of its previous theoretical support (See Table 3).

Table 3. Principal components analysis eigenvalue summary for the personality scale

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	7.46	26.66	26.66
2	4.69	16.76	43.42
3	1.91	6.84	50.27
4	1.25	4.49	54.76
5	1.09	3.91	58.68

Table 4. Factor loadings and communalities based on a principal components analysis with varimax rotation for 28 items from the personality scale (n = 353)

	Neuroticism	Achievement	Extroversion	Diffidence	Thoughtfulness	Communality
4. Aggressive	**0.76**	-0.02	-0.08	-0.03	-0.16	0.60
7. Seeks definiteness	**0.51**	0.41	0.01	-0.09	-0.04	0.44
8. Defensive	**0.77**	-0.02	-0.02	0.14	-0.14	0.63
9. Dominant	**0.81**	-0.10	0.02	0.11	-0.26	0.74
11. Attention-seeking	**0.56**	-0.15	0.36	0.31	-0.12	0.57
13. Impulsive	**0.65**	-0.20	0.13	0.29	-0.08	0.57
20. Anxious	**0.71**	-0.18	-0.21	0.21	0.07	0.63
25. Compulsive	**0.57**	0.34	-0.21	-0.09	-0.07	0.50
26. Authoritarian	**0.78**	-0.04	-0.15	0.01	0.06	0.64
28. Neurotic	**0.66**	-0.27	-0.10	0.25	0.08	0.58
2. Ambitious	-0.05	**0.57**	0.31	-0.01	0.36	0.55
10. Enduring	0.01	**0.62**	0.18	0.19	0.10	0.46
15. Orderly	-0.13	**0.70**	-0.15	0.11	0.19	0.58
19. Intellectually curious	-0.11	**0.59**	0.49	0.02	-0.15	0.62
21. Intelligent	-0.14	**0.65**	0.24	-0.02	-0.16	0.52
23. Shows leadership	0.10	**0.64**	0.29	-0.04	0.09	0.52
24. Objective	-0.35	**0.52**	0.35	-0.01	0.21	0.56
3. Sociable	-0.28	0.31	**0.54**	-0.03	0.41	0.64
5. Independent	0.18	0.16	**0.58**	0.11	-0.15	0.43
6. Changeable	-0.22	0.50	**0.51**	-0.01	0.20	0.60
16. Fun-loving	-0.21	0.23	**0.69**	0.09	0.27	0.65
22. Liberal	-0.06	0.54	**0.58**	0.03	0.05	0.64
27. Extraverted	-0.07	0.11	**0.72**	0.10	0.15	0.57
12. Harm-avoiding	0.20	0.13	-0.14	**0.72**	-0.20	0.63
17. Approval-seeking	0.05	0.16	0.33	**0.69**	0.18	0.65
18. Seeks help and advice	0.25	-0.03	0.14	**0.74**	0.16	0.66
1. Meek	-0.12	0.08	0.10	0.11	**0.78**	0.65
14. Supporting	-0.30	0.32	0.44	-0.05	**0.49**	0.62

10 items loaded on Factor 1. It is clear that all relate to one of the Big Five personality traits, neuroticism (i.e., instable, anxious, moody). Thus, this factor was labeled *neuroticism*. Factor 2 was defined as *achievement* because most of the items loading on this factor (i.e., ambitious, intellectually curious, and intelligent) relate to students regarding their instructor as an achiever. Factor 3 was interpreted as *extroversion* as this factor received

high loadings from such items as sociable, fun-loving, and extroverted. Factor 4 was defined as *diffidence* based on the communality among the three items that loaded on this factor. Factor 5 obtained loadings both from meek and supporting, and thus was called *thoughtfulness*.

3.2 Significant Indicators of General Impressions of the Course

A multiple regression analysis was conducted to evaluate how well the instructional and personality scales predicted the general impression of the course. The predictors were the factor scores of the two instructional factors (interest in class and class management) and the factor scores of the five personality factors (neuroticism, achievement, extroversion, diffidence, and thoughtfulness). The criterion variable was the overall rating. The data collected from the non-major and major groups were analyzed separately as the classes and teachers that they were asked to rate were different.

3.2.1 Non-Major Group

Complete data were available for 160 participants. Correlations between the predictor variables are presented in Table 5. A multiple regression analysis was employed to predict the general impression of the course from interest in class, class management, neuroticism, achievement, extroversion, diffidence and thoughtfulness. These variables significantly predicted the general impression of the course, $F(7, 151) = 45.047$, $p < .00$, $R^2 = .676$. As shown in Table 6, three of the seven variables, namely interest in class, class management, and neuroticism, added statistically significantly to the prediction, $p < .001$. The correlation coefficient for two other personality factors, extroversion and thoughtfulness, were also significant at $p < .05$. The instructional factors together accounted for 61.7% (.593 = .35, .515 = .26) of the variance of the general impression of course whereas the personality factors contributed 48.9% (-.40 = .16, .36 = .12, .45 = .20). The result suggests that when students in the non-major group consider the course as interesting and clear and the teacher as extroverted and thoughtful, they tend to give a higher score on overall impression of the course. On the other hand, when students in this group see the teacher as compulsive and anxious, they are likely to give a lower point to overall impression of the course.

Table 5. Correlations between predictor variables for non-major group

	Interest in class	Class management	Neuroticism	Achievement	Extroversion	Diffidence
Class management	-.038					
Neuroticism	-.168*	-.322**				
Achievement	.435**	.517**	-.109			
Extroversion	.264**	.271**	-.091	.081		
Diffidence	.039	.082	-.027	.168*	.059	
Thoughtfulness	.326**	.342**	-.136	.141	.142	-.036

**$p < .01$, *$p < .05$

Table 6. Summary of multiple regression analysis for non-major group

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	8.474	0.077		110.388	0.000	8.323	8.626
Interest in class	0.667	0.107	0.446	6.250	0.000	0.456	0.878
Class management	0.522	0.124	0.331	4.218	0.000	0.277	0.766
Neuroticism	-0.277	0.082	-0.175	-3.373	0.001	-0.439	-0.115
Achievement	0.196	0.118	0.123	1.667	0.098	-0.036	0.428
Extroversion	0.171	0.089	0.103	1.921	0.049	-0.005	0.346
Diffidence	0.046	0.078	0.028	0.591	0.555	-0.108	0.201
Thoughtfulness	0.237	0.094	0.142	2.508	0.013	0.050	0.423

3.2.2 Major Group

Complete data were available for 193 participants. Correlations between the predictor variables are presented in Table 7. The result of multiple regression analysis suggests that these variables significantly predicted the general impression of the course, $F(7, 184) = 30.223$, $p < .00$, $R^2 = .535$. As shown in Table 8, in case of English major students, only instructional factors, interest in class and class management, added statistically significantly to the prediction, $p < .001$. The instructional scales together accounted for 58% (.596 = .36, .466 = .22) of the variance of the general impression of the course. On the other hand, the correlation coefficients for instructor personality factors were not significant. The result implies that although how interesting the class is and how appropriately the class is managed mattered to this group of students, unlike the non-major group, teacher personality did not influence their final evaluation of the course.

Table 7. Correlations between predictor variables for major group

	Interest in class	Class management	Neuroticism	Achievement	Extroversion	Diffidence
Class management	.095					
Neuroticism	.089	-.337**				
Achievement	.363**	.250**	.087			
Extroversion	.298**	.088	.065	-.056		
Diffidence	-.221**	-.192**	.018	-.124	-.041	
Thoughtfulness	-.157*	.084	.100	-.103	-.095	.023

**$p < .01$, *$p < .05$

Table 8. Summary of multiple regression analysis for major group

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	7.878	0.092		85.861	0.000	7.697	8.059
Interest in class	0.974	0.109	0.527	8.951	0.000	0.759	1.189
Class management	0.804	0.106	0.444	7.609	0.000	0.595	1.012
Neuroticism	0.150	0.097	0.087	1.557	0.121	-0.040	0.341
Achievement	0.030	0.100	0.017	0.301	0.764	-0.167	0.227
Extroversion	0.077	0.092	0.045	0.832	0.406	-0.105	0.258
Diffidence	0.039	0.088	0.023	0.439	0.661	-0.135	0.212
Thoughtfulness	-0.050	0.087	-0.030	-0.570	0.569	-0.222	0.122

4. Discussion and Conclusion

The present study attempted to answer four questions.

1. How is instructional scale correlated with general course evaluation? The multiple regression analyses shown in 3.2.1 and 3.2.2 have indicated that both instructional factors, interest in class and class management, significantly correlated with general course evaluation. This finding is, to a large extent, consistent with the authors' previous study (Tanabe & Mori, 2013) as well as other research such as Marsh and Roche (1997) that summarized SETs as "reliable and stable"(p. 1187). There having been a lot of controversy about the reliability and validity of SETs for many years, the result of the present study statistically proves that students generally rate the overall evaluation of the class based on their observations of the instructor's teaching, not groundlessly. Having said that, it is worth citing the following claim in Mori and Tanabe (2012) which explored the correlations between each instructional item and the overall rating; "students are not likely to give their overall impression of the class based on their inclusive observations of the teacher's instructional ratings" (p. 176). SETs could be regarded as reliable but the overall rating of his/her class evaluation may just be a reflection of limited

aspects of his/her teaching performance.

2. What is the effect of perceived instructor personality on general course evaluation? With regards to this question, the non-major group and the major provided different results, which will further be discussed in Question 4. Although no personality factors were found to correlate with general course evaluation in the major group, the correlation coefficients for both instructional factors and one personality trait, neuroticism, were found to be significant at $p < .001$, and also two other personality traits, extroversion and thoughtfulness, at $p < .05$ in the non-major group. The result is, again, consistent with the authors' previous study (Tanabe & Mori, 2013) and the result provides some support for the claim that "one cannot and should not regard student ratings as a bias-free instrument to evaluate the instructional effectiveness of a teacher" (p. 62) although it is not applicable to the finding with the major group.

3. Is there a strong relationship, based on major, between instructional scale and general course evaluation? The result in the present study revealed no difference between the two groups. In his reviews of the literature on this particular respect, Aleamoni (1999) noted, "no significant differences and no significant relationships exist between student ratings and whether students were taking a course as part of a major" (p. 157). This suggestion is not contradicted by the findings of the study. On the contrary, Benton and Cashin (2012) listed some variables that possibly correlate with SETs, one of which is students' major. They, however, analyzed some factors that might account for these variables and noted they are not necessarily to be considered biases. Thus, it could generally be said that students' major has little or no influence on student ratings.

4. Is there a strong relationship, based on major, between perceived instructor personality and general course evaluation? Unlike Question 3, the non-major and major groups returned different results: the former showed significant correlations between three personality traits, neuroticism, extroversion and thoughtfulness and the general course evaluation while the latter indicated no correlations. The result differs slightly from previous studies on the subject. Spooren et al. (2013), for instance, provides an overview of previous studies that addresses student-related, teacher-related, and course-related characteristics that might affect SETs. They observe some are meaningful indicators of student learning and are therefore logically related to effective teaching and SETs, but according to the overview, the majority of research shows that a number of biasing factors including instructor's personality traits possibly underlie SETs. Clayson and Sheffet (2006) even imply that "students universally are associating perceived personality with instructional effectiveness" (p. 156). It may, thus, be worth clarifying the differences between the non-major and the major groups to seek a possible interpretation for the inconsistency with the previous studies. However, with statistical information on the participants being limited, there is no basis for determining what could be a key factor in explaining the result. Although the authors assume that the most plausible characteristic for the differences between the two groups could be the degree of motivation for studying English and the presence of goal setting, further investigation is required to prove this assumption.

As a conclusion of the present study, it was confirmed that instructional factors significantly correlated with general course evaluation, which consequently assures validity of SETs. It also revealed a need to further investigate why students' perceived teacher personality traits, possible sources of bias, did not influence student ratings in one of the two groups. The data collected in this survey is limited and a further study may be required on a broader scale, but the authors hope that the present study has offered some valuable implications for attempting to grasp the complexity of SETs.

References

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personal Evaluation in Education*, *13*(2), 153-166. http://dx.doi.org/10.1023/A:1008168421283

Basow, S. A., & Montgomery, S. (2005). Student ratings and professor self-ratings of college teaching: Effects of gender and divisional affiliation. *Journal of Personnel Evaluation in Education*, *18*, 91-106. http://dx.doi.org/10.1007/s11092-006-9001-8

Benton, S. L., & Cashin, W. E. (2012). Student ratings of teaching: A summary of research and literature. *IDEA Paper*, *50*, 1-20.

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. http://dx.doi.org/10.1080/02602930500260688

Clayson, D. E. (2013). Initial Impressions and the student evaluations of teaching. *Journal of Education for Business*, *88*, 26-35. http://dx.doi.org/10.1080/08832323.2011.633580

Clayson, D. E., & Sheffet, M. J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education*, *28*, 149-160. http://dx.doi.org/10.1177/0273475306288402

Digman, J. M. (1990). Personality structure: An emergence of the five factor model. *The Annual Review of Psychology*, *41*, 417-440. http://dx.doi.org/10.1146/annurev.ps.41.020190.002221

Divorky, J. J., & Rathermel, M. A. (1988). Student perceptions of the relative importance of dimensions of teaching performance across type of class. *Educational Research Quarterly*, *12*, 40-5.

Greenwald, A. G., & Gillmore, G. M. (1997). No pain, no gain? The importance of measuring course workload in student ratings of instruction. *Journal of Educational Psychology*, *89*, 743-751. http://dx.doi.org/10.1037/0022-0663.89.4.743

Griffin, B. W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*, *29*, 410-425. http://dx.doi.org/10.1016/j.cedpsych.2003.11.001

Hart, J., & Driver, J. (1978). Teacher evaluation as a function of student and instructor personality. *Teaching of Psychology*, *5*, 377-378. http://dx.doi.org/10.1207/s15328023top0504_8

Marsh, H. W., & Roche, L A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias and utility. *American Psychologist*, *52*, 1187-1197. http://dx.doi.org/10.1037/0003-066X.52.11.1187

Maurer, T. W. (2006). Cognitive resonance or revenge? Student grades and course evaluations. *Teaching of Psychology*, *33*, 176-179. http://dx.doi.org/10.1207/s15328023top3303_4

McPherson, M. A., Todd, J. R., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. *Eastern Economic Journal*, *35*, 37-51. http://dx.doi.org/10.1057/palgrave.eej.9050042

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.

Murray, H. G., Rushton, J. P., & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, *82*(2), 250-261. http://dx.doi.org/10.1037/0022-0663.82.2.250

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

Olivares, O. J. (2001). Student interest, grading leniency, and teacher ratings: A conceptual analysis. *Contemporary Educational Psychology*, *26*, 382-399. http://dx.doi.org/10.1006/ceps.2000.1070

Orinau, D. J., & Bush, R. P. (1987). The propensity of college students to modify course evaluations and its impact on course performance information. *Journal of Marketing Education*, *9*, 42-52. http://dx.doi.org/10.1177/027347538700900108

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. http://dx.doi.org/10.1080/02602930903308258

Petchers, M. K., & Chow, J. C. (1988). Sources of variations in students' evaluations of instruction in a graduate social work program. *Journal of Social Work Education*, *24*, 35-42.

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. http://dx.doi.org/10.1080/00223980109603696

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. http://dx.doi.org/10.1080/01411920701492043

Santhanam, E., & Hicks, O. (2001). Disciplinary, gender and course year influences on student perceptions of teaching: Explorations and implications. *Teaching in Higher Education*, *7*, 17-31. http://dx.doi.org/10.1080/13562510120100364

Sauber, M. H., & Ludlow, R. R. (1988). Student evaluation stability in marketing: The importance of early class

meetings. *The Journal of Midwest Marketing*, *3*(1), 41-49.

Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation in higher education: Love me, love my lectures? *Assessment & Evaluation in Higher Education*, *25*, 397-405. http://dx.doi.org/10.1080/713611436

Spooren, P., Brockx, B., & Mortelman, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598-642. http://dx.doi.org/10.3102/0034654313496870

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. http://dx.doi.org/10.5539/ijel.v3n3p53

Ting, K. (2000). A multilevel perspective on student ratings of instruction: Lessons from the Chinese experience. *Research in Higher Education*, *41*, 637-661. http://dx.doi.org/10.1023/A:1007075516271

Tomasco, A. (1980). Student perceptions of instructional and personality characteristics of faculty: A canonical analysis. *Teaching of Psychology*, *7*(2), 79-83. http://dx.doi.org/10.1207/s15328023top0702_3