

## An Investigation of the Use of Self-Assessment with Young Foreign Language Learners

Jirada Wudthayagorn, et al.\*

### ABSTRACT

*This study reports on the use of self-assessment and examines its reliability as a language assessment tool used by American elementary school students who studied Japanese as a foreign language since kindergarten. We examine whether the students in different grades and at different levels of language proficiency provide accurate assessments of their developing language abilities. The results indicate that whether or not the students are able to assess their language abilities accurately seems to be based on at least two factors: the numbers of years in the language program and the level of language proficiency. However, we argue that it may be premature for every student to use self-assessment as a method to identify his/her linguistic proficiency, because the student's linguistic performance may not be well-developed to determine his/her own performance. It is believed that self-assessment is appropriate for informative purposes of language learning not for evaluation.*

---

\* Richard Donato, G. Richard Tucker, and Kanae Igarashi

## **Introduction**

Research attention on the issue of self-assessment of second and foreign language proficiency has recently found its way into the language testing literature over the past few years (e.g., Bachman & Palmer, 1989; Heilenman, 1990; MacIntyre, Noels, & Clement, 1997; Oskarsson, 1980). Peirce, Swain, and Hart (1993) argue that this increasing attention results from a current trend in second/foreign language teaching and learning that shifts its focus from a teacher-centered to a learner-centered classroom. It is not surprising, therefore, that, as a consequence, research on assessment of second language ability has also been expanded to include investigations of the learner's ability to assess the self. These studies of learner-centered assessments have been conducted primarily on young children learning a second language (e.g., Peirce, Swain, & Hart, 1993) or a foreign language (e.g., Donato, Antonek, & Tucker, 1994, 1996; Donato, Tucker, Wudthayagorn, & Igarashi, 1999; Tucker, Donato, & Antonek, 1996).

Parallel to this increasing attention on self-assessment, the recent federal legislation, "Goal 2000: Educate America Act," requires American students to leave grades 4, 8, and 12 with competence over challenging subject matter including foreign languages. According to Tucker et al. (1996), this Act generates five major issues concerning foreign language education in the elementary school: (1) which model of instructions to implement (e.g., an immersion or a FLES model), (2) the appropriate age to begin language study, (3) the choice of languages, (4) realistic proficiency expectations, and (5) how best to assess the language proficiency of young children. While Tucker et al. (1996) have already addressed the first and fourth issues, this study will focus on the fifth issue by examining the use of self-assessment among young children who learn Japanese as a foreign language.

It is hoped that this study will expand the body of knowledge and lead to a better understanding of self-assessment among young children.

## **Review of Self Assessment Literature**

Oscarson (1989) states that students' foreign language ability can be assessed from two perspectives: assessment from teachers or trained examiners and assessment

from the students themselves. In Oscarson's (1989) view, the former assessment is other-directed activity, whereas the latter, self-assessment, is an internal, or self-directed activity, in which the students evaluate themselves from their own experiences and perspectives. Upshur (1975, as cited in Oskarsson, 1980, p.14) believes that "...most of us could give a pretty accurate appraisal of our own abilities in other languages that we know."

Language learners can take advantage of self-assessment because it promotes learning and increases learners' involvement in the process of language learning (Heilenman, 1990; Oskarsson, 1980; Peirce et al., 1993). Oscarson (1989) explains that, through the use of self-assessment, learners are able to evaluate and adjust their progress, which ultimately affects language learning and achievement. In addition, self-assessment practices appear to enhance learners' motivation (Blanche & Merino, 1989), because learners take active roles in examining their own progress and future needs in learning. Like language learners, language teachers can make use of self-assessment.

Heilenman (1990) points out that, within a short time and limited budget, the use of self-assessment can indicate substantial language ability. Oscarson (1989) considers self-assessment a mutual responsibility between teachers and adult learners because teachers and adult learners frequently see things differently and establish different priorities. Thus, as Oscarson (1989) claimed, mutual responsibility is likely to lead to the democratic development of classroom instruction.

Janssen-van Dieten (1989) reminds us that "the main purpose of self-assessment... is not its application for selection, but its positive influence on the learning process" (p.44). In contrast, Heilenman (1990) argues that self-assessment has been used as a satisfactory estimation for language ability as assessed by objective measures. Oscarson (1997) concludes that the use of self-assessment by mature learners for diagnostic monitoring of progress is appropriate, whereas for grading, promotion, and certification purposes it seems to be inappropriate.

Thus, based on the thinking of previous researchers, it appears that self-assessment has multiple purposes. For language learners, it is a self-evaluation

instrument which increases learners' involvement and motivation in language learning. For language teachers and testers, self-assessment can be used as an approximation to learners' language ability. In this vein, if learners are well-trained and able to assess their language ability accurately, the application of self-assessment for selection and placement is beneficial (Heilenman, 1990).

In spite of the encouraging discussions of many researchers concerning self-assessment and its ability to motivate students' learning and provide indices of language achievement, the research findings concerning the relationship between self-assessment and other independent measures such as an oral interview or instructor's assessment are contradictory. In several studies, Pearson-product moment correlation coefficients were calculated to indicate the relationship between self-assessment and other measures. Blanche (1990) states that the correlation coefficients ranging from 0.50 to 0.60 are common, and higher values are rare. For example, Oscarson (1978) found that self-assessment and instructor's assessment were correlated at 0.60, while self-assessment and test proficiency were correlated at only 0.50.

Additionally, LeBlanc and Painchaud (1985) reported that among 500 Canadian university students, English self-assessment and placement scores were correlated at 0.53. Peirce et al. (1993) concluded that self-assessment correlated weakly with objective measures. Moreover, Peirce et al. (1993) noted that self-assessment on specific tasks was more highly correlated with tested proficiency than were global self-assessment measures. This finding contradicts the recent investigation of Donato et al. (1999) on elementary school children's self-assessment. In this study, significant correlation was found on global scores only. When subsections of the self-assessment (e.g., sentence-level production, expression of one's needs and feelings, or story-telling ability) were compared to sections of an oral interview, discrepancies between student self-assessment and actual performance were found. On the other hand, Janssen-van Dieten (1989) found no correlation between a test of Dutch as a second language and a parallel version of that test in self-assessment format.

Despite disconfirming evidence, there are a few studies which are more optimistic about self-assessment. For example, Bachman and Palmer (1989) examined the trait

structure of a self-rating test of communicative language ability through the use of the multitrait multimethod (MTMM) design with confirmatory factor analysis (CFA). With this particular design and analysis, they discovered that self-ratings were a reliable and valid measure of language ability.

In this study, Bachman and Palmer (1989) measured three linguistic abilities (i.e., grammatical competence, pragmatic competence, and sociolinguistic competence) with three types of questions: (a) "ability" questions (e.g., Do you use different kinds of English depending on the person you are using it with such as a child, a close friend, or a teacher?), (b) "difficulty with production" questions (e.g., How often do you think you don't know enough English words?), and (c) "recognition of input" questions (e.g., Can you tell how polite English speaking people are by the kind of English they use?). The participants were 116 adult non-native speakers in Utah. Of the three types of questions, they found that the most effective were the "difficulty with production" questions and the least effective were the "recognition of input" questions.

Bachman and Palmer's (1989) study suggests that types of questions play a role in determining the effectiveness of the use of self-assessment. Blanche and Merino (1989) also points out that the questions that seem to have yielded the most accurate answers contain descriptions of concrete linguistic situations that the learner can assess in behavioral terms.

Interestingly enough, Blanche and Merino (1989) and Blanche (1990) note that good learners appear to underrate their language abilities. Blanche (1990) also states that cases of overestimation involve low achieving students more often than high achieving students. Heilenman (1990) concludes that novice learners often overestimate themselves because they have little or no way of being aware of what they do not know or are able to do. She also explains that the more experience in language learning the learners have, the more they are aware of the limits of their skills and knowledge in the target language. Corroborating this assertion, Peirce et al. (1993) contend that early immersion students tend to give higher self-assessments of ability than middle immersion students.

In summary, research studies supporting self-assessment as a reliable and valid indicator of language ability present some important caveats. Specifically, the accuracy of self-assessment seems dependent upon the learner (e.g., the achievement levels--high vs. low, young vs. old), the nature of the assessment tasks given (can do vs. can't do, peers vs. unidentified others), and the modality assessed (production vs. comprehension).

### **Purpose of the Study**

The purpose of this study is to report on the use of self-assessment and examine its reliability as an assessment tool used by American elementary school students who studied Japanese as a foreign language since kindergarten. This study is guided by three research questions.

1. How do the students in grades 4, 5, and 6 assess their language skills?
2. How do the students at different language proficiency levels (i.e., novice-high (NH), novice-mid (NM), and novice-low (NL)) assess their language skills?
3. How do the self-assessment scores relate to language proficiency levels identified by independent interviewers?

We hope that this study will make a useful contribution to other schools that offer similar programs and are concerned about assessment issues and to language testing researchers and second/foreign educators who are interested in self-assessment issues in young children.

### **The Study**

#### **Participants**

The participants of this study were American elementary school students at Falk School, the laboratory school of the University of Pittsburgh. These students were in a Japanese FLES (Foreign Language in Elementary School) program, that is, they had been studying Japanese as a foreign language since kindergarten.

Forty-eight elementary school students (25 boys, 23 girls) participated in this study. The students in grades 4, 5, and 6 had studied Japanese continuously since

kindergarten; therefore, they had been in the Japanese FLES program for 5, 6, and 7 years, respectively. Nine of them (5 boys, 4 girls) were in grade 4; 19 of them (8 boys, 11 girls) were in grade 5; and 20 of them (12 boys, 8 girls) were in grade 5. Eleven grade 5 students (2 boys, 9 girls) and 18 grade 6 students (10 boys, 8 girls) were selected to receive the Pro-I.

### **Materials and Procedures**

**Student Self-Assessment.** A student self-assessment questionnaire used in this study was modified from the one used in Donato et al.'s (1999) study. The self-assessment questionnaire consists of 13 items on a 4-point Likert scale and three open-ended questions (see Appendix 1). Twelve items on a 4-point Likert scale are classified into seven developing language skills and the other one is classified as comfort in a Japanese class.

The seven developing language skills are (1) Comprehension at a word level (Com-W), (2) Comprehension at a sentence level (Com-S), (3) Comprehension at a discourse level (Com-D), (4) Pronunciation (Pronun), (5) Production at a word level (Pro-W), (6) Production at a sentence level (Pro-S), (7) Production at a discourse level (Pro-D). Also, the item coping with comfort in Japanese is also included in this self-assessment.

The students were asked to think about their own ability in Japanese--whether they can understand and say different kinds of things in Japanese. They were reminded that they did not have to speak like the Japanese teacher. The student self-assessments were administered by the Japanese teacher during the fourth week of May 1999. (The number of items and prompts for each language skill can be seen in Appendix 2.)

**Prochievement interview (Pro-I).** The Pro-I was divided into five sections: (1) warm-up, (2) grammaticality judgement, (3) comprehension, (4) production, and (5) wrap-up. The warm-up and wrap-up sections were not taken into account in the scoring. Similar to the Pro-I procedures that were used in the past years (see Donato et al., 1994, 1996; Tucker et al., 1996; or Donato et al., 1999), interviewers used a "Student Observation Form" (SOF) to assign scores ranging from one to five on each

dimension of linguistic abilities: comprehension, fluency, vocabulary, pronunciation, and grammar.

For example, a score of one on grammar indicated that the students responded primarily with one-word answers. When the students attempted longer answers, grammar and word order problems made speech very difficult to understand. A score of five on grammar indicated that the students frequently attempted multiple-word answers. These answers generally exhibited correct grammar and word order.

The students were divided into two groups. Each group was interviewed in a separate room by two interviewers. While one interviewer interviewed a student, the other observed, took notes, and made comments on the SOF. Afterwards, both interviewers discussed and identified the students' level of language proficiency (i.e., novice-low, novice-mid, novice-high, intermediate-low, intermediate-mid, and intermediate-high). The interview lasted 15 to 20 minutes.

Twenty-nine out of forty eight students (about 60%) received the Pro-I at the end of May 1999. The levels of students' language proficiency are as follows:

The highest level of language proficiency, identified by the interviewers, was novice-high (NH) which included four students from grade 6 only. Seventeen students (7 in grade 5, 10 in grade 6) were novice-mid (NM). Eight students (4 in grade 5, 4 in grade 6) were novice-low (NL).

### **Data Analysis**

Mean scores of each language skill from 13 items were calculated and then the line graphs were plotted. Line graphs were used to present the data analysis because they allow us to compare the mean scores across groups of NH, NM, and NL students. Thus, we are able so see the differences of their self-assessment mean scores of each language skill more clearly.

It is important, however, to note that the findings may not be able to generalize to all early language learners. Self-assessment should be further studied in various contexts.



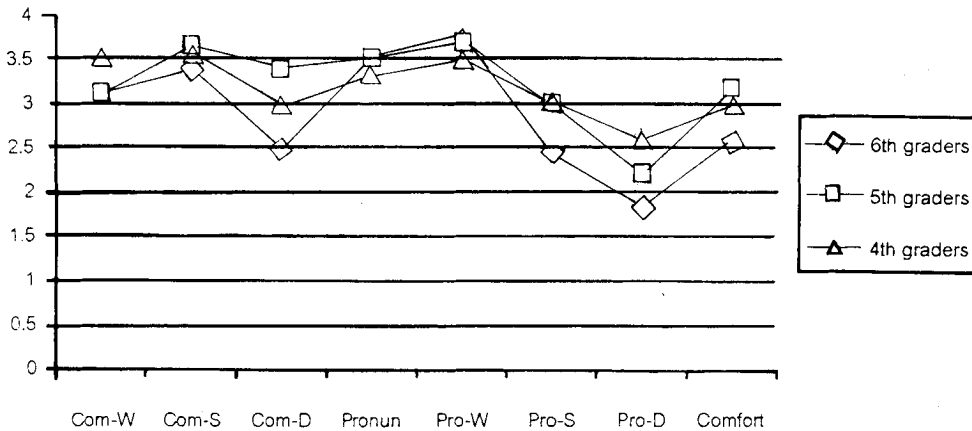
In addition, in order to be relevant to the research questions, we chose not to analyze the answers from the open-ended questions.

## Findings

In this section, the researchers will discuss the findings according to three research questions described previously.

### 1. How do the students in grades 4, 5, and 6 assess their language skills?

The following figure presents self-assessment scores of students in grades 4, 5 and 6 on their language skills



**Figure 1:** Self-assessment scores of students in grades 4, 5, and 6 on different language skills

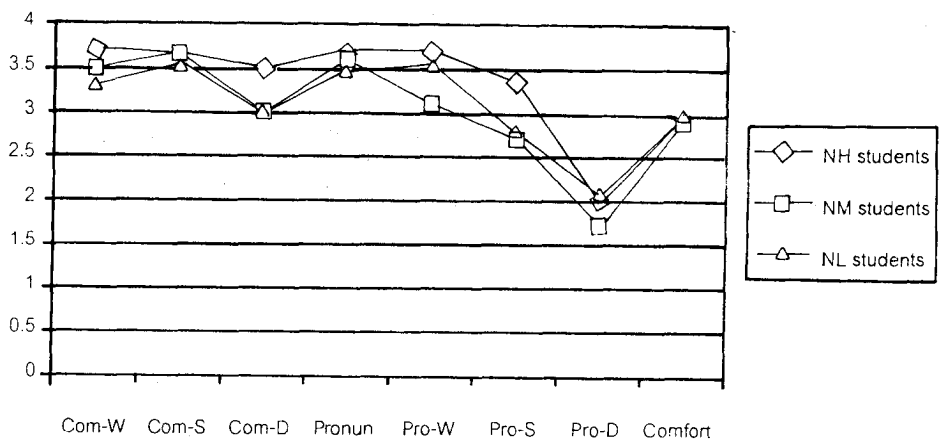
Forty-eight students in grades 4, 5, and 6 who had participated in the Japanese FLES program since kindergarten completed self-assessment questionnaires. Based on Figure 1, a clear pattern of self-assessment among students in grades 4, 5, and 6 cannot be observed. In general, the students in grade 6 assess their language abilities less positively than the students in grades 4 and 5. The pattern is reversed in the case of pronunciation (Pronun) in which the students in grade 6 assess themselves as equal to those in grade 5 at the score of 3.53, which is higher than those in grade 4 whose scores are at 3.33. The reversed pattern is also found in word-level production (Pro-W)

in which the students in grade 6 assess themselves more positively than those in grades 5 and 4, respectively. A distinctive case of production at discourse level (Pro-D) can be observed in that students in grade 4 assess themselves most positively, followed by students in grade 5 and 6 respectively. Moreover, across seven language abilities, students in grade 6 assess themselves most positively on production at word level (Pro-W) and students in grades 4 and 5 assess themselves most positively on comprehension at sentence level (Com-S). Students in all grade levels give the lowest scores on production at discourse level (Pro-D).

Interestingly enough, the number of years that the students in grade 6 have been exposed to Japanese is related negatively to the way they assess their language abilities. The more they have been in the Japanese FLES program, the less they give self-assessment scores. By comparing self-assessment scores of students in grade 4 and 5, those in grade 5 assess themselves more positively than those in grade 4 on every language skills, except for production at discourse level (Pro-D) in which those in grade 4 give higher scores.

**2. How do students with different language proficiency levels (e.g., novice-high, novice-mid) assess their language skills?**

The following figure presents self-assessment scores of NH, NM, and NL students on different language skills.



**Figure 2 :** Self-assessment scores of NH, NM, and NL students on different language skills

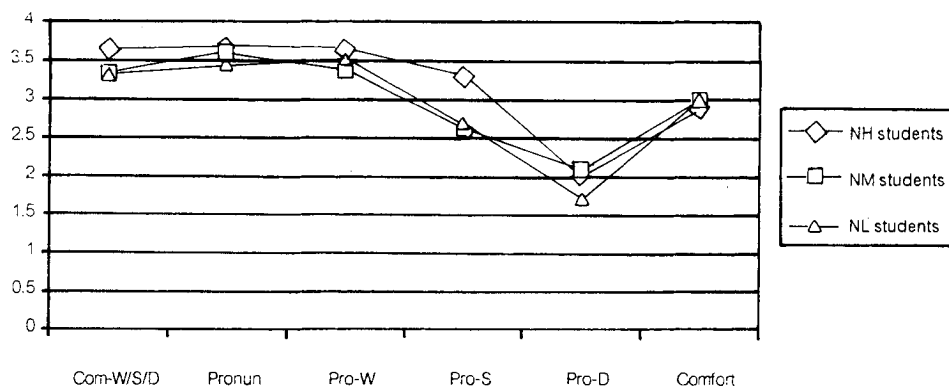
In general, the self-assessment scores of NH, NM, and NL students on different language skills occur to some extent in a patterned way. That is to say, across students' levels of proficiency, the self-assessment scores indicating comprehension at word level (Com-W), and at sentence level (Com-S) are higher than those for comprehension at discourse level (Com-D). The self-assessment scores of NH, NM, and NL students for pronunciation (Pronun) are higher than those for production at word level (Pro-W), at sentence level (Pro-S), and at discourse level (Pro-D). Moreover, their self-assessment scores begin to drop at word-level production (Pro-W) and drop sharply at discourse-level production (Pro-D). Furthermore, NH, NM, and NL students are in agreement that they are comfortable speaking in a Japanese class (Comfort).

Concerning each group of students' proficiency, NH students rate themselves the highest scores across different language skills, except for the case of production at discourse level (Pro-D) in which NH students assess themselves less positively than NM, but higher than NL students. NM students give themselves the lowest scores for production at word level (Pro-W), at sentence level (Pro-S), and at discourse level (Pro-D). NM students can assess themselves higher than NL students in some language skills such as comprehension at word level (Com-W) and at sentence level (Com-S). Self-assessment scores of NL students occur in an unpredictable way. For example, while NL students gave self-assessment scores on comprehension at word level (Com-W) and at sentence level (Com-S) lower than NM students, NL students gave self-assessment scores on production at word level (Pro-W) and at sentence level (Pro-S) higher than NM students.

It is interesting to note that NH students (i.e., four students in grade 6) identify their language skills differently from and higher than the rest of the students. That is, some students who had been in the Japanese FLES program longest and achieved a higher rating on the Pro-I assessed their language skills generally higher than and differently from other students who had not been in the Japanese program as long and had achieved a relatively lower level of language proficiency.

### 3. How do the self-assessment scores relate to language proficiency levels identified by interviewers?

These questions can be answered by Figure 3 and Figure 4. Figure 3 presents the self-assessment scores of NH, NM, and NL students on different language skills. Figure 4 presents the Pro-I scores of NH, NM, and NL students.



**Figure 3 :** Self-assessment scores of NH, NM, and NL students on different language skills

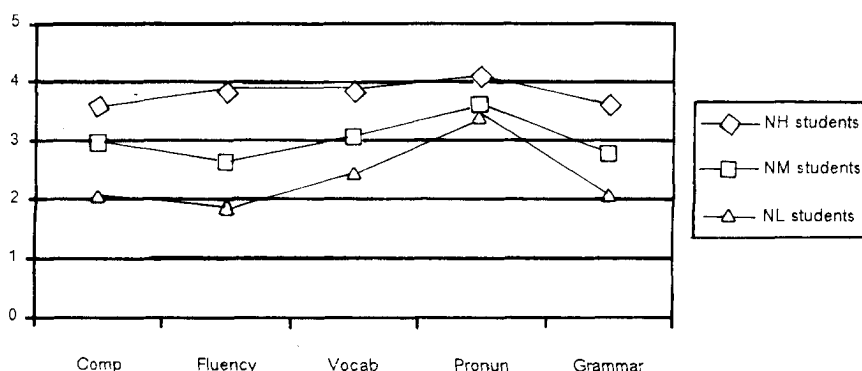
[Note: Comp-W/S/D = Comprehension at word level, sentence level, and discourse level]

Across proficiency levels, the students assess their language skills in an identical way in that they report that they are good at comprehension (Com-W/S/D), pronunciation (Pronun), and production at word level (Pro-W), but weak at production at sentence level (Pro-S) and at discourse level (Pro-D). Although self-assessment scores of NH, NM, and NL students are identical, they are not comparable because they are not ranked orderly from low to high.

That is to say, self-assessment scores on comprehension (Com-W/S/D) and pronunciation (Pronun) are classified systematically and hierarchically from NH to NM to NL students. Beyond comprehension (Com-W/S/D) and pronunciation (Pronun), such a systematic and hierarchical classification from NH to NL students does not longer exist. NL students assess themselves more positively than NM students on production at word level (Pro-W) and production at sentence level (Pro-S), and

NM students assess themselves more positively than NH students on production at discourse level (Pro-D). In addition, three groups of the students assessed themselves quite similarly in terms of comfort in speaking Japanese in class (Comfort).

Now we will turn our attention to how the independent interviewers assigned the proficiency scores on different language skills to NH, NM, and NL students.



**Figure 4 :** Pro-I scores of NH, NM, and NL students on different language skills

The Pro-I scores of different language abilities were rated by the independent interviewers. The classification of language abilities of NH, NM, and NL students occurs in a patterned way. NH students receive the highest scores, followed by NM students and NL students across five language skills: (a) comprehension (Comp), (b) fluency (Fluency), (c) vocabulary (Vocab), (d) pronunciation (Pronun), and (e) grammar (Grammar). Across three groups of the students, pronunciation scores are the highest. Within NM and NL groups, fluency scores are the lowest. Within NH group, grammar scores are the lowest. However, such a clear pattern of classification was not found when NH, NM, and NL students assessed themselves across different language skills, except for the case of comprehension and pronunciation in which self-assessment scores of NH, NM, and NL students are systematically and hierarchically sequenced. (see Figure 3).

## Discussion

Thus far, we have investigated the students' self-assessment from two perspectives. First, we examined how the students in different grades assessed themselves. Put differently, we examined how the number of years exposed to Japanese was related to their self-assessment. Second, we took a different perspective to examine how the students assessed themselves in relation to their language proficiency identified by the interviewers. Then, we investigated how students' self-assessment scores related to language proficiency levels as identified by the independent interviewers.

As mentioned previously, the students who have been exposed to Japanese the longest (e.g., grade 6 students) assessed their developing language skills the lowest (see Figure 1), while those in grades 4 and 5 who gave higher scores for all skills. Peirce et al. (1993) noted that early immersion students assessed themselves higher than middle immersion students. Our findings agree with Peirce et al. (1993) in the sense that older students (grade 6) assessed themselves lower than younger students (grades 4 and 5). Heilenman (1990) also noted that novice learners often overestimate their language skills. If we consider students in grades 4 and 5 as beginning learners because they have been exposed to Japanese instructions relatively less than grade 6 students, Heilenman's (1990) notion receives support that novice learners often overestimate their language skills.

A distinctive case of production at discourse level (in Figure 1) in which students in grade 4 gave themselves higher scores than those in grades 5 and 6 requires comment. Length of time that students in grade 4 have been in the Japanese FLES is shorter than that of students in grades 5 and 6. Students in grade 4 may not be fully exposed to Japanese and complicated language tasks such as retelling a story. Thus, they may not realize how limited their language skills and knowledge are and they have not experienced the difficulty of such tasks. As a consequence, their self-assessment scores on production at discourse level were inflated. Unlike students in grade 4, students in grade 6, who have been in the program longest, have been already experienced a variety of language tasks, are aware of the limits of their skills and

knowledge in the Japanese language. As such, they underrated their language skills. In short, length of time in the program, or in other words-- time of task, plays a role to students' self-assessment.

In terms of proficiency levels, NH students, as identified by independent interviewers, generally assessed their developing language skills the highest (see Figure 3). In other words, these NH students (e.g., four grade 6 students) were able to identify their developing language skills in general higher than and differently from NM and NL students. Blanche and Merino (1989) and Blanche (1990) stated that good learners appear to underrate their language abilities. Unlike Blanche and Merino (1989) and Blanche (1990), our findings reveal that the students who had been in the program longest (e.g., grade 6 students) underrated their language skills, but the combination of achieving high levels of proficiency (e.g., novice-high) and being exposed to Japanese FLES program relatively long (e.g., 7 years) enables some of the students to assess themselves higher than and differently from the other students.

It seems that whether or not the students are able to assess their language abilities accurately is likely to depend on at least two factors: (1) the numbers of years in the program and (2) the levels of language proficiency. The facts that they have been in the Japanese FLES program relatively long and their language proficiency was relatively high may be crucial in helping them be aware of their own language skills. The follow-up study of self-assessment by students in grades 4 and 5 is important if we are to understand whether the number of years and levels of proficiency play an interactive role in determining the accuracy of self-assessment.

We also found that, in general, across different language skills, self-assessment scores of NH, NM, and NL students did not occur in a patterned way comparing to the Pro-I scores, as identified by the independent interviewers. As discussed, only NH students appeared to be able to rate their language skills accurately when compared with the ratings assigned by the independent interviewers. NM and NL students were not able to do so. Heilenman (1990, p. 177) reasons that "...less proficient learners who, in general, will also have less information upon which to base their answers than will their more proficient counterparts."

It is, however, noteworthy that when NH, NM, and NL students assessed their less complicated language skills such as comprehension (Com-W/S/D) and pronunciation (Pronun), the classified pattern was found. When the skills are more complicated (e.g., production), the pattern was no longer found (see Figure 3). Briefly stated, the independent interviewers were able to identify students' language proficiency from high to low across language skills in a patterned way. NH, NM, and NL students could do so only when the skills they assessed were less complicated. In other words, the level of proficiency can assist the students in assessing their pronunciation and production at word level-language skills that do not require complex processing. When the language skills they assessed become more complicated, the more advanced students (in this case, NH students who have been in the program longest) accurately assessed themselves. NM and NL students who have not developed enough linguistic competence could not assess accurately language skills that require complex processing beyond pronunciation and comprehension.

Interestingly, self-assessment scores given by the students cannot be generally compared with the Pro-I scores assigned by the independent interviewers. In contrast, Donato et al.(1999) found a significant correlation between self-assessment scores of the students in grade 4, 5, and 6 and their Pro-I scores. This means that the students were able to globally evaluate their language abilities similarly to the interviewers. Such a correlation, however, was not found in this study. But a prominent exception was found that the students who have been in the program longer (e.g., 7 years) and have reached a certain level of language proficiency (e.g., novice-high) were able to assess themselves accurately compared with the interviewers' assessment on the Pro-I. It is noted that inconsistent relationships between self-assessment scores and Pro-I scores over the years of the Japanese FLES study suggest that self-assessment may not be appropriate as a proficiency estimator. Self-assessment needs to be viewed as one piece of information in a student's profile of achievement.



## Conclusion

So far, we have seen that, across levels of language proficiency, the students assessed their language skills similarly in that they reported that they were good at comprehension, pronunciation, and word-level production. The self-assessment scores dropped at the sentence level, and dropped sharply at the discourse level (i.e., retelling a story). Moreover, the students were in agreement that they were comfortable speaking in class. Therefore, language teachers may take such information into consideration when making a lesson plan. For example, language teachers may pay close attention to speaking instruction at the sentence and/or discourse levels.

Even though self-assessment has been used as a satisfactory approximation to language abilities (Heilenman, 1990), we argue that it is premature for all students, particularly those in the Japanese FLES program, to use self-assessment as a method to identify their linguistic proficiency. This is because the students' linguistic competence may not be sufficiently developed to determine their own performance (Heilenman, 1990). Todd (2002) also concludes that some self-assessment is appropriate for learning purposes not for evaluation purposes.

## References

- Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing* 6 (1), 14-25.
- Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: The DLIFLC study. *Language Testing* 7 (2), 202-229.
- Blanche, P. & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning* 39, 313-340.
- Donato, R., Antonek, J. L., & Tucker, G. R. (1994). A multiple perspectives analysis of a Japanese FLES program. *Foreign Language Annals* 27 (3), 365-377.
- Donato, R., Antonek, J. L., & Tucker, G. R. (1996). Monitoring and Assessing a Japanese FLES program: Ambiance and Achievement. *Language Learning* 46 (3), 497-528.

- Donato, R., Tucker, G. R., Wudthayagorn, J., & Igarashi, K. (1999). Converging evidence: Attitudes, achievement and instruction in the later years of FLES. **Foreign Language Annals** 33, 377-394.
- Heilenman, L. K. (1990). Self-assessment of second language ability: The role of response effect. **Language Testing** 7 (2), 174-201.
- Janssen-van Dielen, A. (1989). The validity of self-assessment by inexperienced subjects. **Language Testing** 6(1), 30-46.
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. **TESOL Quarterly** 19, 673-687.
- MacIntyre, P. D., Noels, K. A., & Clement, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. **Language Learning** 47 (2), 265-287.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rational and applications. **Language Testing** 6 (1), 1-13.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham & D. Corson (Eds.), **Encyclopedia of language and education** (pp. 175-187). The Netherlands: Kluwer Academic Publishers.
- Oskarsson, M. (1980). **Approaches to self-assessment in foreign language learning**. NY: Pergamon Press.
- Peirce, B. N., Swain, M., & Hart, D. (1993). Self-assessment, French immersion, and locus of control. **Applied Linguistics** 14 (1), 25-42
- Todd, R. W. (2002). Using self-assessment for evaluation. **Forum** 40, 17-19.
- Tucker, G. R., Donato, R., & Antonek, J. L. (1996). Documenting growth in a Japanese FLES program. **Foreign Language Annals** 29, 539-550.

## Appendix 1: Self-assessment questionnaire

### PART ONE

We would like to know what YOU think about your ability in Japanese. In Part One, please respond to the statements by checking the answer that best describes what you think you can or cannot do in Japanese.

In Part Two, you should write your answers. When thinking about your ability in Japanese, remember that you do not have to speak like Sensei! There are no right or wrong ways to respond to the statements and questions -- just your own opinion of your ability to understand and say different kinds of things in Japanese.

- 1) I can follow instructions in Japanese, for example "Sit down," "Raise your hand," and "Look at me."  
 DEFINITELY YES  
 PROBABLY YES  
 SORT OF BUT NOT TOTALLY  
 NOT AT ALL
- 2) I can understand the names of lots of things in Japanese (for example, classroom objects, members of the family, colors, numbers, animals, etc.).  
 DEFINITELY YES  
 PROBABLY YES  
 SORT OF BUT NOT TOTALLY  
 NOT AT ALL
- 3) I can say "hello," and tell someone my name in Japanese.  
 DEFINITELY YES  
 PROBABLY YES  
 SORT OF BUT NOT TOTALLY  
 NOT AT ALL

- 4) I can say the names of lots of things in Japanese (for example, colors, numbers, family members, common objects).
- \_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 5) I can say sentences in Japanese, for example “I go to Falk School,” or “There is a pencil on the table.”
- \_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 6) I can give instructions in Japanese (e.g., “sit down,” “be quiet”).
- \_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 7) I can look at a picture of everyday life (for example, a classroom, a house, a school scene) and describe in Japanese what I see.
- \_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 8) I can retell a story in Japanese that I am familiar with, such as a fairy tale like The Peach Boy, Kintaro, or The Monkey and The Crab.
- \_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL

- 9) I feel comfortable speaking Japanese in class.  
\_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 10) I can talk about how I am feeling in Japanese (for example, maybe you can say something like "I am hot," "I have a headache," "I have a stomachache," or "I am tired.")  
\_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 11) I can ask questions in Japanese to a student from Japan who is visiting my class (for example, What is your name?, Where do you live?, Who is your favorite music group?, When is your birthday?, etc.)  
\_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 12) I can pronounce Japanese the way my teacher has taught me.  
\_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL
- 13) I can understand a story when told to me in Japanese.  
\_\_\_ DEFINITELY YES  
\_\_\_ PROBABLY YES  
\_\_\_ SORT OF BUT NOT TOTALLY  
\_\_\_ NOT AT ALL

## PART TWO

Please read the following questions and write your responses. There are no right or wrong responses. Just tell us what you think!

- 1) What do you think you know BEST in Japanese?
- 2) What do you think you still need to learn so that you can understand and speak Japanese better?
- 3) Can you talk about Japanese culture in Japanese? Can you give a few examples of what you can discuss (for example, stories, Sumo wrestlers, Japanese food, holidays, and songs)?

### Appendix 2: Language skills, number of items, and prompts in self-assessment questionnaires.

Language Skills	# of Items	Prompts
1. Comprehension at a word level (Com-W)	1	"I can understand the names of lots of things in Japanese..." (Item 2)
2. Comprehension at a sentence level (Com-S)	1	"I can follow instructions in Japanese, for example, 'sit down,' 'raise your hand,' and 'look at me.' (Item 1)
3. Comprehension at a discourse level (Com-D)	1	"I can understand a story when told to me in Japanese." (Item 13)
4. Pronunciation (Pronun)	1	"I can pronounce Japanese the way my teacher taught me." (Item 12)
5. Production at a word level (Pro-W)	2	"I can say 'hello' in Japanese and tell someone my name," (Item 3) and "I can say the names of lots of things in Japanese" (Item 4)
6. Production at a sentence level (Pro-S)	5	"I can say sentences in Japanese..." (Item 5) "I can look at a picture of everyday life ... and describe in Japanese what I see." (Item 7)

“I can give instructions in Japanese ...”

(Item 6)

“I can talk about how I am feeling in Japanese ...” (Item 10)

“I can ask questions in Japanese...”

(Item 11)

- 7. Production at a discourse level      1  
(Pro-D)
- 8. Comfort in a classroom              1  
(Comfort)

“I can retell a story in Japanese after I hear it...” (Item 8)

“I feel comfortable speaking Japanese in class.” (Item 9)