

Is Self-Assessment Possible? : Japanese Students' Self-Assessment in their Course Work

Noriko IWAMOTO

1. Introduction

Although self-assessment has been used in many fields such as psychology, sociology, and business, it was only during the 1980s that self-assessment began to gather attention in L2 learning contexts. Nevertheless, many researchers emphasize the importance and usefulness of self-assessment in language learning. For example, Todd describes five reasons for using self-assessment. First, self-assessment is a prerequisite for a self-directed learner. Second, it can raise learners' awareness of language. Third, it increases motivation and goal orientation in learning. Fourth, some aspects of language learning, such as effort and learner beliefs, can only be assessed through self-assessment. Finally, it can reduce teacher's assessment burden. In addition to these benefits, Oscarson states that asking questions such as "How well have I done?" stimulates learners to consider assessment in the course more keenly, and this fosters important evaluative attitude in learners (*Language Proficiency*, 4). Furthermore, Kusnic and Finley claim that "self-evaluation helps students make meaning, derive relevance, and build coherence through their educational experience. This form of active engagement seems an important antidote to passivity in the classroom" (13).

However, some researchers question the appropriateness of self-

assessment because of the doubt that learners are capable of assessing their own abilities. Jordan (cited in Blanc and Painchaud) asserts that it is wrong to have students assess their own English abilities because they have no yardstick with which to compare themselves. McLead (cited in Blanc and Painchaud) also claims that good students tend to underestimate their abilities, while students who arrived at a plateau tend to overestimate theirs. In this paper, I would like to examine if Japanese students' self-assessment has any relationship with their true achievement in the course.

2. Literature Review

2.1. Correlation with True Ability

Self-assessment is considered to be an alternative assessment to traditional teacher-centered assessment, which has been mainly conducted at schools. Brown defines the self-assessment in terms of second or foreign language as "any items wherein students are asked to rate their own knowledge, skills, or performances. Thus, self-assessments provide the teacher with some idea of how the students view their own language abilities and development"(58).

When we consider introducing self-assessment into our classroom or utilizing it as a placement tool, we question whether self-assessment is actually reflecting learners' true abilities. Some studies were conducted to explore the validity of self-assessment. Le Blanc and Painchaud investigated whether students have the ability to evaluate their own proficiency in their second language and whether self-assessment can be used as a placement instrument. The participants were 200 freshmen at University of Ottawa, whose second language was either English or French. They took English or French standard proficiency tests (listening & reading), and also answered self-assessment questionnaires about their second language proficiency. The correlation between the two was 0.53. Next, the authors modified the self-assessment questionnaire to be closely tied to the students' situations as second language users. This improved the correlation to 0.80. Accordingly, this study suggests that if self-assessment is closely related with students' situations, it has a high

correlation with students' tested proficiency. In fact, University of Ottawa used self-assessment instead of proficiency tests for placing students and the authors reported that it worked very well.

Ross conducted a self-assessment validity study for a language training program at a Japanese electronics company. Two hundred fifty four subjects at the beginning and elementary level of English completed a self-assessment of English skills. Eight teachers also provided assessments of each learner. Learners then took the listening test which included the sections that the subjects had experienced directly in the coursebook, and the section using a different format from the coursebook. The regressions of the students' self-assessment with the modified format test section showed a correlation of 0.39, while students' self-assessment with the exact match test sections had a correlation of 0.50. Consequently, this study revealed that the accuracy of self-assessment was larger when the criterion was achievement (experienced) than when self-assessment was based on proficiency (abstract) criteria.

Jassen-van Dieten's study involved 730 adult learners of Dutch as a second language who came from 73 different countries, mainly nonwestern. She compared a test of Dutch for language courses with a parallel version of that test in self-assessment format which involved reading, listening, speaking and writing tests. Students also took C-test. The correlation between self-assessment and criterion test ranged from 0.33 to 0.69, while a correlation between C-test and the criterion test was 0.44 to 0.76. From this result, it was found that the self-assessment predicted students' English skills, but C-test was a better predictor of the criterion.

Although many studies reported that there exists medium to high correlations between a self-assessment and a criterion test, some studies failed to find a relationship between the two. For example, Pierce et al.'s study investigated 500 Grade 8 students in French immersion programs in Canada. The students took French tests of four skills and answered a self-assessment questionnaire. The results showed that the correlations between self-assessments of language proficiency and tested proficiency were weak (0.01 to 0.25). Thus, self-assessment was only a weak indicator of tested proficiency.

Moreover, Brantmeier explored whether self-assessment is an accurate predictor for language placement and classroom performance in reading. The study involved 34 American university students who enrolled in an advanced Spanish class. At the beginning of the semester, they took OPLE (Online Placement Exam) which consisted of multiple-choice questions. They also answered the self-assessment questionnaire of Spanish reading skill. In order to assess subsequent reading performance, in the third week of the class, the same students answered a self-assessment questionnaire and took three kinds of reading tests: the written recall of the reading passage, sentence completion items and multiple-choice questions. The results of the regression analysis revealed no significant relationships between pre-self-assessment and OPLE, and between pre-self-assessment and three tests. Likewise, no significant association was found between post-self-assessment and OPLE, or between post-self-assessment and three tests. On the other hand, there was a significant association between the following: OPLE and recall (0.20), and OPLE and multiple-choice (0.32). Therefore, this study indicated that the OPLE was better than self-assessment for predicting actual reading performance in the course.

2.2. Culture and Self-Assessment

Some researchers point out the problem on the side of students when conducting a self-assessment questionnaire. The degree of self-assessment may vary according to culture because how people view themselves differs in different cultural contexts. For example, in European-American cultural contexts, the characteristic of high self-esteem is often seen because this is both a consequence and a prerequisite for participating in the independent, mutually approving relationship. On the other hand, in East Asian cultural contexts, self-criticism is indispensable to engage in the independent, mutually sympathetic relationship (Kitayama and Markus). Likewise, Heine et al. maintain that unlike North Americans who regard themselves positively, Japanese self-perceptions tend to be critical and self-effacing. This is because Japanese people believe that this self-criticism serves for future improvement

and achievement.

For instance, in Matsumoto and Kitayama's (cited in Kitayama and Markus) study, they asked both Japanese and American college students to judge what proportion of others in their own colleges were better than themselves in each of 20 different domains such as "considerate," "intelligent," "tolerant," "creative," and "attractive." The results indicated that Americans judged only 32.4 percent of others to be better than themselves, while Japanese had the overall means of 50.2 percent. According to Kitayama and Markus, this is because since when they are small, Americans quickly develop the habits of identifying the positive features and believe that they are slightly better than their peers. In contrast, Japanese often have negative, self-critical views due to the fact that they believe successful experiences are likely to cause negative consequences such as envy by others.

According to Farh et al., the positive view of the self brings about "the leniency effects" in terms of self-assessment, and this effect is well known in the field of human resource management in the western society. Therefore, Farh et al. investigated whether the received doctrine of leniency is supported outside the bounds of Western culture. This is because Eastern people are likely to be modest in self-ratings and refrain from exaggerating their achievements. The participants were 982 leader-subordinate pairs who were drawn from eight different organizations in Taiwan, such as an oil company, a hospital, and an automobile manufacturer. The self-ratings of their job performance were compared with supervisory rating of subordinate performance. The results showed that Taiwanese workers evaluated themselves less favorably than did their supervisors. These findings were contrary to the typical U. S. findings that self-ratings of performance were higher than supervisory ratings. Therefore, the authors concluded that Taiwanese workers exhibited a modesty bias, and that culture played an important role in shaping workers' perceptions of their own work performance.

Yu and Murphy replicated Farh et al.'s study in order to investigate whether the same modesty bias can be observed in the samples from mainland China. To provide a broader test of the modesty bias hypothesis, Yu and Murphy

obtained peer ratings as well as supervisory ratings, and compared them to self-ratings of job performance. The participants were 453 supervisor-subordinate pairs. They were industrial workers in China. The rating-scale used in Farh et al. study was utilized in this study. The results indicated that unlike Farh et al., the same pattern that is typically found in Western research was found. That is to say, self-ratings were significantly higher than supervisor or peer ratings. Accordingly, these findings casted doubt on the cultural relativity hypothesis, which was used to explain the results of Farh et al's findings.

Yu and Murphy's study is worth noticing in that their findings criticized a cultural dichotomy between East and West. Indeed, as Hwang points out, Chinese who received more formal education and are more exposed to Western culture have attitudes more consistent with individualism. This tendency is also explained by Yokoyama, who reported that nowadays many of his students in *Yobiko* (cram school) do not understand Japanese way of communication such as "*ishin-denshin*," tacit understanding. He commented that due to much exposure to Western culture, those young Japanese tend to adopt a Western communication style.

3. Purpose

Since there are few studies on Japanese students' self-assessment of their coursework, I would like to investigate if there is a relationship between Japanese students' self-assessment and their achievement in the course. In addition, I would also like to focus on a modesty bias by examining if there is difference in the accuracy of the self-assessment between students who self-assessed high and those who self-assessed low. I hypothesize that the latter group shows a modesty bias in their self-assessment, so that they are less accurate in their self-ratings than the former group.

4. Research Questions

My research questions are:

- (1) How does Japanese students' self-assessment relate to their true achievement in the course?
- (2) Is there any difference in accuracy of the self-assessment between those who self-assessed high and those who self-assessed low?

5. Method

5.1. Participants

The participants were 54 freshmen (49 males and 5 females) at a private university in Japan. All of them belonged to a science and engineering department. They were taking an English Reading/Writing class in the spring semester of 2006 taught by the author. In this class which met once a week for 90 minutes, students reviewed English grammar and translation, and mainly learned how to write English paragraph essays. No returnee was involved in this study, so all the students received a typical English education in Japan, which began from junior high school.

5.2. Materials

Twelve self-assessment questions were developed according to what the students have learned in the class. According to Oscarson, one of the features of self-assessment is that "self-assessments appear to be more accurate when subjects use assessment tools written in their respective native languages rather than in the target language" (*Foreign and Second*, 183). Therefore, I decided on the questions to write in Japanese (See Appendix for English translation version). Likert scale was used from 1 (I do not agree with the statement at all) to 5 (I strongly agree with the statement). Thus, possible scores ranged from 12 to 60.

The final test of the course was utilized to assess students' true achievement in the course. The test was created by the author who taught the course. All the questions in the exam were related to the self-assessment questions. The test consisted of grammar multiple-choice, (which relates to self-assessment questions 1 and 2), translations from Japanese into

English (Q3), summarizing an English passage in Japanese(Q4), finding a topic sentence in a paragraph (Q5), writing topic sentences and concluding sentences (Q6, 7, 8), and writing one paragraph essay in English (Q 6,7, 8,9,10,11,12). The test scores ranged from 0 to 100.

5.3. Procedures

At the last class of the course (12th week), students were asked to fill out the self-assessment questionnaire for about ten minutes. The following week, they took a final examination for 60 minutes.

The self-assessment score was an independent variable. The final test score was a dependent variable. In order to predict the test score from self-assessment, I examined data using regression. The regression analysis examined the strength of the relationship between the self-assessment and the test in addition to the amount of variables explained by the self-assessment. The Alpha level for statistical significance was set at 0.05.

6. Results

Mean score and standard deviation of each SA (self-assessment questionnaire) is shown in Table1. Total mean score and *SD* of SA are $N=54$, $M= 36.28$. $SD=8.03$. Total M and SD of the test are $N=54$, $M=65.87$, $SD=12.97$. A linear regression analysis was conducted to evaluate whether the students' self-assessment score is useful in predicting their achievement in the course, that is, the score of the final exam. More specifically, it evaluated a null hypothesis: the population correlation is equal to zero.

<i>M</i>		<i>M</i>		<i>M</i>	
Q1	3.04	Q5	3.52	Q9	3.06
Q2	2.76	Q6	3.06	Q10	3.11
Q3	3.22	Q7	3.00	Q11	2.44
Q4	3.22	Q8	3.07	Q12	2.83

Table 1. Mean scores of each self-assessment question.

Is Self-Assessment Possible? :
Japanese Students' Self-Assessment in their Course Work

The scatter plot for the two variables (Figure 1) indicates that the two variables are linearly related such that as the score of the self-assessment questionnaire (SA) increases, the test score also increases. The regression equation for predicting test score is

$$\text{Predicted test score} = 0.59 \text{ self-assessment score} + 44.60$$

Since F is significant: $F(1,52) = 7.89, p < 0.05$, and the 95 % confidence interval for the slope, 0.17 to 1.01, does not contain the value of zero, the null hypothesis should be rejected at the 0.05 level. Thus, the self-assessment is significantly related to the test and students who assessed their achievements highly tend to have better test scores. Accuracy in predicting the test score was moderate. The correlation between the self-assessment and the test was 0.36, and 13.2 % of the variance of the test was accounted for by its linear relationship with the self-assessment.

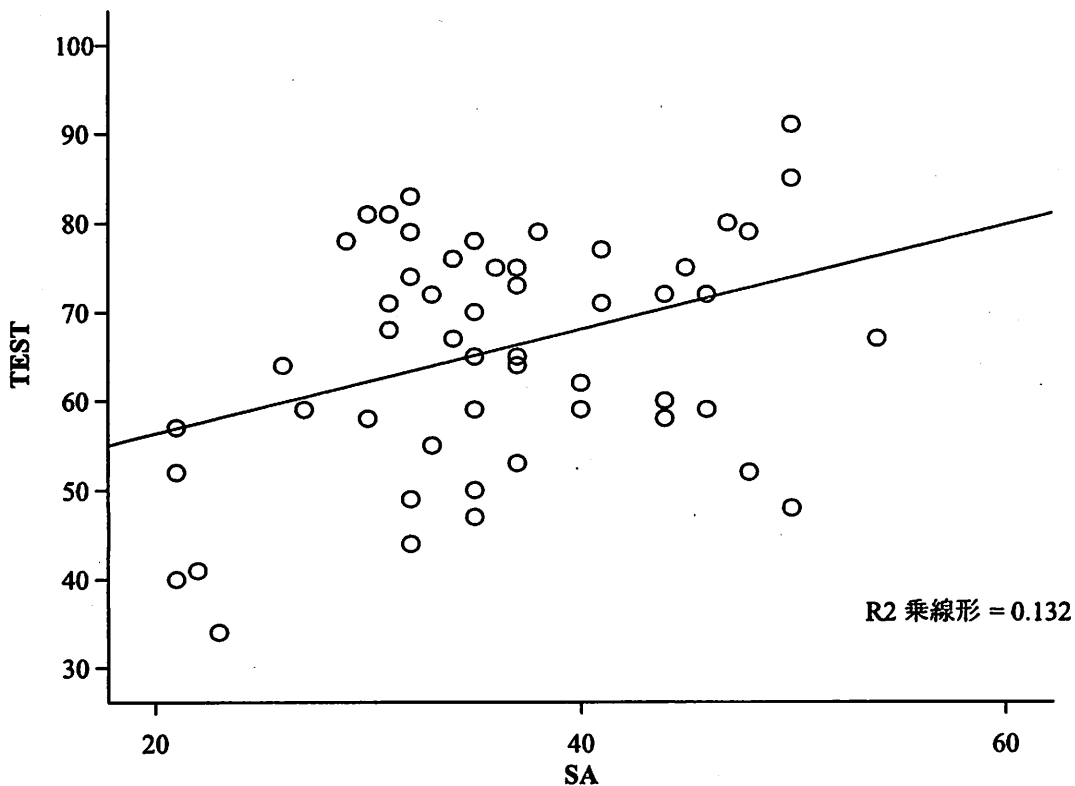


Figure 1. Scatterplot between test scores and self-assessment scores.

In order to answer the second research question, I divided the students into two groups according to their scores in self-assessment: those who scored

low, (low self-assessment group, SAL), ($N=27$, $M=30.07$, $SD=4.72$) and those who scored high, (high self-assessment group, SAH), ($N= 27$, $M= 42.48$, $SD=5.40$). The mean test score of the former group is $M=62.30$, $SD=14.33$, while that of the latter group is $M=69.44$, $SD=10.53$. When I examined each data using a regression analysis, there was an interesting result. In terms of 27 students who scored SA lower than others (SAL Group), the scatter plot is shown in Figure 2. F is significant: $F(1, 25) = 6.19$, $p < 0.05$, and the confidence interval for the slope is 0.17 to 1.01. Therefore, the self-assessment of the low group is related to the test score. The correlation increased to 0.45, and SAL can account for 20 % of the variation in test scores.

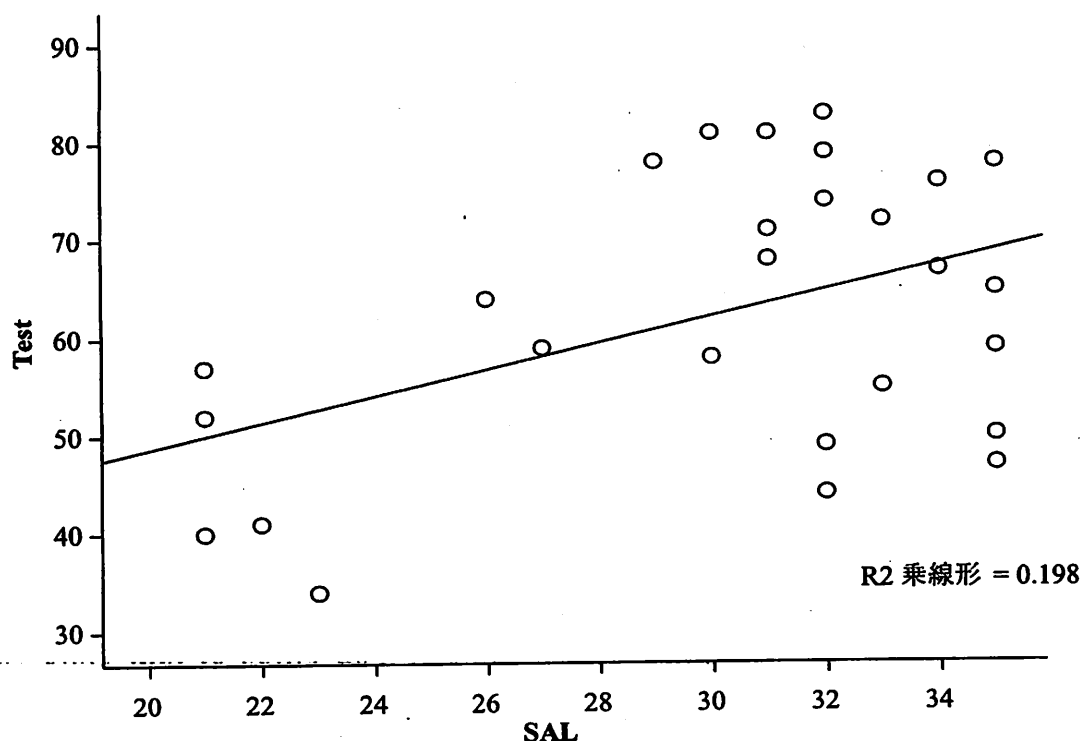


Figure 2. Scatterplot between test scores and low self-assessment scores

In contrast, SAH group (those who scored high in SA) had a different result. The scatter plot is shown in Figure 3. F is not significant for the test: $F(1, 25) = 0.005$, $p = 0.94$. Therefore, we failed to reject the null hypothesis, and SAH was not useful in predicting test scores.

Is Self-Assessment Possible? :
Japanese Students' Self-Assessment in their Course Work

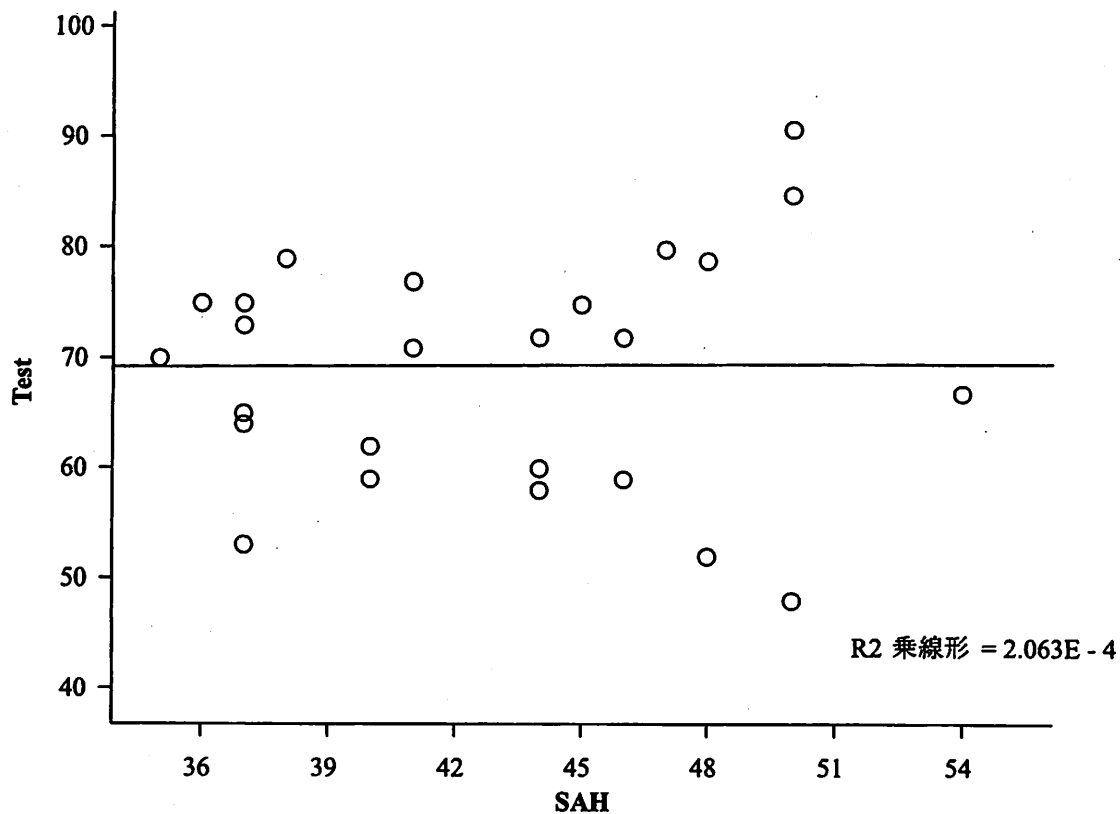


Figure 3. Scatterplot between test scores and high self-assessment scores

7. Discussion

The results of the regression analysis revealed that although moderate, the self-assessment questionnaire had a positive linear relationship with the test score. Therefore, we may state that to some degree Japanese students were able to self-assess their achievement in the course. Next, when we compared the regression analysis of the two groups: SAL and SAH, there was an interesting outcome. The test score of SAL group could be predicted by the self-assessment questionnaire, while those of SAH group could not. This result was not what I expected. If the overwhelming view that Japanese people tend to self-criticize their ability is correct, the degree of accuracy in self-assessment should be higher in SAH group than SAL group. This is because when so-called self-criticizing Japanese people self-assess their ability, they will self-assess it lower than their true ability. As they tend to give lower

scores to the things that they can do actually, low self-assessment scores may not be a good predictor of their true ability. In contrast, if their self-assessment score is high, this seems to mean that they have a great confidence in that area and that they can do it by all means. Thus, high self-assessment scores are likely to be very reliable.

However, the results of this study showed an opposite result. Those who self-assessed low were able to judge their true achievement correctly, whereas high self-assessment scores failed to show a relationship with test scores due to the fact that those who self-assessed high did not necessarily received high scores on the exam, and some of them overestimated their ability. Interestingly, this overestimation does not fall into the generalized cultural image of self-criticizing Japanese people. Consequently, these results seem to suggest that the recognition of "I cannot do it" is more trustworthy than the feeling of "I can do it." This finding relates to Bachman and Palmer's study, which showed that ESL students in the U.S. seemed to be more aware of the areas in which they have difficulty than they find easiest.

The findings of the present study seem to be consistent with Yu and Murphy, rather than Farh et al., which found that Chinese workers showed no evidence of modesty bias in self-ratings. I am not sure whether this result is due to the assumption that young Japanese came to have a western attitude and individualism through much exposure to Western culture. We need further investigation on this topic.

8. Conclusion

In the present study, a positive correlation was found between a self-assessment questionnaire and a final exam. However, it should be noted that the number of the participants was small, so it may be difficult to generalize the result. Moreover, some students' overestimation may be due to the use of a final examination. The test might have been somewhat more difficult than what they learned during the class.

Despite these limitations, the results of the study showed that the self-

assessment score was reflecting the test score. Since it was found that the recognition of "I cannot do it" is likely to be accurate, the self-assessment can be used sometime during the course to check students' understanding of each unit. It takes a smaller amount of time to conduct a questionnaire. Additionally, although many students hate taking tests or quizzes, they will not mind answering a questionnaire. Indeed, when I conducted the self-assessment questionnaire in this study, many students completed it less than five minutes, and no students complained about answering it.

Since the value of correlation was not very high in this study, it is necessary to consider how to make a self-assessment questionnaire and a test which can assess students' achievement much more accurately. Furthermore, this study dealt with only reading and writing areas, especially focusing on English paragraph essays. For the future direction, I would like to focus attention on the self-assessment of listening and speaking skills, which most Japanese students feel that they are poor at.

This study is based on the oral presentation at the Japan Association of English Linguistics and Literature Conference on March 10, 2007.

Works Cited

- Bachman, Lyle, F., and Palmer, Adrian. S. "The Construct Validation of Self-ratings of Communicative Language Ability." *Language Testing*, 6.1(1989): 14-29.
- Brantmeier, Cindy. "Advanced L2 Learners and Reading Placement: Self-assessment, CBT, and Subsequent Performance." *System*, 34 (2006): 15-35.
- Brown, James, Dean. *Testing in Language Programs*. New York: McGraw Hill, 2005.
- Farh, Jiing-LiH., Dobbins, Gregory, H., and Cheng, Bor-Shiuan. "Cultural Relativity in Action: A Comparison of Self-Ratings made by Chinese and U.S. Workers." *Personnel Psychology*, 44, (1991): 129-47.
- Green, Samuel, B, & Salkind, Neal, J. *Using SPSS for Windows and Macintosh*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.
- Heine, Steven, J., Markus, Hazel, Rose, Lehman, Darrin. R., and Kitayama, Sinobu. "Is There a Universal Need for Positive Self-regard?" *Psychological Review*, 106.4 (1999): 766-94.

- Hwang, CH. "A Study of the Personal Preference of Chinese University Students by Edwards Personal Preference Schedule." *Psychology and Education*, 1(1967): 52-67.
- Janssen-van Dieten, Anne-Mieke. "The Development of a Test of Dutch as a Second Language: the Validity of Self-assessment by Inexperienced Subjects." *Language Testing*, 6 (1989): 1-13.
- Kitayama, Sinobu, and Markus, Hazel, Rose. "The Pursuit of Happiness and the Realization of Sympathy: Cultural Patterns of Self, Social Relations, and Well-being." *Culture and Subjective Well-being*. Ed. Diener, & Suh. Cambridge : The MIT Press, 2000, 113-61.
- Kusnic, Edith, and Finley, Mary, Lou. "Student Self-evaluation: An Introduction and Rationale." *Student Self-evaluation: Fostering Reflective Learning*. Ed. MacGregor. San Francisco: Jossey-Bass Publishers, 1993, 5-14.
- LeBlanc, Raymond, and Painchaud, Gisele. "Self-assessment as a Second Language Placement Instrument." *TESOL Quarterly*, 19.4 (1985): 673-87.
- Oscarson, Mat. "Self-Assessment of Language Proficiency: Rationale and Applications." *Language Testing*, 6.1 (1989): 1-13.
- . "Self-Assessment of Foreign and Second Language Proficiency." *Encyclopedia of Language and Education. Volume 7. Language Testing and Assessment*. Ed. Clapham and Gorson. Dorrecht: Kluwer Academic Publisher, 1997, 175-87.
- Pierce, Bonny, N., Swain, Merrill., & Hart, Doug. "Self-assessment, French Immersion, and Locus of Control." *Applied Linguistics* 14 (1993): 25-42.
- Ross, Steven. "Self-Assessment in Second Language Testing: A Meta-analysis and Analysis of Experimental Factors." *Language Testing* 15, (1998): 1-20.
- Todd, Richard, Watson. "Using Self-Assessment for Evaluation." *English Teaching Forum*, (2002): 16-19.
- Yamamoto, Masahiko. *Kokoseinō tameno Ronri Sikou Training. (Logical Thinking Training for High School Students)*. Chikuma Shinsho, 2006.
- Yu, Jiayuan, and Murphy, Kevin. R. "Modesty Bias in Self Ratings of Performance: A Test of the Cultural Relativity Hypothesis." *Personnel Psychology*, 46 (1993): 357-63.

Is Self-Assessment Possible? :
Japanese Students' Self-Assessment in their Course Work

Appendix

Self-Assessment Questionnaire

(1= I do not agree the statement at all, 5= I totally agree with the statement)

- | | | | | | |
|---|---|---|---|---|---|
| 1. I can choose the most appropriate grammatical form from four choices. | 1 | 2 | 3 | 4 | 5 |
| 2. I can find an inappropriate grammatical form from four choices. | 1 | 2 | 3 | 4 | 5 |
| 3. I can translate Japanese sentences into English correctly. | 1 | 2 | 3 | 4 | 5 |
| 4. I can summarize one English paragraph essay in Japanese. | 1 | 2 | 3 | 4 | 5 |
| 5. I can find a topic sentence from a paragraph. | 1 | 2 | 3 | 4 | 5 |
| 6. I can write a topic sentence based on a given topic. | 1 | 2 | 3 | 4 | 5 |
| 7. I can write a topic sentence based on a given paragraph. | 1 | 2 | 3 | 4 | 5 |
| 8. I can write a concluding sentence based on a given paragraph. | 1 | 2 | 3 | 4 | 5 |
| 9. I can write supporting sentences which fully support a topic sentence. | 1 | 2 | 3 | 4 | 5 |
| 10. I can use appropriate connectors in a paragraph essay. | 1 | 2 | 3 | 4 | 5 |
| 11. I can write English sentences with few grammatical mistakes. | 1 | 2 | 3 | 4 | 5 |
| 12. I can write English sentences with few misspellings. | 1 | 2 | 3 | 4 | 5 |