Article

Relationship Between L2 Proficiency and Psychological Traits With Self-Assessment Bias Among L2 Speakers

Noriko Iwamoto
Toyo University

Self-assessment is sometimes used to assess second language (L2) skills, but it has a degree of error that is possibly caused by L2 learners themselves, resulting from their L2 proficiency level and psychological traits. This study, involving 196 Japanese university students, calculated the self-assessment bias of L2 speakers using many-facet Rasch measurement. Correlation analysis explored the relationship between self-assessment bias and L2 speaking proficiency with psychological traits including self-esteem, English speaking anxiety, and English speaking motivation. The results showed that self-assessment bias was related to the belief in effortism, in which a person correlates ability with intensity of effort, and the Dunning-Kruger effect, in which high-proficiency learners tend to underestimate and low-proficiency learners tend to overestimate their abilities. Specifically, overestimation related to low proficiency prevents L2 learners from accurately assessing their performance, whereas underestimation related to L2 learners' belief that they lacked the effort to improve their speaking skills.

自己評価は時折第二言語技能の評価に使用されるが、ある程度の誤差が生じる。おそらくこの誤差には、第二言語 (L2) 習熟度や心理的特徴など、学習者自身によってもたらされる誤差が含まれる。本研究では196名の日本人大学生を対象に、多相ラッシュ分析を使ってL2スピーキング自己評価のバイアスを測定した。そして相関分析を用いて、自己評価バイアス、スピーキング習熟度、心理的特徴（自尊心、L2スピーキングの不安とモチベーション）の関係を調査した。その結果自己評価バイアスは、努力主義の信念とダニング・クルーガー効果と関連があることが明らかとなった。特に英語習熟度の低い学習者は自身のスピーキングパフォーマンスを過大評価

https://doi.org/10.37546/JALTJJ45.2-1

JALT Journal, Vol. 45, No. 2, November 2023

する傾向があり、一方でスピーキング技能を伸ばす努力が不足していると考える学習者は過小評価する傾向にあることがわかった。

Keywords: effort; English speaking anxiety; English speaking motivation

Self-assessment is sometimes used to assess second language (L2) skills. Self-assessment is considered beneficial for teachers because it provides them with "some idea of how the students view their own language abilities and development" (Brown, 2005, p. 58). It is also beneficial for learners because, by reflecting on their own skills and improvements, they increase their autonomy and motivation for learning (Brown & Hudson, 1998; Oscarson, 1989).

The question then arises whether learners can accurately evaluate their own L2 skills. High correlations greater than .70 between self-assessment and L2 proficiency have been reported in some studies (AlFallay, 2004; Babaii et al., 2016; Bachman & Palmer, 1989; Le Blanc & Painchaud, 1985). Other studies have found low correlation coefficients below .30 or no correlations (Brantmeier, 2006; Jafarpur, 1991; Trofimovich et al., 2016). In a meta-analysis involving ten studies, Ross (1998) reported a correlation between self-assessment and language performance of .633. Li and Zhang (2020), covering 67 studies, reported .466 and maintained that criterion type, training, and instruments are factors that exert a moderating effect.

Apart from the factors mentioned by Li and Zhang (2020), the inaccuracy of self-assessment can be attributed to L2 learners themselves, including their L2 proficiency and psychological traits. For example, L2 proficiency may influence self-assessment because of the Dunning-Kruger effect (Kruger & Dunning, 1999), in which high-proficiency learners tend to underestimate and low-proficiency learners tend to overestimate their abilities. L2 learners' psychological traits, such as self-esteem, L2 anxiety, and L2 motivation, have also been found to be related to their self-assessment (AlFallay, 2004; MacIntyre & Doucette, 2010; MacIntyre et al., 1997; Masgoret & Gardner, 2003). However, studies investigating the relation between self-assessment and psychological traits have often utilized L2 learners' self-assessment raw scores and tend to ignore self-assessment bias, or the degree of deviation from actual ability.

The present study used the many-facet Rasch measurement (MFRM) to calculate self-assessment bias-size measures of L2 speaking performance. The research examined L2 speaking skill because among the four skills, speaking seems to be most strongly connected to self-assessment, for

Underhill (1987) stated that when people talk with others, consciously or unconsciously, they are constantly assessing themselves in terms of how successfully they are communicating. Regarding psychological traits, the study chose self-esteem, L2 anxiety, and L2 motivation because self-esteem is considered to be related to the act of self-assessment (Heine et al., 2001), whereas anxiety and motivation have been extensively investigated in L2 acquisition literature. Therefore, this work dealt with the self-assessment bias of L2 speaking to investigate how the degree of deviation in the self-assessment of L2 speaking skills would relate to L2 learners' psychological traits, including self-esteem, L2 speaking anxiety, and L2 speaking motivation, as well as L2 speaking proficiency.

Literature Review

Most studies on the self-assessment of L2 skills have focused on the validity of self-assessment and often found overestimation of self-assessments compared with teacher ratings (Barrot, 2015; Hung et al., 2016; Jafarpur, 1991; Jassen-van Dieten, 1989; Suzuki, 2015; Trofimovich et al., 2016), probably owing to the "above-average effect," in which people tend to overestimate their own abilities and rate themselves above average (Dunning et al., 1989). Other researchers reported on the underestimation of self-assessments, especially among Asian students probably due to a modesty bias of their cultures (Aryadoust, 2015; Chen, 2008; Matsuno, 2009; Rian et al., 2014; Suzuki, 2009).

Markus and Kitayama (1991) explained that, in European and American cultural contexts, high self-esteem is a prerequisite for participating in independent and mutually approving relationships, and people in these relationships tend to view themselves positively. By contrast, in East Asian contexts, people tend to have lower self-esteem and tend to be self-critical, which is indispensable for mutually sympathetic relationships (Heine et al., 2001). Therefore, the modesty bias caused by lower self-esteem appears to be related to self-assessment. Indeed, because of their lower self-esteem, Asian participants were often found to evaluate their traits, abilities, or performance lower than Western counterparts (Farh et al., 1991; Heine et al., 2001). Therefore, self-esteem appears to be related to the act of self-assessment. However, few studies have investigated the correlation between self-esteem and self-assessment of L2 skills. Only AlFallay (2004) investigated the correlations between self-esteem and self-assessment of L2 presentation skills, reporting that participants with high self-esteem give higher self-assessment scores than the teacher-assessment, whereas participants

with low self-esteem are the most accurate in their self-assessment.

The overestimation and underestimation of L2 self-assessments can also be explained in terms of proficiency. Higher- and lower-proficiency learners tend to underestimate and overestimate their abilities, respectively (Barrot, 2015; Saito et al., 2020; Suzuki, 2015). Kruger and Dunning (1999) called this tendency the "Dunning-Kruger effect" in self-assessment and stated that those in the bottom quartile tend to overestimate their abilities because their incompetence prevents them from accurately evaluating their abilities, whereas top-quartile participants underestimate their abilities because they tend to consider their proficiency as similar to that of their peers.

To some extent, some researchers have investigated the influence of psychological traits on self-assessment of L2 skills. For L2 acquisition, L2 anxiety is one of the most important psychological factors influencing learners' L2 learning and performance. MacIntyre and Gardner (1994) defined it as "the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening, and learning" (p. 284). Horwitz et al. (1986) developed the Foreign Language Classroom Anxiety Scale (FLCAS) as a measure of anxiety specific to foreign language learning. L2 speaking is considered the most anxiety-provoking skill (Horwitz et al., 1986; Phillips, 1992), and L2 anxiety has shown a negative relation with L2 learners' self-assessment of L2 speaking (Clément et al., 1994; Gardner & MacIntyre, 1993; MacIntyre et al., 1997).

L2 motivation is another factor that strongly influences L2 learning. Gardner and MacIntyre (1993) described the motivated individual as "one who wants to achieve a particular goal, devotes considerable effort to achieve this goal, and experiences satisfaction in the activities associated with achieving this goal" (p. 3). Based on the socio-educational model, Gardner (1985) developed a multicomponential motivation questionnaire called the Attitude/Motivation Test Battery (AMTB) and found that motivation is positively correlated with L2 skills. Masgoret and Gardener (2003) conducted a meta-analysis of the studies by Gardner and his associates that used the AMTB. By investigating 75 independent samples in their meta-analysis, Masgoret and Gardener found that motivation is more strongly correlated with the self-rating of L2 skills compared with the other two achievement measures of course grades and objective measures.

Studies investigating self-assessment of L2 skills in relation to psychological traits have often calculated correlations between self-assessment raw scores and psychological traits. However, they have not focused on bias, or the degree to which self-assessment deviates from criterion measures.

Meanwhile, Saito et al. (2020) used overconfidence scores, calculated by subtracting the mean external listeners' score from their self-assessment score; they found no significant correlation between overconfidence scores and the promotional orientation variable, which represents more enjoyment and less anxiety. Their findings differ from those of previous studies that used self-assessment raw scores, in which psychological traits are often correlated with self-assessment. An implication is that self-assessment and self-assessment bias-size measures may not be the same.

Therefore, this study calculated the self-assessment bias of L2 speaking skills using MFRM and investigated how the degree of self-assessment inaccuracy is related to psychological traits, including self-esteem, English speaking anxiety, and English speaking motivation, as well as L2 speaking proficiency. The study posed the following research questions:

RQ1. To what degree do Japanese students' self-assessments of their L2 oral performance differ from teacher-assessments?

RQ2. To what degree do self-esteem, English speaking anxiety, English speaking motivation, and L2 speaking proficiency relate to self-assessment bias-size measures of L2 oral performance?

Method

Participants

The participants in this study included L2 learners as well as raters. As for the L2 learners, 196 students (53 females and 143 males) participated. They majored in science and engineering or information science at two private universities in Japan. Most participants had the English skills of level A2 or B1 of the Common European Framework of Reference for Languages. The participants had studied English for at least six years in English courses in Japanese schools, where reading and grammar were more often the focus of instruction than communication skills. Given the EFL context, most of the participants had had few opportunities to speak English in their daily lives.

Next, four raters participated in the study. Two raters were L1 English speakers, and two raters were L1 Japanese speakers. All raters were English teachers at Japanese universities and had experience testing and marking Japanese students' speaking abilities. A summary of the raters' profiles is presented in Table 1. Rater 2, an L1 Japanese speaker, earned his M.A. and Ph.D. degrees from an American university and had experience teaching at an American university for 20 years, thus he helped to back-translate the questionnaire.

Table 1
Raters' Profiles

Rater	Gender	Age	Nationality	Educational background	Teaching position	Teaching experience
1	Female	30s	American	M.A. in English Literature	Adjunct professor	3 years
2	Male	60s	Japanese	Ph.D. in Philosophy	Professor	35 years
3	Male	50s	Canadian	M.A. in TESOL	Adjunct professor	20 years
4	Female	40s	Japanese	M.A. in TESOL	Associate professor	13 years

Instruments

This study used a questionnaire to measure self-esteem, English speaking anxiety, and English-speaking motivation. The self-esteem scale was based on the self-esteem scale of Rosenberg (1965). The English-speaking anxiety scale was based on the FLCAS (Horwitz et al., 1986). The English-speaking motivation scale was based on the motivation items from the AMTB (Gardner, 1985); modifications were made based on the questionnaire used by Gardner et al. (1997) and Irie (2005). The English-speaking motivation scale consisted of three components: attitude toward speaking English, desire to learn to speak English, and motivational intensity. These English questionnaire items were translated to Japanese by the author, and the Japanese translation was backtranslated to English by a bilingual professor, who also served as Rater 2. The back-translated questionnaire was compared with the original, and some modifications were made. The items were rated using a six-point Likert scale, where 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, and 6 = strongly agree.

The participants' speaking abilities were assessed using a two-part oral proficiency test consisting of an interview and picture task. Each participant took part in a two-part five-minute oral interview. During the first part, the participant was asked to respond to five questions, such as "What is your hobby?" and "What did you do last weekend?" In the latter part, the participant was asked to tell a story in English while looking at a four-panel cartoon adapted from the pre-first level interview questions of the Eiken Test in

Practical English Proficiency (Obunsha, 2010). This test was used because it is the most widely known English proficiency test in Japan, supported by Japan's Ministry of Education, Culture, Sports, Science and Technology; many Japanese students are familiar with it.

The research assessed the participants' performance using the Kanda English Proficiency Test (KEPT) scale used by Bonk and Ockey (2003). This scale was selected because it was developed to assess the speaking skills of Japanese university students. KEPT has been successfully used as a diagnostic and placement test for Japanese university students (Ockey, 2009, 2011; Van Moere, 2006). The original KEPT scale consists of five categories designed to assess oral discussion skills in a group of four students. The present study excluded one category, communicative strategies, which assesses how students respond to and interact with other students. This study only conducted individual oral interviews in which no interactions with other students were involved. Thus, the remaining four categories of grammar, vocabulary, fluency, and pronunciation were used. The English descriptions were translated to Japanese using a back-translation method with the help of Rater 2. Although the original KEPT has six levels, the first level, "Does not discuss," was deleted because all participants spoke English during the individual interviews. Therefore, the remaining five levels were used: 1 = very weak, 2 = weak, 3 = fair, 4 = good, and 5 = very good. Each level is accompanied by a description of the performance for that level. Because raters can use half points from levels 1 to 4, KEPT is a nine-point scale. The same scale was used for both the students' self-assessment and the teachers' assessment.

Data Collection

Data were gathered from May to August 2011. The study obtained written consent from those who agreed to participate in the study. They completed the questionnaire and submitted it during the individual oral interview with the author, a Japanese teacher of English. The interviews were audio-recorded with an IC recorder. After completing the oral interview, each participant was given a self-assessment sheet and was instructed to evaluate the four categories of grammar, vocabulary, fluency, and pronunciation by finding the description of the KEPT scoring levels that best matched their performance. The participants did not receive self-assessment training, which often improves the accuracy of self-assessment (Babaii et al., 2016; Chen, 2008), given that the objective of this study was to determine how the degree of self-assessment inaccuracy is related to psychological traits and proficiency.

Before conducting the teacher-assessment, each rater received an explanation of the oral interview and rating scale by the author. Using the audio-recorded data, four raters independently assessed each participant's oral proficiency using the same scoring rubric. MFRM does not require every rater to assess the complete data set, only that there is sufficient overlap (Linacre & Wright, 2002). Thus, to save time and labor, Rater 1 assessed students 1 to 146, Rater 2 assessed students 26 to 196, Rater 3 assessed students 97 to 196, and Rater 4 assessed students 1 to 25 and 97 to 196.

Data Analysis

The study analyzed the collected questionnaire data using Winsteps 3.80.1 (Linacre, 2013). The Rasch rating scale model provides several advantages in analyzing Likert-scale data over using raw scores (Apple, 2013). First, it changes ordinal raw scores to interval measures called logits. When the average is set at 0 logits; positive logits represent higher than average scores and negative logits, lower (Bond et al., 2021). The Rasch model also indicates the relative difficulty of each item and places both persons and items on the same single logit scale (Bond et al., 2021). Second, the Rasch model provides fit statistics that allow the identification of poorly performing items and raters. Reasonable infit and outfit MNSQ values fall within .5 and 1.5 (Linacre, 2007), which was used for this study. Third, researchers can check the dimensionality of the items hypothesized to measure the same trait using Rasch principal components analysis (PCA) of item residual analysis (Bond et al., 2021) which is generally used to determine unidimensionality. The criteria for determining unidimensionality are that over 50% is necessary for the variance explained by measures, and that the first contrast should account for either less than 10% of the variance and/or the eigenvalue should be less than 3.0 (Linacre, 2007). The present study used these criteria.

Participants' speaking data were examined using MFRM, which is an extension of the Rasch model. In addition to person ability and item difficulty, MFRM can assess other variables such as tasks and raters (Linacre, 2014). MFRM was used for L2 speaking assessment in this study because of its advantages over conventional approaches. First, MFRM can provide estimates of ability adjusted for rater bias; in contrast, speaking scores in a conventional approach using raw scores are likely to be degraded owing to differences in rater severity/leniency (Bond et al., 2021). Second, the joint calibration of facets allows rater severity to be placed on the same scale as ratee performance and task difficulty. This enables researchers to "draw

useful, diagnostically informative comparisons among the various facets" (Myford & Wolfe, 2003, p. 404).

The study calculated students' ability measures using four teachers' ratings, processed via Facets 3.80.0 (Linacre, 2017). Next, the self-assessment ratings were separated from the teacher-ratings by weighing teacher-assessment scores at .001, and the self-assessment measures were calculated. Then, bias-size measures were calculated by subtracting teacher-assessment measures from self-assessment measures. A positive bias-size measures indicated a more lenient self-rater relative to the teacher raters. A negative bias-size measures indicated a more severe self-rater compared with the teacher raters.

Preliminary Analysis

The study applied a Rasch measurement model for the questionnaire items using Winsteps. Three self-esteem items did not fit the Rasch model and were deleted. To check for unidimensionality, the study conducted Rasch PCA for each construct, which were found to meet the criteria. Therefore, five factors were used in the main analysis: self-esteem, English speaking anxiety, attitude toward speaking English, desire to learn to speak English, and motivational intensity (see Appendix A for the descriptions of items with the Rasch analysis results).

Results

Research Question 1

The study used MFRM; the modeled facets were the raters, students, and assessment categories. Scale 9 was not used by any rater, probably because the participants were all science majors whose English-speaking abilities were not as high as English language majors for whom Bond and Ockey (2003) originally created the KEPT rubric. The eight-point scale met Linacre's (2002) criteria for effective category functioning: at least 10 responses were made for each category, the outfit MNSQ was below 2.00, and the step difficulty of each category advanced by at least .25 logits (Wolfe & Smith, 2007). Therefore, an eight-point scale was used in this study.

Table 2 provides a rater measurement report for teacher- and self-raters, and Table 3 provides that for teacher-raters only. The current study adopted Linacre's (2007) infit and outfit MNSQ criterion of .50 to 1.50, which indicates that the items do not greatly diverge from Rasch model expectations. As shown in Tables 2 and 3, the infit and outfit MNSQ values in the present

study met this criterion. The reliability estimate of .99 means that the raters were separated into different levels of severity. The significant chi-square value showed that all raters were not equally severe.

Table 2

Calibration Report for Teacher-Raters and Self-Raters

Rater	Logit Measure	SE	Infit MNSQ	Outfit MNSQ
1	.53	.04	.80	.81
2	-.22	.04	.90	.92
3	.59	.05	.78	.78
4	.39	.04	.86	.86
Self	1.24	.03	1.42	1.46

Note. Fixed (all same) chi-square = 823.3; df = 4, p < .001; separation = 11.29; reliability = .99.

Table 3

Calibration Report for Teacher-Raters

Rater	Logit Measure	SE	Infit MNSQ	Outfit MNSQ
1	.63	.05	.89	.91
2	-.39	.04	1.08	1.09
3	.69	.06	.97	.97
4	.42	.05	.98	.98

Note. Fixed (all same) chi-square = 358.1; df = 3, p < .001; separation = 8.70; reliability = .99.

Bias-size measures were calculated by subtracting the teacher-assessment measures from the self-assessment measures. Pearson correlations between teacher-assessments (TA) and self-assessments (SA) were calculated. The results showed that TA and SA were moderately correlated (r = .44, p < .001).

Research Question 2

To answer Research Question 2, two groups were created based on the bias-size measures from both ends. AlFallay (2004) used subjects with the

highest and lowest 25% scores to represent the opposing groups. This grouping method was used in the present study. An overestimation (OE) group and an underestimation (UE) group were created based on the highest and lowest 25% bias-size measures. The OE group included the top quarter students, or 51 students with bias-size measures above 1.4 logits, and the UE group included the bottom quarter students, or 50 students with bias-size measures below -.85 logits. The descriptive statistics of each variable for each group, as well as for all students, are shown in Table 4. A MANOVA was conducted to determine the effect of bias-size measures (overestimation or underestimation) on the TA, SA, and five variables. Wilks's Λ was significant, $F(7, 93) = 15.97$, $p < .01$, $\eta^2 = .55$. As a follow-up test, a series of ANOVAs were conducted for each dependent variable. Using the Bonferroni method, each ANOVA was tested at the .007 level (.05/7). The results showed that the ANOVAs for TA and SA were significant: TA, $F(1, 99) = 18.15$, $p < .007$, $\eta^2 = .16$; SA, $F(1, 99) = 50.06$, $p < .007$, $\eta^2 = .34$. Thus, UE students had significantly higher average teacher-assessment measures than OE students, whereas their self-assessment measures were significantly lower than those of OE students. However, the results showed no significant differences in psychological traits between the two groups, indicating that those who overestimated and underestimated their abilities tended to have similar levels of self-esteem, anxiety, and motivation.

Table 4

Descriptive Statistics for All, Overestimation, and Underestimation Groups

	All		Overestimation		Underestimation	
	M	*SD*	*M*	*SD*	*M*	*SD*
TA	-.03	1.54	-.68	.22	.63	.22
SA	-.00	2.79	1.66	.41	-2.50	.42
SE	.04	1.64	.11	.25	.03	.25
ANX	.54	1.49	.34	.23	.85	.23
ASE	.42	1.59	.47	.24	.40	.24
DLSE	.67	1.37	.63	.19	.88	.19
MI	-.95	1.87	-.95	.28	-1.19	.29

Note. TA = Teacher-Assessment; SA = Self-Assessment; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

Next, the correlations between the eight variables were calculated for all students and for each group, and the results are shown in Tables 5, 6, and 7. Several differences were observed between the two groups. First, the bias-size measures of the OE group were negatively correlated with TA (r = -.54, p < .01), whereas those of the UE group were positively correlated with TA (r = .64, p < .01). The negative relation for the OE group indicated that those with lower speaking abilities were likely to overestimate their own performance. Meanwhile, the positive relation for the UE group indicated that those with lower speaking ability tended to assess themselves lower. Second, unlike the results of OE students, which showed no correlations between bias-size measures and SA, UE students had a high correlation between them (r = .95, p < .01). An implication is that the greater their negative bias, the lower their self-assessment measures, whereas the smaller their negative bias, the higher their self-assessment. However, such a regular and systematic relation was not observed among the OE students. Finally, no correlations were found between the bias-size measures and psychological traits of OE students, meanwhile the bias-size measures of UE students were significantly correlated with the two variables of desire to learn to speak English and motivational intensity.

Table 5

Correlations Among Teacher-Assessment, Self-Assessment, Bias-Size Measures, and Five Psychological Traits (N = 196)

Variable	1	2	3	4	5	6	7	8
1. TA	——							
2. SA	.44**	——						
3. Bias	-.12	.84**	——					
4. SE	.08	.05	.00	——				
5. ANX	-.09	-.18**	-.15	-.18*	——			
6. ASE	.24**	.21**	.09	.29**	-.23**	——		
7. DLSE	.33**	.20**	.02	.15*	-.05	.55**	——	
8. MI	.34**	.32**	.15*	.04	-.18*	.36**	.44**	——

Note. TA = Teacher-Assessment; SA = Self-Assessment; Bias = Bias-Size Measures; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

*p < .05. **p < .01.

Table 6

Correlations Among Teacher-Assessment, Self-Assessment, Bias-Size Measures, and Five Psychological Traits with Overestimation Group (N = 51)

Variable	1	2	3	4	5	6	7	8
1. TA	——							
2. SA	.85**	——						
3. Bias	-.54**	-.00	——					
4. SE	.12	.22	.13	——				
5. ANX	-.12	-.16	-.02	-.12	——			
6. ASE	.33**	.36*	-.05	.36**	-.06	——		
7. DLSE	.43**	.46**	-.07	.21	-.01	.65**	——	
8. MI	.28*	.32**	-.01	-.02	-.15	.24	.47**	——

Note. TA = Teacher-Assessment; SA = Self-Assessment; Bias = Bias-Size Measures; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

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

Table 7

Correlations Among Teacher-Assessment, Self-Assessment, Bias-Size Measures, and Five Psychological Traits with Underestimation Group (N = 50)

Variable	1	2	3	4	5	6	7	8
1. TA	——							
2. SA	.84**	——						
3. Bias	.64**	.95**	——					
4. SE	.01	-.08	-.12	——				
5. ANX	-.14	-.13	-.10	-.31*	——			
6. ASE	.24	.24	.20	.20	-.31*	——		
7. DLSE	.27	.33*	.32*	.13	-.07	.55**	——	
8. MI	.43**	.42**	.35*	-.08	-.12	.48**	.44**	——

Note. TA = Teacher-Assessment; SA = Self-Assessment; Bias = Bias-Size Measures; SE = Self-Esteem; ANX = English Speaking Anxiety; ASE = Attitude toward Speaking English; DLSE = Desire to Learn to Speak English; MI = Motivational Intensity.

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

Discussion

Research Question 1

Research Question 1 compared students' self-assessment of L2 oral performance with teacher-assessment. The present study identified a correlation of .44. The moderate correlation suggested that self-assessments of L2 oral performance by EFL learners may not be very reliable. The results of rater severity showed that self-assessment (1.24 logits) was more severe than teacher-assessments (-.22 to .59 logits). An examination of the bias-size measures showed that the number of students who overestimated their ability was larger than those who underestimated it. For example, 71 students had bias-size measures above 1.0, whereas 41 students had measures below -1.0 logits. Although more students overestimated their proficiency, some students made excessive underestimation: The least ability measures were -6.23 logits by teacher-assessment and -9.73 logits by student-assessment. Twelve students assessed themselves lower than -6.23 logits. In contrast, the highest ability measures were not greatly different between the teacher-assessment (3.92 logits) and self-assessment (4.28 logits). Thus, excessively severe self-assessment measures made by some students might have caused a greater severity of self-assessment logits compared with the teacher-assessment logits.

Research Question 2

Research Question 2 examined the relation of self-assessment bias with L2 proficiency and psychological traits. As teacher-assessment fit the Rasch model (Table 3), teacher-assessment in this study is considered to be a reliable measure of learners' L2 speaking proficiency. This is because MFRM can produce person ability measures that are adjusted for rater bias as long as raters are internally consistent (Bond et al., 2021). First, the results for all students were checked. The correlation results, shown in Table 5, revealed that self-assessment was correlated with L2 proficiency and all psychological variables except self-esteem, whereas bias-size measures were correlated with anxiety and motivational intensity only. Self-assessment and bias-size measures showed different results, which could indicate that these two are not the same.

Regarding psychological traits, English speaking anxiety was negatively correlated with bias-size measures. This showed that those with greater anxiety tended to underestimate their own performance. This finding is in accordance with past studies that found a negative relationship between

anxiety and self-assessment of L2 speaking (Clément et al., 1994; Gardner & MacIntyre, 1993; MacIntyre et al., 1997).

In addition to anxiety, bias-size measures were correlated with motivational intensity. The positive correlation indicated that those who said they extended more effort tended to overestimate their own speaking performance. This seems to imply Japanese people's belief in effortism, a tendency to value effort (Okawa, 2016). Kariya (1995) explained that unlike Western societies, Japanese society tends to believe that students' academic achievements are mainly derived from their efforts rather than their innate abilities. For example, Sudo's (2015) empirical study of 3,436 Japanese elementary school students revealed that although only one-third of the students' academic achievements were significantly related to their studying hours, 76.5% believed that anyone could be good at school subjects if they put in enough effort. From this result, Sudo (2015) highlighted Japanese people's tendency to blame students' lack of effort for their poor grades while paying little attention to their innate abilities. Likewise, the participants in the present study probably connected their L2 speaking ability to how much effort they had made.

Next, the participants from the top and bottom quartiles, 51 OE students and 50 UE students, were compared to investigate the students whose self-assessment measures greatly deviated from the teacher-assessment. The UE group had significantly higher average L2 speaking ability than the OE group, but the former evaluated their performance significantly lower than the latter. SA and TA were correlated highly in both groups (r = .85 for OE and .84 for UE, p < .01), compared with all students (r = .44, p < .01), which was an unexpected result because the OE and UE students had extreme bias. A possible explanation is that the OE and UE students each had the same tendency of either overestimation or underestimation for their self-assessments, which might have caused greater correlational values. This may imply that high correlational values do not necessarily indicate the accuracy of self-assessment. Bond et al. (2021) also pointed out the problem with intercorrelations between judge ratings because "they can demonstrate only consistency among the rank orders of candidates. They do not tell us anything about the severity or leniency differences between judges" (p. 147). This also supports the importance of calculating bias-size measures; the inaccuracy of self-assessment may not be detected from the results of correlation analysis.

Several different correlation results were found between the OE and UE groups. First, the correlation coefficient for teacher-assessment and bias-

size measures was -.54 for OE students. This negative relation indicated that among OE students, those with lower L2 speaking proficiency tended to overestimate their own performance. This demonstrated the Dunning-Kruger effect, and because the OE students had lower proficiency than average, as Kruger and Dunning (1999) stated, their incompetency might have prevented them from accurately evaluating their performance. On the contrary, the teacher-assessment and bias-size measures for UE students had a positive relation of .64, indicating that among UE students, those with lower proficiency tended to underestimate their performance. Thus, the Dunning-Kruger effect, where lower-proficiency participants tended to overestimate their ability, was not observed among UE students. A possible explanation may be that the participants conducted self-assessments immediately after their oral performance had finished. Other studies conducted prior practice and/or peer-assessment, after which participants could think about their performance more objectively. Meanwhile, the self-assessments in the present study might have been more influenced by participants' immediate subjective feelings. If the participants were not satisfied with their performance during the interview, their negative feelings on their performance might have caused them to evaluate themselves significantly lower. Indeed, after the interview, some participants who could not speak English well hung their heads or lamented, "Oh, my English ability is so poor!" The disheartening feeling that they could not speak English as expected might have caused them to have a greater negative bias toward their own L2 performance. Thus, among UE students, those with lower proficiency tended to assess themselves lower than necessary.

Second, the correlation between bias-size measures and self-assessment differed between the two groups. The UE group had a high correlation of .95, whereas the OE group showed no significant correlation. UE students with a positive bias toward their L2 oral performance tended to give a higher self-assessment, whereas UE students with a negative bias tended to give a lower self-assessment. Meanwhile, OE students did not show such systematic relations. The inconsistency in the bias size for OE students could signify their incompetence in evaluating L2 oral performance properly owing to their lower proficiency (Kruger & Dunning, 1999).

Finally, the relation between bias-size measures and psychological traits differed between the two groups. Although the bias size of the OE group had no significant relation with any psychological traits, that of the UE group showed a weak relation with desire to learn to speak English (DLSE) and motivational intensity (MI), implying that when they had little desire

or made little effort to improve their spoken English, they were likely to underestimate their performance. Therefore, the belief in effortism seems to be related especially to underestimation. Although the degrees of DLSE and MI were not significantly different between the UE and OE students, only the UE students' DLSE and MI were related to self-assessment bias size. Thus, UE students may be more likely to feel that their lack of a great desire or sufficient effort to improve in speaking English could indicate their low level of speaking ability, leading them to underestimate their ability. In other words, they seem to believe that if they have a stronger desire and make more effort, then their English-speaking ability will improve. This idea is in accordance with a previous finding that Japanese self-perception tends to be critical because Japanese people strongly believe in improvement and achievement (Heine et al., 1999). Hung et al. (2016) stated that highly proficient learners underestimate their L2 abilities because they set high standards for themselves. Thus, for UE students, although their speaking ability is higher than average, they are critical of and underestimate their own L2 speaking skills, which may represent their dissatisfaction with the intensity of their desire and efforts and their belief in the possibility of improvement. Although lower self-esteem is considered to be related to self-criticism (Markus & Kitayama, 1991), in the present study, no significant correlations were found between self-esteem and bias-size measures in either OE or UE group, as both had the same level of self-esteem. Thus, contrary to the hypothesis, lower self-esteem showed no relation to underestimation, but belief in effortism seemed more related to underestimation.

This study's results have several implications. First, this study calculated bias-size measures, instead of raw scores for self-assessments, to elucidate the degree to which self-assessment could deviate from actual ability. Indeed, the use of bias-size measures obtained results different from those obtained using self-assessment measures. For example, all students' self-assessment measures were correlated with L2 proficiency, whereas the bias-size measures were not. Their self-assessment measures were correlated with most psychological traits, but the bias-size measures were correlated only with anxiety and motivational intensity. Research on the relation between self-assessment and psychological traits has mainly utilized self-assessment raw scores to explore for correlations. These previous studies revealed that L2 learners with certain psychological traits give higher or lower self-assessment scores for their abilities. By calculating self-assessment bias-size measures, the present study could explore how the degree of overestimation and underestimation is related to one's psychological traits. Therefore, in-

vestigation of self-assessment bias-size measures will provide new insights into complex self-assessment behaviors among L2 learners with different degrees of anxiety and motivation, along with other psychological traits.

Second, L2 proficiency was negatively correlated with the bias-size measures of OE students, indicating that lower-proficiency students tended to have a greater overestimation of their speaking abilities. As Kruger and Dunning (1999) explained, students' lack of competence probably prevents them from assessing their own ability accurately. Overestimation of one's own L2 ability based on one's incompetence is a problem requiring attention. If learners cannot realize their weaknesses, then future improvements could be stalled. As some studies have reported that the accuracy of self-assessment improves after training (Babaii et al., 2016; Chen, 2008), lower-proficiency students especially require training that may help them correctly view their own L2 ability.

The third implication is that L2 learners' self-assessment bias may be related to their (lack of) effort in speaking English. In particular, higher-proficiency learners had a tendency to view their L2 speaking performance negatively when they considered that they had not made sufficient efforts to improve their English speaking skill. Japanese people's belief in effortism encourages and motivates students to study harder, but at the same time it can be problematic because it diverts attention from actual abilities (Sudo, 2015). Although teachers should emphasize the significance of making an effort, they should also teach learners to view their actual abilities more accurately, without taking into account how much effort has been made.

Conclusion

This study analyzed the self-assessment bias-size measures for L2 oral performance of 196 students and considered their relations with self-esteem, English speaking anxiety, English speaking motivation, and L2 speaking proficiency. The results showed that self-assessment was moderately correlated with teacher-assessment. The self-assessment bias is believed to be attributed to weak influences of the belief in effortism and the Dunning-Kruger effect.

This study had two limitations. First, it involved only science majors from two Japanese universities; the inclusion of more participants with different majors and/or proficiencies is needed for generalization. For example, those majoring in English may have more confidence in their L2 abilities and evaluate their English skills differently. Second, because culture can influence the evaluative attitude of one's own abilities (Heine et al., 2001), participants

from different backgrounds should be included to elucidate the influences of culture, such as modesty bias and effortism, on self-assessment.

Despite these limitations, this research is among the few studies that have investigated the relation between self-assessment bias, L2 speaking proficiency, and psychological traits. Self-assessment is considered a highly complex metacognitive task (Butler & Lee, 2006) that cannot be explained by only a few factors. Thus, other variables might influence self-assessment. For example, qualitative data, such as interviews, might shed light on the factors that cause inaccuracy in self-assessment. Accurate assessment of one's own language skills is important for L2 acquisition; by recognizing their own strengths and weaknesses, learners can become more autonomous and monitor their own L2 skills. Given the EFL context of Japan, L2 learners have few opportunities to speak English outside their classrooms, unlike L2 learners in ESL contexts who can use their L2 in daily life and monitor their own speaking skills every day. If EFL learners learn to assess their own speaking proficiency, such as by becoming better informed of their biases, then they can gauge their own progress. Further research on the self-assessment of speaking skills is needed to enhance the L2 oral proficiency of EFL learners.

Noriko Iwamoto is a professor of English as a Foreign Language (EFL) in the Faculty of Science and Engineering at Toyo University. Her research interests include self-assessment of L2 skills and the influence of L2 affective factors on second language learning.

References

AlFallay, I. (2004). The role of selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System, 32*(3), 407–425. https://doi.org/10.1016/j.system.2004.04.006

Apple, T. M. (2013). Using Rasch analysis to create and evaluate a measurement instrument for foreign language classroom speaking anxiety. *JALT Journal, 35*(1), 5–28. https://doi.org/10.37546/JALTJJ35.1-1

Aryadoust, V. (2015). Self- and peer assessments of oral presentations by first-year university students. *Educational Assessment, 20*(3), 199–225. https://doi.org/10.1080/10627197.2015.1061989

Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing, 33*(3), 411–437. https://doi.org/10.1177/0265532215590847

Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing, 6*(1), 14–29. https://doi.org/10.1177/026553228900600104

Barrot, J. (2015). Self- and teacher assessment of speaking performances: An investigation of interrater consistency and agreement. *Journal of Language and Literature, 6*(4), 193–200.

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge. https://doi.org/10.4324/9780429030499

Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing, 20*(1), 89–110. https://doi.org/10.1191/0265532203lt245oa

Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. *System, 34*(1), 15–35. https://doi.org/10.1016/j.system.2005.08.004

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment.* McGraw-Hill.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4)*,* 653–675. https://doi.org/10.2307/3587999

Butler, Y. G., & Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal, 90*(4), 506–518. https://doi.org/10.1111/j.1540-4781.2006.00463.x

Chen, Y.-M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research, 12*(2), 235–262. https://doi.org/10.1177/1362168807086293

Clément, R., Dörnyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning, 44*(3)*,* 417–448. https://doi.org/10.1111/j.1467-1770.1994.tb01113.x

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*(6), 1082–1090. https://doi.org/10.1037/0022-3514.57.6.1082

Farh, J., Dobbins, G. H., & Cheng, B. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U.S. workers. *Personnel Psychology, 44*(1), 129–147. https://doi.org/10.1111/j.1744-6570.1991.tb00693.x

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitude and motivation.* Edward Arnold.

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contribution to second language learning. Part II: Affective variables. *Language Teaching, 26*, 1–11. https://doi.org/10.1017/S0261444800000045

Gardner, R. C., Tremblay, P. F., & Masgoret, A. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal, 81*(3), 344–362. https://doi.org/10.1111/j.1540-4781.1997.tb05495.x

Heine, S. J., Kitayama, S., & Lehman, D. R. (2001). Cultural differences in self-evaluation: Japanese readily accept negative self-relevant information. *Journal of Cross-Cultural Psychology, 32*(4)*,* 434–443. https://doi.org/10.1177/0022022101032004004

Heine, S. J., Markus, H. R., Lehman, D. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*(4), 766–794. https://doi.org/10.1037/0033295X.106.4.766

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.2307/327317

Hung, Y., Samuelson, B. L., & Chen, S. (2016). Relationships between peer- and self-assessment and teacher assessment of young EFL learners' oral presentations. In M. Nikolov (Ed.), *Assessing young learners of English: Global and local perspectives* (pp. 317–338). Springer. https://doi.org/10.1007/978-3-319-22422-0_13

Irie, K. (2005). *Stability and flexibility of language learning motivation: A multimethod study of Japanese junior high school students* (Publication No. 3176834) [Doctoral dissertation, Temple University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/dissertations-theses/stability-flexibility-language-learning/docview/305393271/se-2

Jafarpur, A. (1991). Can native EFL learners estimate their own proficiency? *Evaluation and Research in Education, 5*(3), 145–157. https://doi.org/10.1080/09500799109533306

Janssen-van Dieten, A.-M. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing, 6*(1), 30–46. https://doi.org/10.1177/026553228900600105

Kariya, T. (1995). *Taishū kyōiku shakai no yukue* [Prospect of public education society]. Chukou Shinsho.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi:10.1037/0022-3514.77.6.1121

Le Blanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*(3), 673–687. https://doi:10.2307/3586670

Li, M., & Zhang, X. (2020). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing, 38*(2), 189–218. https://doi.org/10.1177/0265532220932481

Linacre, J. M. (2002). Optical rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85–106.

Linacre, J. M. (2007). *A user's guide to WINSTEPS: Rasch model computer program*. https://www. winsteps.com

Linacre, J. M. (2013). *WINSTEPS* (version 3.80.1) [Computer software]. Beaverton. https://www. winsteps.com

Linacre, J. M. (2014). *A user's guide to FACETS: Rasch model computer program.* https://www. winsteps.com

Linacre, J. M. (2017). *FACETS* (version 3.80.0) [Computer software]. Beaverton. https://www. winsteps.com

Linacre, J. M., & Wright, B. D. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement, 3*(4), 486–512.

MacIntyre, P. D., & Doucette, J. (2010). Willingness to communicate and action control. *System, 38*(2), 161–171. https://doi.org/10.1016/j.system.2009.12.013

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning, 47*(2), 265–287. https://doi.org/10.1111/0023-8333.81997008

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. https://doi.org/10.1037/0033295X.98.2.224

Masgoret, A., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning, 53*(1), 123–163. https://doi.org/10.1111/1467-9922.00227

Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing, 26*(1)*,* 75–100. https://doi.org/10.1177/0265532208097337

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of Applied Measurement, 4*(4), 386–422.

Obunsha. (2010). *Eiken jun ikkyū nijishiken mensetsu taisaku kanzen yosōmondai* [The Eiken test in practical English proficiency pre-1 level interview test questions]. Obunsha.

Ockey, G. J. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test scores. *Language Testing, 26*(2), 161–186. https://doi.org/10.1177/0265532208101005

Ockey, G. J. (2011). Self-consciousness and assertiveness as explanatory variables of L2 oral ability: A latent variable approach. *Language Learning, 61*(3), 968–989. https://doi.org/10.1111/j.1467-9922.2010.00625.x

Okawa, K. (2016). *Ganbaru koto/ganbaranai koto no shakai-gaku: Doryoku shugi no yukue* [Sociology of making effort/not making effort: Prospect of effortism]. Harbestosha.

Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*(3), 1–13. https://doi.org/10.1177/026553228900600103

Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal, 76*(1), 14–26. https://doi.org/10.2307/329894

Rian, J., Hinkelman, D., & Cotter, M. (2014). Self-, peer, and teacher assessments of student presentation videos. In P. Clements, A. Krause, & H. Brown (Eds.), *JALT2014 Conference Proceedings* (pp. 688–697). JALT.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press. https://doi.org/10.1515/9781400876136

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing*, *15*(1), 1–20. https://doi.org/10.1177/026553229801500101

Saito, K., Trofimovich, P., Abe, M., & In'nami, Y. (2020). Dunning-Kruger effect in second language speech learning: How does self-perception align with other perception over time? *Learning and Individual Differences*, *79*, 101849. https://doi.org/10.1016/j.lindif.2020.101849

Sudo, K. (2015). *Shōgakusei no doryoku shugi no keisei yōin to kiketsu* [Formation factors and consequences of elementary school students' effortism]. Benesse. https://berd.benesse.jp/up_images/research/Survey_on_learning_report_3.pdf

Suzuki, M. (2009). The compatibility of L2 learners' assessment of self- and peer revisions of writing with teachers' assessment. *TESOL Quarterly, 43*(1), 137–148. https://doi.org/10.1002/j.1545-7249.2009.tb00233.x

Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing, 32*(1), 63–81. https://doi.org/10.1177/0265532214541885

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition, 19*(1), 122–140. https://doi.org/10.1017/S1366728914000832

Underhill, N. (1987). *Testing spoken language*. Cambridge University Press.

Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing, 23*(4), 411–440. https://doi.org/10.1191/0265532206lt336oa

Wolfe, E. W., & Smith, E. V. Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II-Validation activities. *Journal of Applied Measurement, 8*(2), 204–234.

Appendix

Questionnaire Item Descriptions and Rasch Results

Items	Descriptions	logits
Self-Esteem		
person reliability = .83, item reliability = .98		
SE1	I believe that I have a number of good qualities.	-.98
SE2	I am able to do things as well as most other people.	-.95
SE3	I feel useful most of the time.	-.01
SE4	I feel that I am a person of worth.	-.11
SE5	I respect myself.	1.03
SE6	I am able to do things better than other people.	.34
SE7	I have more good points than weak points.	.69

English Speaking Anxiety person reliability = .77, item reliability = .95		
ANX1	I would feel nervous speaking English with native speakers of English.	-.46
ANX2	I feel nervous about speaking English in class activities.	.13
ANX3	I lack confidence in my English-speaking abilities.	-.51
ANX4	I worry that my English teacher thinks that my English-speaking level is low.	.72
ANX5	I worry that I will make mistakes when I speak English	-.01
ANX6	I feel nervous having a conversation in English.	.13
Attitude Toward Speaking English person reliability = .82, item reliability = .89		
ASE1	I enjoy speaking English.	-.25
ASE2	I enjoy speaking English more than reading English.	-.11
ASE3	I enjoy speaking English more than writing English.	-.44
ASE4	I look forward to my English-speaking classes.	.38
ASE5	I enjoy English speaking classes more than other classes.	.17
ASE6	I look forward to opportunities to speak English.	.25
Desire to Learn to Speak English person reliability = .79, item reliability = .97		
DLSE1	Speaking English is important for engineers.	-.81
DLSE2	I would take an English conversation course in school, even if it were not required.	.54
DLSE3	I wish I had more classes in which I could speak English.	.74
DLSE4	I really want to learn to speak English better.	-.44
DLSE5	I believe that Japanese students should be taught to speak English at school.	-.34
DLSE6	My desire to learn to speak English is increasing.	.30

Motivational Intensity person reliability = .84, item reliability = .88		
MI1	I think I try to speak English more than other students.	.08
MI2	I look for opportunities to speak English outside of class.	-.24
MI3	I spend a long time studying English.	.13
MI4	I study English more than most of my classmates.	.44
MI5	I often think about how I can improve my English-speaking skills.	-.48
MI6	I work hard to become an excellent speaker of English.	.09