Features of Oral Proficiency in Task Performance by EFL and JFL Learners

Noriko Iwashita
The University of Queensland

1. Introduction[1]

Many foreign language learners study a foreign language wishing to become fluent and accurate, including using a range of sophisticated structures and vocabulary. Yet second language acquisition (SLA) specialists are still unsure of what makes up the myriad factors that contribute to second language (L2) proficiency. To establish a baseline view of the complex interactions amongst traits, the researchers (Iwashita, Brown, McNamara, & O'Hagan, 2008; Norris & Ortega, 2000; Ortega, 2003) focused on four key traits: syntactic complexity, grammatical accuracy, lexical diversity, and fluency. However, measures for these four traits vary from study to study, as do the diverse definitions of traits used by different researchers. An additional problem is that most investigations have concerned the acquisition of English. Because of this, it is not clear to what extent research findings from English may be applicable to languages which are typologically distant from English, such as Japanese. This article presents the results of in-depth analyses of the above four traits of oral proficiency of learners of English and Japanese as a foreign language, and how these traits contribute to oral proficiency.

2. Oral proficiency

Characteristics of speakers who are regarded as proficient are often termed as "good," "fluent," "knowledgeable," "bilingual," "competent" and so on. However, it is not always clear what speaking proficiency entails; the term may be used differently from researcher to researcher (Galloway, 1987; McNamara, 1996). A considerable volume of literature in both language assessment and SLA has investigated characteristics of oral proficiency. Some studies such as Adams (1980), Higgs and Clifford (1982) have explored proficiency based on scores awarded from rating scales and feedback on ratings collected from teachers and experts employing qualitative approach, while others such as Magnan (1988) and Larsen-Freeman (2006) have conducted in-depth analyses of learner performance through objective assessment.

Adams (1980) investigated the relationship between the five factors that were identified in assessing the Foreign Service Institute (FSI) Oral Interview Test of Speaking (i.e., accent, comprehension, vocabulary, fluency and grammar) and the global speaking score (e.g., on a scale of 1 to 5) by analysing analytic and overall score data drawn from test performances in various languages.

[1] The study reported in this paper was financially supported by the Australian Research Council under the Discovery Grant Project Scheme, The University of Melbourne's Collaborative Research Grant Scheme, the National Foreign Language Centre at the University of Hawai'i and Kwansei Gakuin University in Japan. I am grateful to the co-researchers of the large-scale study, Lourdes Ortega, Sara Rabie and John Norris, for their work. I also acknowledge with gratitude the support of Dick Schmidt, the Director of the National Foreign Language Center, Alan Davies, Cathie Elder, Tim McNamara, and Mike Long, who helped my co-researchers and me to set up the joint research. I also acknowledge the contribution of Ken Hashimoto, Reiko Komiyama, Jun Matoba, Sally O'Hagan, Steph Power, and Lyn Tang for their tireless work in coding, and other graduate students at the University of Hawai'i and The University of Melbourne for their assistance in data collection and transcription. I am also thankful for my colleagues and students at Melbourne and Queensland including Paul Gruba. I also gratefully acknowledge the assistance of the people who participated in this study.

 Selected Proceedings of the 2008 Second Language Research Forum, ed. Matthew T. Prior et al., 32-47. Somerville, MA: Cascadilla Proceedings Project.

The main discriminating factors were found to be vocabulary and grammar; accent and fluency failed to discriminate at several levels. Higgs and Clifford (1982) suggested that different factors contribute differently to overall language proficiency at the different levels defined in the FSI scale, and proposed the Relative Contribution Model (RCM) to describe rater perceptions of the relative role of each of five component factors making up global proficiency (i.e., vocabulary, grammar, pronunciation, fluency and sociolinguistics.) In their hypothesized model, vocabulary and grammar were considered to be the most important across all levels, but as the level increased, other factors such as pronunciation, fluency and sociolinguistic factors also became important. The hypothesized RCM was then presented to a panel of experienced teachers, whose opinions were elicited on the question of the relative contribution of factors at different levels. The results showed that teachers perceived vocabulary and pronunciation factors to be most important at lower levels and that fluency and grammar factors contributed little; contributions from fluency and grammar increased as the proficiency level went up. At higher proficiency levels, four factors (vocabulary, grammar, pronunciation and fluency) contributed equally, with the sociolinguistic factor contributing relatively less.

De Jong and Van Ginkel (1992) used speaking test data from 25 secondary-school students of French to investigate the relative contributions of different aspects of oral proficiency to the global proficiency score. The results revealed that the pronunciation category contributed most to global proficiency at the lower level, but, as the level went up, fluency became more important. The contribution of accuracy and comprehensibility did not vary across the levels. McNamara (1990), validating the speaking sub-test of the Occupational English Test (OET), a specific purpose test for health professionals, investigated the relationship between the global score (Overall Communicative Effectiveness) and the five analytic scales (Resources of Grammar and Expression, Intelligibility, Appropriateness, Comprehension and Fluency). An analysis using Rasch Item Response Modelling identified Resources of Grammar and Expression as the strongest determinant of the score for Overall Communicative Effectiveness; it was also the most "difficult," that is, the most harshly rated criterion (Comprehension was scored most leniently). On the whole, the studies cited above appear to show that, across levels, grammatical accuracy is the principal determining factor for raters assigning a global score, with variations in the contribution of other factors depending on level.

The studies cited above examined the features of proficiency based on the ratings and feedback from teachers and experts, but the features of proficiency were not identified according to what learners produced. As mentioned above, a number of researchers have undertaken in-depth analyses of test-taker performance and cross-referenced this with the ratings. Some studies focused on one or two traits of oral proficiency (e.g., grammatical accuracy, syntactic complexity etc.) and others investigated a variety of features. For example, Magnan (1988) examined the number of different types of grammatical errors in the transcripts of oral proficiency interviews conducted with 40 students studying French at college level, and then co-referenced this to oral proficiency interview (OPI) ratings. A significant relationship was found between the percentage of grammatical errors and the OPI score, but the relationship was not always linear. According to Magnan, (a) the relationship of errors to proficiency varies considerably depending on the category of error; (b) at higher levels, learners attempt more complex grammatical notions, and consequently make more errors. More recently, Iwashita et al. (2008) investigated the nature of speaking proficiency in English as a second language in the context of a larger project to develop a rating scale. Spoken test performances representing five different tasks and five different proficiency levels (200 performances in total) were analysed using a range of measures of grammatical accuracy and complexity, vocabulary, pronunciation and fluency. The results showed that features from each of these categories helped distinguish overall levels of performance, especially features of vocabulary and fluency. Unlike the two studies above, Larsen-Freeman (2006) focused her investigation on the development of proficiency by analysing the oral and written data of five Chinese learners of English in terms of complexity, accuracy, lexical complexity and fluency over a four-month period. Qualitative and quantitative analyses revealed that the five learners became more fluent, accurate and complex, but that each learner followed a different path in terms of rate of improvement.

The following three studies, in investigating the nature of oral proficiency, also explored the usefulness of the T-unit as a unit of analysis. Halleck (1995) examined the relationship between holistic judgments of oral proficiency and objective measures of syntactic maturity in oral proficiency interviews of 107 students of English as a Foreign Language (EFL) in China. The comparison was

based on three separate tasks. Halleck found learner performance differed according to proficiency level and task type. The syntactic complexity measures used in the study were mean T-unit length, mean error-free T-unit length, and percentage of error-free T-units. Based on L1 research findings (e.g., Witte & Sodowsky, 1978), it was assumed that syntactic complexity increases as L2 learners gain experience in the L2 (Farhady, 1979; Scott & Tucker, 1974). Halleck also discussed the rationale of using T-unit length as a complexity measure and referred to the studies conducted in languages such as French, German, Spanish and Arabic as well as English as a Second Language (ESL). Similarly, Harrington (1986) examined speaking performances of 14 learners of Japanese using T-units to see whether T-unit analysis is a reliable measure of oral proficiency of Japanese as a foreign language (JFL). The results showed that the average T-unit length (the number of words per T-unit) and the average length of error-free T-units (the number of words per error free T-unit) serve in some degree to discriminate among learners. However, significant differences were obtained only after the number of proficiency levels was reduced. Harrington concluded the T-unit measures are of limited usefulness as an index of oral proficiency. Like Larsen-Freeman (2006), Tamaru, Yoshioka, and Kimura (1993) used various T-unit measures (e.g., the number of words per T-unit, the number of clauses per T-unit) to investigate the development of oral proficiency of six JFL learners, and found significant improvement in terms of the length and complexity of learner speech over 18 months. Although the result is promising in terms of the use of T-unit measures and syntactic complexity (measured in terms of the number of words per T-unit, the number of clauses per T-unit, the number of words per error-free T-unit, and the number of error-free clauses per T-unit), the sample size is small and there is no mention of the relationship with independent proficiency measures. Furthermore, it is noted that the T-unit was developed by Hunt (1970) for use with written language development in Primary Schools and for the mother tongue, not foreign languages. Any unexpected findings, therefore could be attributed to the use of the unit for analysis of written language produced by native speakers.

The languages investigated in these studies are predominantly English and other European languages (e.g., French). To date very few studies have investigated languages other than English, with the exception of Harrington (1986) and Tamaru et al. (1993). Furthermore, although various aspects in the four traits of oral proficiency were investigated and some interesting findings produced, the definition of each trait remains unclear. In the following section, each of the four main language traits will be reviewed briefly in order to clarify the understanding of the four main traits.

2.1. Syntactic complexity

Investigation of syntactic complexity in learner language has a long tradition in L2 writing studies. More recently, a growing number of SLA studies have examined syntactic complexity through analysing speech samples (e.g., Iwashita, McNamara, & Elder, 2001; Ortega, 1999; Robinson, 1995; Skehan & Foster, 1999). Common to both writing and speaking studies and in line with interpretations of syntactic complexity, various measures have been employed to investigate syntactic complexity in learner language. For example, Foster and Skehan (1996) referred to syntactic complexity as the elaboration and variation of syntactic patterning. Wolf-Quintero, Inagaki, and Kim (1998) explained that grammatically complex language involved varied and sophisticated structures. For Ortega (2003), "syntactic complexity (syntactic maturity or linguistic complexity) refers to the range of forms that surface in language production and the degree of sophistication of such forms" (p. 492). The measures used to examine syntactic complexity include length of production unit (e.g., T-units, clauses, verb phrases and sentences), amount of embedding, subordination and coordination, range of structural types, and structural sophistication. These measures are usually reported in terms of frequency, ratios and indices.[2]

Reviewing thirty-nine L2 writing studies, Wolf-Quintero et al. (1998) investigated how second language development is evaluated in L2 writing research and which measures are the best indicators of developmental levels in writing. In their review, all grammatical complexity measures (including frequency, ratio and index measures) were summarised and the strength of the relationship between a number of measures and independently-determined proficiency levels in terms of statistical correlation was examined. They identified five measures as the most satisfactory in predicting developmental levels, but they cautioned that those measures do not always discriminate between adjacent levels (e.g.,

[2] Index measure is based on a formula that yields a numerical score (e.g., Falhive & Snow, 1980; Perkins, 1980).

Levels 1 and 2, Levels 5 and 6 etc.) nor are they necessarily strongly correlated with other proficiency measures (e.g., holistic ratings, program levels). Ortega (2003) analysed twenty-five L2 writing studies at college level over twenty-five years, and investigated the impact of instructional setting and the proficiency sampling criterion on the variations in results. The syntactic complexity measures used in the twenty-five studies ranged from length (MLTU: mean length of T-unit) to embedding, while production units also varied from T-units to clauses and sentences.

Many studies investigating syntactic complexity in L2 speech data have also employed the measures used in L2 writing studies (e.g., Crookes, 1989; Ortega, 1999; Skehan & Foster, 1999). However, Foster, Tonkyn, and Wigglesworth (2001), reviewing several approaches to spoken language analysis, pointed to problems in employing measures used to analyse written data in terms of the fragmentary and elliptical characteristics of speaking data. They suggested a new unit, the Analysis of Speech Unit (AS-unit), as a way of providing sufficiently detailed data for analysis. An AS-unit consists of 'an independent clause or sub-clausal unit, together with any subordinate clauses(s) associated with either' (Foster et al., 2001, p. 365). It includes units common in speech, such as an independent sub-clausal unit (e.g., responding 'London to the question 'Where are you from?'). Although the AS-unit offers a solution to the systematic analysis of complex oral language, the process of coding AS units is itself rather complex and so far few studies have used it.

2.2. Lexical diversity

For lexical diversity, like syntactic complexity, the most commonly-used measure is ratio-based (type/token ratio). Despite the widespread use of the measure, a crucial question has been raised as to whether type-token ratio really measures lexical richness (e.g., Malvern & Richards, 2002; Vermeer, 2000). Iwashita et al. (2008), for example, found that even if the number of word tokens and types increased alongside an increase in proficiency level, the *rate* of increase differed. Accordingly, the type-token ratio of lower proficiency learners was found to be higher than that of higher proficiency learners. Vermeer (2000) reviewed a number of lexical measures used in both first- and second-language studies and found none satisfactory for the assessment of lexical diversity. Further complicating matters is the fact that despite a number of attempts by researchers (e.g., indices of function and content words, nouns and verbs by Broader et al., 1993; the use of a vocabulary profile by Laufer & Nation, 1995) no solution has been offered to gauge the development of lexical richness. Malvern and Richards (2002) developed a measure of vocabulary diversity, *D*, based on mathematically modeling how new words were introduced into larger and larger language samples. In their study, they investigated the relationship between *D* and other measures of foreign language proficiency by analysing oral interview data, and found a significant correlation between *D* and other measures.

2.3. Grammatical accuracy

It is generally agreed that accurate speech does not contain errors, or, put differently, it is error-free. However, empirical studies in both SLA and language assessment have reported measures of grammatical accuracy in terms of global accuracy (i.e., identifying any and all types of errors; e.g., Foster & Skehan, 1996; Skehan & Foster, 1999), or specific types of error (e.g., Ortega, 1999; Robinson, 1995; Wigglesworth, 1997). The *global accuracy* approach has the potential to be the most comprehensive because all errors are considered. On the other hand, it is also the most difficult to code consistently. Indeed, studies of global accuracy by Foster and Skehan (1996) and Skehan and Foster (1999) report no inter-coder reliability measures. Related studies (Elder & Iwashita, 2004; Elder, Iwashita, & McNamara 2002; Iwashita et al., 2001) have also found that coders tend not to agree on error types in such work. *Specific types of error* approaches, on the other hand, do not encounter inter-coder reliability problems. Unfortunately, they are also narrower and less inclusive of all potential features related to accuracy found in learner discourse.

2.4. Fluency

The definition of fluency also varies widely. Some researchers focus solely on the temporal features of speech (e.g., words or syllables per minute, the length or number of pauses; Lennon 1990).

Others look exclusively at the automaticity of language use (i.e., to what extent learners are able to produce a second language without attending to rules of the target language grammar; e.g., Schmidt, 1992; Towel, Hawkins, & Bawergui, 1996). Counting production features may be more straightforward than counting instances of automaticity, but distinguishing these features from other key traits is problematic. It is not clear, for example, whether temporal features in speech have much to do with the extent of speech sophistication. Furthermore, assessing fluency across cross-linguistic differences needs to be taken into consideration, as stress-timed languages (e.g., English) differ from syllable-timed languages (e.g., Japanese and Spanish).

2.5. Summary

A considerable number of studies have investigated features of oral proficiency using various methods. The results differ, however, depending on the data type and the methodology. That is, from studies that use data in the form of ratings and feedback on ratings, grammatical accuracy is the principal determining factor for raters assigning a global score, with some variation in the contribution of other factors depending on proficiency level. On the other hand, in studies that conduct in-depth analyses of learner performance, vocabulary and fluency are the principal factors, but, depending on the level, other features come into play. Nevertheless, the studies above showed that different traits contribute to the global proficiency differently according to the proficiency level. A small number of studies have investigated the usefulness of the T-unit, but these were mostly investigated in terms of T-unit length. Other features of syntactic complexity are not explored with the exception of a few studies. Furthermore, the learner performance analysed in the studies above is relatively short (i.e., 3-5 mins at most) and collected under testing condition. Therefore it is not clear whether the limited length of speech and the condition under which learners produced may have some impact on the overall characteristics of oral proficiency. Across the studies, a variety of languages were investigated, but they were mainly European; very few published studies of Japanese and other non-European languages are available to date.

4. The study

4.1. Research questions

The present study addresses the following two research questions.

1. In what ways does performance on narrative tasks differ by proficiency level?
2. What are the distinguishing features in the four traits of oral proficiency at each of two levels?

These two research questions will be examined by analyzing the data of three narrative story retelling task performances by participants of two different proficiency levels who studied EFL or JFL at the time of data collection.

4.2. Participants

Data were drawn from oral performances by 72 learners of EFL (in Japan) and JFL (in the US) at two levels of proficiency[3] (EFL Low N = 20, High N = 19; JFL Low N = 13, High N = 20). Of the EFL group, all were native speakers of Japanese at a university in Japan. The mean age was 19.64 with a *SD* of 1.30. All had studied English for at least six years prior to university and many had spent some time in an English-speaking country (mean 0.43 years with a *SD* of 1.5). They were divided into two proficiency groups (Low and High) according to their score on the institutional version of the TOEFL test. There were 16 males and 23 females.

Of the JFL group, the low proficiency learners had completed three semesters at the time of data collection. The learners in the high proficiency group had completed four semesters. The learners were native speakers of English learning Japanese at tertiary level in the US. There were 17 males and 16 females. The mean age was 23.4 with a *SD* of 5.24. Most of the students had studied Japanese at high

[3] The two proficiency groups in the study were referred as High and Low in this article, but in fact the learner proficiency of both language groups are lower intermediate considering the TOEFL (EFL) and SOPI (JFL) scores and their learning experience.

school (mean 4.84 years with a *SD* of 2.49) and spent some time in Japan (mean 1.23 years with a *SD* of 2.5).

4.3. Proficiency measures

In order to establish a comparison framework across measures, four independent proficiency measures[4] were used together with a bio data survey: Elicited Imitation Task (EI task) and a self-assessment (SA). Also, a Simulated Oral Proficiency Interview (SOPI) was used for the JFL group and an institutional version of TOEFL was given to the EFL group.[5] We used TOEFL for EFL group and SOPI for JFL group among the three proficiency measures. We did this to validate the two proficiency groups in the first place because these two measures are used widely and regarded as a highly reliable measure. EI tasks have been shown to correlate with other external criterion measures, such as oral proficiency ratings (e.g., Bley-Vroman & Chaudron, 1994; Cartier, 1980; Clark, 1980; Hendricks et al., 1980; Henning, 1983; Radloff, 1991). The task consists of 30 sentences varying in length and syntactic complexity. The participants were asked to repeat as much of the sentence as they could after hearing the sentence on tape. For the JFL group, a tape-mediated oral proficiency test was administered using the Simulated Oral Proficiency Interview (SOPI) developed by the Center for Applied Linguistics (CAL) in Washington DC (1992). All SOPI performances were assessed by accredited SOPI raters. The self-assessment instrument used in the study was also developed by CAL. In addition, for the EFL group, as mentioned above, an institutional version of the TOEFL test was administered. The results of these proficiency measures are summarized below. With the exception of the self-assessment scores for the JFL group, all measures were significantly different between the two proficiency levels.

Table 1. *Proficiency Measures (EFL)*

	High *N* = 20		Low *N* = 19		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
EI	54.89	11.37	39.20	15.05	3.66	0.00
SA	59.79	11.46	51.80	10.24	2.30	0.03
Inst TOEFL	492.79	8.33	403.45	11.65	27.42	0.00

Note. EI = Elicitated Imitation Task, SA = Self assessment task; Inst TOEFL = Institutional Version of TOEFL

Table 2. *Proficiency Measures (JFL)*

	High *N* = 20		Low *N* = 13		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
EI	93.15	15.55	71.92	19.54	3.46	0.00
SA	70.25	15.75	63.62	11.43	1.31	0.20
SOPI	1.89	0.57	1.29	0.18	3.68	0.00

Note. EI = Elicited Imitation Task, SA = Self assessment task, SOPI = Simulated Oral Proficiency Interview

4.4. Tasks

Three oral narrative story-telling tasks (monologic tasks) typically used in oral proficiency interviews, language classrooms in SLA and language assessment research were chosen to elicit learners' speech. In the first task, learners were asked to retell the story after viewing Chaplin's short silent film *Alone and Hungry*. In the second task, learners listened to a story in their L1 (for the EFL

[4] Four measures in total (i.e., Elicited Imitation task, a self-assessment, Simulated Oral Proficiency Interview and an institutional version of TOEFL) were used in the study, but all participants in each language group took three measures.

[5] The version of the TOEFL all EFL learners took in the present study is the same.

group Japanese and for the JFL group English) on tape and recounted the story in L2 using the accompanying pictures. The last task was divided into three sections. In each section, students told a story from pictures. These tasks had all been used in other studies and had proved satisfactory in eliciting a range of language from learners (e.g., Ortega, 1999). Learners were given approximately fifteen minutes to tell each story. Three minutes of planning time for the first task and two minutes for the second and third tasks were given in order to ensure that learners produced at their maximum level of complexity. All speech samples were collected in a language laboratory, audio-recorded and transcribed using guidelines developed for the larger study (Iwashita et al., in progress).

4.5. Analyses

The features analysed in the present study were grouped according to the four main traits of oral proficiency mentioned above: syntactic complexity, grammatical accuracy, lexical diversity, and fluency.

All 72 speech samples were transcribed using transcription guidelines described in a study by Iwashita et al. (in progress). For the analyses, the segmented and coded speech samples were entered into a database for use with the CLAN (Computerized Language Analysis) program developed as part of the *CHILDES* project (MacWhinney, 1999). The CLAN program allows a large number of automatic analyses to be performed on data, including frequency counts, word searches, co-occurrence analyses, and calculation of type/token ratios. The measures of each of the four language traits are summarized below. In this paper, the results of the data analyses of the selected features of the four traits are reported.

4.5.1. Syntactic complexity

The transcribed data were coded according to the various syntactic complexity measures used in the empirical studies. The measures used in the study are summarised in Table 3.

Table 3. *Production Units and Syntactic Complexity Measures Used in the Present Study*

Production unit	T-unit, clause, word
Syntactic complexity category	Syntactic complexity measure
Length	Number of words per T-unit, number of words per clause
General complexity measure	Number of clauses per T-unit
Subordination measure	Number of DCs per T-unit

Note. DC = Dependent clause

Production units employed in this study were T-unit, clause, and word. In coding T-units, features of spoken languages mentioned by Foster et al. (2001) (e.g., independent sub-clausal unit, minor utterances) and non-target-like features were included. Following Hunt (1970) and Polio (1997), a T-unit is an independent clause and all its dependent clauses. While Hunt (1970) confirmed the usefulness of the T-unit for widely-spaced school levels, Ishikawa's study (1995) revealed that the clause is more helpful than the T-unit for analysing data of beginner learners as it provides a smaller context. As this study included learners beyond beginner level, it was decided to use both T-units and clauses.

There are four types of complexity measures: general complexity, subordination and length (both T-unit and clause). The 'general complexity' measure (the number of clauses per T-unit) considers the depth of a T-unit. There is an assumption that the more clauses per unit, the more complex the production. Subordination concerns the degree of embedding. It is assumed that more advanced learners will produce more dependent clauses per T-unit and that the proportion of dependent clauses to the total number of clauses will increase. The length of clauses and T-units was determined by calculating the number of words in each production. As shown in past studies, it was assumed that learners combine short, simple sentences into longer and complex sentences as their language develops. The examples in each language are given in the Appendix. All examples are drawn from the data in the present study.

4.5.2. Grammatical accuracy

As mentioned earlier, empirical studies in both language assessment and SLA have reported measures of the grammatical accuracy of learners' speech either in terms of *global accuracy* (i.e., identifying any and all types of errors; e.g., Foster & Skehan, 1996; Skehan & Foster, 1999) or in terms of *specific types of errors* (e.g., Ortega, 1999; Robinson, 1995; Wigglesworth, 1997). In the present study global accuracy was examined by calculating error-free T-units as a percentage of the total number of T-units. Error-free T-units are T-units free from any grammatical errors including both the specific errors defined above as well as other grammatical errors (e.g., word-order, omission of pronouns etc.)

4.5.3. Lexical diversity

Lexical diversity was examined with the measure *D* developed by Malvern and Richards (2002) using their *VocD* analysis program in CHILDES program.

4.5.4. Fluency

The following features were identified as suitable measures of fluency: *unfilled pauses*, *repairs*, *repeats*, and *speech rate*. The number of *unfilled pauses* was calculated by counting the number of pauses of one second or more that occurred in the speech (Mehnert, 1998). In order to enable comparisons, instances of *unfilled pauses* and *repair/repeat* were counted per 60 seconds of speech because the actual speaking time of individual learners varied (as a function of the amount of pause time and filled pauses). *Repair* refers to the repetition of exact words, syllables or phrases; replacement; reformulations (grammatical correction of structural features); false starts; and partial repetition of a word or utterance (Freed, 2000). *Speech rate* was calculated by dividing the total number of syllables produced in a given speech sample by the total time expressed in seconds (Ortega, 1999). First, the transcribed speech was pruned by excluding features of repair; then the resulting total number of syllables was divided by the total speech time, excluding pauses of three or more seconds.

4.6. Inter-coder reliability

The data were coded by two coders. Inter-rater reliability was calculated for the segmentation of T-units and clauses and the identification of error-free T-units by the percentage of agreement between the coders. A high level of agreement (greater than 95%) was reached for all measures.

5. Results

The results of the analyses of the four proficiency traits are reported separately in each language group below. Because the proficiency of participants in the two language groups are not quite the same, and therefore direct comparison of the results was not undertaken.

5.1. EFL data
5.1.1. Syntactic complexity

In this section, the frequency of production units in the two proficiency groups will also be reported as well as the results of syntactic complexity measures. The two proficiency levels were compared for the frequency of production units (i.e., T-unit, clause, and word token), the length of the T-units and clauses, and the percentages of clauses per T-unit (general complexity measure) and of dependent clauses per clause (subordination measure). The descriptive statistics and the t-test statistics are summarised in Table 4. As is shown in Table 1, although the higher proficiency learners received significantly higher scores for all three proficiency measures (i.e., Elicited Imitation Task, institutional

version of TOEFL, and self assessment), as shown in Table 4 below, participants in the lower proficiency group were more productive than the higher proficiency group in terms of the number of words, T-units and clauses. These differences, however, were not statistically significant.

Descriptive statistics of the four complexity measures (i.e., general complexity measure, subordination measure and length of T-unit and clause) showed little difference between the two proficiency groups. The mean length of clause was longer in the lower proficiency group than in the higher proficiency group. Nevertheless, the difference was not significant. It should be noted that standard variations of some measures are quite large, which indicates a large individual variation.

Table 4. *Production Units and Syntactic Complexity Measures (EFL)*

	High $N = 20$		Low $N = 19$			
	M	*SD*	*M*	*SD*	*t*	*p*
Production units						
Words	590.11	179.54	609.35	217.86	-0.30	ns
T-units	71.16	16.75	76.95	22.70	-0.90	ns
Clauses (Cl)	92.32	26.89	94.35	31.50	-0.22	ns
General complexity measure:						
Number of Cl per T-unit	1.29	0.17	1.22	0.14	1.55	ns
Subordination: No. of DCs per Cl	0.19	0.07	0.19	0.05	0.14	ns
Length						
Number of words per T-unit	8.28	1.27	7.6	1.5	1.71	ns
Number of words per Cl	4.09	2.46	5.39	1.67	-1.90	ns

Note. Cl = clause, DC = dependent clause

5.1.2. Grammatical accuracy

The descriptive statistics of the percentage of error-free T-units are summarized in Table 5. The percentage of error-free T-units produced was just above 30 percent for the higher proficiency learner group and under 30 percent for the lower proficiency learner group. The difference was found to be significant.

Table 5. *Grammatical Accuracy (EFL)*

	High $N = 20$		Low $N = 13$		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
% of EFT	0.34	0.08	0.27	0.08	2.52	0.02

Note. EFT = error free T-unit

5.1.3. Lexical diversity

Table 6 shows that lexical density (percentage of content words in the total number of word tokens) was found to be not significantly different between the two groups, but the other measure shows the effect of the proficiency level. That is, the high proficiency group participants produced a significantly greater variety of words than the low proficiency group participants. A large individual variation in *D* measure was observed in the high proficiency group.

Table 6. *Lexical Diversity (EFL)*

	High $N = 20$		Low $N = 19$		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
Lexical density	0.46	0.03	0.47	0.03	-0.83	ns
D	56.20	15.01	47.29	9.24	2.24	0.03

5.1.4. Fluency

The results for speech rate showed a clear relationship with proficiency level. That is, the higher level EFL learners spoke significantly faster than the lower level participants, but more frequent instances of pause and repeat/repair were observed in the higher level learners than the lower level learners. The difference in the frequency of pauses was significant.

Table 7. *Fluency (EFL)*

	High *N* = 20		Low *N* = 19		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
Unfilled pauses per 60 sec.	26.37	2.77	22.75	2.87	3.90	0.00
Speech rate	1.20	0.33	0.89	0.34	2.72	0.01
Number of repeats & repairs per 60 sec.	4.29	1.64	3.97	2.14	0.51	ns

5.1.5. Summary

Overall, grammatical accuracy and some features of lexical diversity and fluency provided evidence that features of task performance varied according to proficiency level, but no measures of syntactic complexity showed any proficiency effect. Unexpected findings were that lower proficiency learners produced more speech and significantly fewer unfilled pauses than higher proficiency learners.

5.2. JFL data

5.2.1. Syntactic complexity

First of all, as in the previous section, the frequency of production units in the two proficiency groups will also be reported as well as the results of syntactic complexity measures. As is shown in Table 8, the amount of speech measured by three production units (i.e., words, T-units and clauses) was significantly different between the high and low proficiency groups for word tokens and number of clauses, but not number of T-units. This means that high proficiency learners produced a significantly larger number of smaller units and words, but when the speech samples were measured with a larger unit (T-unit) the difference was not significant.

Expected direction was observed in all complexity measures, but a statistical difference between the two groups was observed for the three measures. Subordination measured by the number of dependent clauses per clause was not found to be significant between the two groups. As observed in the results of the EFL group, the individual variations in all measures were found to be quite large, as is shown by the large standard deviation.

Table 8. *Production Units and Syntactic Complexity Measures (JFL)*

	High *N* = 20		Low *N* = 13		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
Production units						
Words	822.35	295.33	580.23	188.97	2.62	0.01
T-units	62.25	23.69	50.92	16.34	1.5	ns
Clauses (Cl)	102.65	37.13	72.85	25.05	2.54	0.02
General complexity measure:						
Number of Cl per T-unit	1.72	0.26	1.43	0.18	3.41	0.001
Subordination:						
Number of DCs per Cl	0.28	0.10	0.22	0.08	1.79	ns
Length						
Number of words per T-unit	12.06	2.21	9.47	1.44	3.74	0.001
Number of words per Cl	7.00	0.42	6.63	0.62	2.06	0.05

Note. Cl = clause, DC = dependent clause

5.2.2. Grammatical accuracy

As summarized in Table 9, the mean percentage of error-free T-units was 50 percent for the higher proficiency group and just under 50 percent for lower proficiency group, but the difference was not found to be significant.

Table 9. *Grammatical Accuracy (JFL)*

	High *N* = 20		Low *N* = 13		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
% of EFT	0.50	0.14	0.45	0.13	1.13	ns

Note. EFT = error free T-unit

5.2.3. Lexical diversity

The same pattern as for the EFL group was observed in lexical diversity. That is, while there was no significant difference in lexical density measured by the percentage of content words in the total word tokens between the two groups, a significantly greater variety of words was produced by the participants in the higher proficiency group than the lower proficiency group. It should be noted that individual variations in the higher proficiency group for lexical variety are quite large.

Table 10. *Lexical Diversity (JFL)*

	High *N* = 20		Low *N* = 13		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
Lexical density	0.52	0.04	0.50	0.02	1.06	ns
D	42.84	10.10	35.77	5.63	2.30	0.03

5.2.4. Fluency

Like the EFL learners, the higher proficiency JFL learners spoke significantly faster than the lower proficiency learners, but the frequency of pauses was not different between the groups. The frequency of other features of fluency repair was significantly different; repairs/repeats occurred much more often among the lower proficiency learners than the higher proficiency learners. The results are summarized below.

Table 11. *Fluency (JFL)*

	High *N* = 20		Low *N* = 13		*t*-test	
	M	*SD*	*M*	*SD*	*t*	*p*
Number of unfilled pauses per 60 sec.	29.02	8.61	33.64	5.48	-1.72	ns
Speech rate	2.19	0.72	1.70	0.46	2.16	0.04
Number of repeats & repairs per 60 sec.	2.06	1.20	3.78	2.05	-3.04	0.00

5.2.5. Summary of JFL data

Overall, most features of oral proficiency showed a proficiency effect, with the exception of a few features. That is, the higher proficiency learners produced significantly faster and more complex speech with a significantly wider variety of words, but their speech was not more accurate than that of the lower proficiency learners, and the frequency of their pauses was not different from that of the lower proficiency learners.

5.3. Overall summary

An overall summary of the results of the statistical analyses is given in Table 12 below. Considerable differences in the two proficiency groups between EFL and JFL learners were observed. That is, while the proficiency effect was observed in the frequency of production units and syntactic complexity measures in the JFL group, no difference was observed in any of the measures in the EFL data. The opposite was the case in the grammatical accuracy analysis. Higher proficiency learners in the EFL group produced significantly more accurate speech than did lower proficiency learners, but the difference in the JFL group was not significant. The same result was observed in lexical diversity: the proficiency effect was observed in the *D* measure, but not in lexical density. As for fluency, the higher proficiency learners in both the EFL and JFL groups spoke significantly faster than did the lower proficiency learners, but while the proficiency effect in the frequency of unfilled pauses was observed in the EFL group (in reverse direction), the proficiency effect in the frequency of repairs and repeats was noted in the JFL group. It should be noted that the large *SD*s show large individual variations across the measures.

Table 12. *Summary of the Results*

	EFL	JFL
Production units		
Words		✓
T-units		
Clauses		✓
Complexity		
Number of clauses per T-unit		✓
Number of DCs per clause		
Number of words per T-unit		✓
Number of words per clause		✓
Accuracy	✓	
Lexical diversity		
Density		
D	✓	✓
Fluency		
Number of unfilled pauses per 60 sec.	✓	
Speech rate	✓	✓
Number of repeats/repairs per 60 sec.		✓

Note. DC = dependent clause

6. Discussion

In the present study, various features in the four traits of oral proficiency in English and Japanese as a foreign language were examined via in-depth analyses of narrative performances in order to see in what ways task performance differed by level and what features distinguished levels more clearly than did others. As is summarised in Table 12 above, the results of the data analysis of the two language groups differed markedly, except for one feature of fluency (speech rate) and one of lexical diversity (measured by *D*). These findings of the study will be discussed in light of the two research questions, for each language.

6.1. EFL data

First, the proficiency effect was observed in the features of grammatical accuracy, lexical diversity measured by the formula *D*, and one aspect of fluency (i.e., speech rate). These results are similar to the findings of a large-scale study by Iwashita et al. (2008). As noted by Iwashita et al. (2008), more macro-level categories, such as speech rate, the main vocabulary measures, the global grammatical

accuracy measure, appear to have had the most influence on the difference in performance of the two proficiency groups. An unexpected finding was that the frequency of all of the three production units (i.e., words, clauses and T-units) was larger for the higher than for the lower proficiency learners, though the difference was not significant. Also, the higher proficiency learners produced significantly more pauses than did the lower proficiency learners.

These unexpected findings may be explained in terms of the proficiency of the EFL learners in the present study and some methodological issues. As shown in Table 1, the differences in the three proficiency measures between the two proficiency groups are highly significant, but the differences reported in the test scores were not as obvious for smaller aspects of the features of oral proficiency under study. As mentioned in the methodology section, all EFL learners in the current study were homogenous in terms of length of study, in-country experience and age. All had studied English for six years at secondary school and were in the first or second of year of university. The TOEFL scores ranged between the low and high 400s (Table 1). Considering the maximum score of TOEFL (i.e., 677), all participants in the two EFL groups were at the lower end of proficiency. As Iwashita et al. found no differences in some features of proficiency between the adjoining levels, the participants in the current study might be placed in the adjoining level if there are more than two levels. If this is the case, it is plausible to conclude that differences at adjacent levels are not always distinguished by measures of the features under investigation; also, that features of syntactical complexity show little difference among learners at the lower end of proficiency.

As explained in the methodology section, in coding all transcribed data with T-units and clauses, the guidelines we developed for analysis considered the features of spoken discourse and learner language which contains similar features identified in the A-S unit (Foster et al., 2000). Nonetheless, it may be worthwhile to code the data with A-S unit to see whether different results might be produced. Lastly, unusual and unexpected findings of the fluency analysis might be limited features of fluency we were able to analyze due to the monologic tasks used for the study. Features observed in interactive tasks performance such as turn taking could not be examined.

6.2. JFL data

Compared with the EFL data, the results of the analyses of the JFL data were generally as expected. As shown in Table 12, most features were found to be significantly different between the high and low proficiency groups. Non-significant differences were observed in some features (e.g., Number of T-units, subordination, accuracy, lexical density and the number of unfilled pauses).

The different findings from the EFL data may be attributable to the difference in proficiency of the two language groups. As explained above, the proficiency of the EFL group was relatively low based on the scores of the institutional version of the TOEFL test. Comparing the self-assessment and elicited imitation task scores of the two language groups (see Tables 1 and 2), the scores of the JFL group are considerably higher than those of EFL group. Also, the mean SOPI scores of the JFL learners (1.89 for the high proficiency group and 1.29 for the low proficiency group) show that their proficiency is in the range from Low Intermediate to High Intermediate High according to the ACTFL proficiency rating.[6] These differences explain the different results in the two language groups, especially in the complexity features.

7. Conclusion

Overall, this study advances our understanding of the features of oral proficiency in learner production. The findings confirm earlier studies that the principal feature that determines proficiency is vocabulary range (measured using *D* in the current study), while other features that characterize oral proficiency vary according to proficiency level. The study needs to be extended to a broader range of proficiency levels of learners and to learners engaged in a wider range of speaking tasks, particularly interactive ones; and a wider range of discourse features needs to be considered. The proficiency range

[6] The SOPI follows the general structure of the oral proficiency interview (OPI) used by the American Council on the Teaching of Foreign Languages (ACTFL) and SOPI scores can be cross-referenced with the ACTFL ratings (Dandonoli & Henning, 1990).

of the participants in the present study was narrow (i.e., two levels in each language) and the task types were specific (i.e., narrative retelling). Future investigations of a range of discourse features with learners of different proficiency levels using a wider variety of speaking tasks are required.

The current study focused on quantitative analyses using the same methods of analysis in order to compare the findings of the data of the two language groups. Use of the four independent proficiency measures was especially helpful in explaining the different findings between the two languages. However, the specific features of each language were largely ignored in the study. For example, subordination and length may not be the only features of complexity in the languages studied. Qualitative analyses of the language data may reveal other features that determine oral proficiency.

Appendix

Examples of T-units and clauses

1. EFL data

T-unit

- And they it come to it comes to them **and** it sits at the foot of her. (2 T-units)
- They have a children, **but** her mother and father the woman's parents want live want to live with with them. (2 T-units)
- The woman stand up **and** run away from the street. (1 T-unit)

Clause

Independent clause [IC]

- Cat is still crying [IC] and it looks like very lonely [IC] and want to have something to eat [IC] (three independent clauses)

Dependent clause [DC]

- He wished [IC] for him to go away [DC]
- Thus he thinks [IC] in the future he and she get married [DC]

2. JFL data

T-unit

- saisho wa sono **futari no otoko no ko** wa hazukashigatteta **kedo** sugu hokano kodomotachi to tokekon de it ta. (two T-units)
  *At first **these two boys** were shy, but soon **(they)** mixed with other children well.*

Independent clause [IC]

- Soshitara minna kodomotachi ga ie na niwa ni de te [IC] booru o mi ni ikimashita [IC] (# 201).
  Then all children got out of the house to the garden and went to see the ball.

Dependent clause [DC]

- {onnanohito wa totemo totemo onaka suiteita **node** [DC]} panya kara pan o nusun de shimaimashita . (# 201)
  As a woman was very hungry, she stole bread from the bakery.

- {demo tatoe ookina kazoku de konna semai ie de sundeite ironna mondai ya yana koto ga okot**temo** [DC]} {kitto kore ga hontoo no shiawase da [DC]} to omoimasu . (#201)
  Even if the size of the family is large and they live in a big house, I think this is real happiness for them.

References

Adams, Marianne L. (1980). Five co-occurring factors in speaking proficiency. In James R. Firth (Ed.), *Measuring spoken language proficiency* (pp. 1–6). Washington, DC: Georgetown University Press.

Bley-Vroman, Robert & Chaudron, Craig (1994). Elicited imitation as a measure of second-language competence. In Elaine E. Tarone, Susan M. Gass, & Andrew D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245–261). Hillsdale, NJ: Lawrence Erlbaum.

Cartier, Francis A. (1980). Alternative methods of oral proficiency assessment. In James R. Frith (Ed.), *Measuring spoken language proficiency* (pp. 7–14). Washington, DC: Georgetown University Press.

Center for Applied Linguistics (1992). *Japanese speaking test*. Washington, DC.

Center for Applied Linguistics (n.d.). *Self assessment questionnaire*. Washington, DC.

Clark, John L. D. (1980). Toward a common measure of speaking proficiency. In James R. Firth (Ed.), *Measuring spoken language proficiency* (pp. 15–25). Washington, DC: Georgetown University Press.

Crookes, Graham (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, *11*, 367–383.

Dandonoli, Patricia & Henning, Grant (1990). An investigation of the construct validity of the ACTFL proficiency guidelines and oral interview procedure. *Foreign Language Annals*, *23*, 11–22.

de Jong, John H. A. L. & van Ginkel, Lieneke W. (1992). Dimensions in oral foreign language proficiency. In John H. A. L. de Jong (Ed.), *The construct of language proficiency* (pp.187–205). Philadelphia: John Benjamin.

Elder, Cathreine & Iwashita, Noriko (2005) Planning for test performance: Does it make a difference? In Rod Ellis (Ed.), *Planning and task performance in a second language* (pp. 219–238). Amsterdam: John Benjamins.

Ellis, Rod (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition, 9*(1), 1–20.

Farhady, Hossein (1979). The disjunctive fallacy between discrete-point and integrative tests. *TESOL Quarterly, 13*, 347–357.

Foster, Pauline & Skehan, Peter (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition, 18*(3), 299–323.

Foster, Pauline, Tonkyn, Alan, & Wigglesworth, Gillian (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354–374.

Freed, Barbara (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In Heidi Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). Ann Arbor: The University of Michigan Press.

Galloway, Vicki B. (1987). From defining to developing proficiency. A new look at the decision. In Heidi Byrnes (Ed.), *Defining and developing proficiency: Guidelines, implementations, and concepts* (pp. 25–73). Lincolnwood, IL: National Textbook Company.

Halleck, Gene B. (1995). Assessing oral proficiency: A comparison of holistic and objective measures. *The Modern Language Journal*, *79*(2), 223–234.

Harrington, Michael (1986). The T-unit as a measure of JSL oral proficiency. *Descriptive and Applied Linguistics 19*, 49–56.

Hendricks, Debby, Scholz, George, Spurling, Randon, Johnson, Marianne, & Vandenburg, Lela (1980). Oral proficiency testing in an intensive English language program. In John W. Oller & Kyle Perkins (Eds.), *Research in language testing* (pp. 77–90). Rowley, MA: Newbury House.

Henning, Grant (1983). Oral proficiency testing: Comparative validities of interview, imitation and completion methods. *Language Learning, 33*, 315–332.

Higgs, Theodore V. & Clifford, Ray (1982). The push towards communication. In Theodore V. Higgs (Ed.), *Curriculum, competence, and the foreign language teacher* (pp. 57–79). Lincolnwood, IL: National Textbook Company.

Hunt, Kellogg W. (1970). Syntactic maturity in school children and adults. *Monographs of the Society for Research in Child Development, 35*(1), Serial No. 134.

Ishikawa, Sandra (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*(1), 51–69.

Iwashita, Noriko, Brown, Annie, McNamara, Tim, & O'Hagan, Sally (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24–49.

Iwashita, Noriko, McNamara, Tim, & Elder, Catherine (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning, 21*(3), 401–436.

Iwashita, Noriko, Ortega, Lourdes, Rabie, Sara, & Norris, John M. (in progress). *Syntactic complexity and oral proficiency in crosslinguistic perspective* Honolulu, Hawai'i: University of Hawai'i, National Language Resource Center.

Larsen –Freeman, Diane (2006). The emergence of complexity, fluency and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590–619.

Laufer, Batia & Nation, Paul (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–322.

Lennon, Paul (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387–417.

MacWhinney, Brian (1999). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

McNamara, Tim (1990). Item response theory and the validation of an ESP test for health professionals. *Language Testing, 7*(1), 52–75.

McNamara, Tim (1996). *Measuring second language performance*. London: Longman.

Magnan, Sally, Sieloff (1988). Grammar and the ACTFL oral proficiency interview: Discussion and data. *The Modern Language Journal, 72*(3), 266–276.

Malvern, David D. & Richards, Brian J. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*, 85–104.

Mehnert, Uta (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*(1), 83–106.

Norris, John M. & Ortega, Lourdes (2003). Defining and measuring SLA. In Catherine Doughty & Michael H. Long (Eds.), *Handbook of second language acquisition* (pp. 716–761). London: Blackwell.

Ortega, Lourdes (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21*(1), 109–148.

Ortega, Lourdes (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 4*(24), 492–518.

Ortega, Lourdes, Iwashita, Noriko, Norris, John, M., & Rabie, Sara (October, 2002). *An investigation of elicited imitation tasks in cross-linguistic SLA research.* Paper presented at the Second Language Research Forum at Courtyard Marriott Downtown, Toronto Canada.

Perkins, Kyle (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly, 14*(1), 61–69.

Polio, Charlene (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47*(1), 101–143.

Radloff, Carla F. (1991). *Sentence repetition testing for studies of community bilingualism*. Arlington, TX: Summer Institute of Linguistics and the University of Texas at Arlington.

Robinson, Peter (1995). Task complexity and second language narrative discourse. *Language Learning*, *45*(1), 99–140.

Scott, Margaret, Sue & Tucker, Richard G. (1974). Error analysis and English-language strategies of Arab students. *Language Learning, 24*(1), 69–97.

Schmidt, Richard (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357–385.

Skehan, Peter & Foster, Pauline (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*(1), 93–120.

Tamaru, Yoshiko, Yoshioka, Kaoru, & Kimura, Shizuko (1993). A longitudinal development of sentence structures: A study of JSL adult learners. *Nihongo kyoiku*: *Journal of Japanese Language Teaching, 81*, 43–54.

Towell, Richard, Hawkins, Roger, & Bazergui, Nives (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84–119.

Vermeer, Anne (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65–83.

Wigglesworth, Gillian (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*(1), 85–106.

Witte, Stephen P. & Sodowsky, Roland E. (1978). Syntactic maturity in the writing of college freshmen. Paper delivered at the Conference on College Composition and Communication. Denver, Colorado (ERIC Document No. ED 163 460).

Wolfe-Quintero, Kate, Inagaki, Shunji, & Kim, Hae Young (1998). *Second language development in writing: measures of fluency, accuracy and complexity* (Technical report #17). Honolulu, HI: Second Language Teaching & Curriculum Center, The University of Hawai‘i.

Selected Proceedings of the 2008 Second Language Research Forum: Exploring SLA Perspectives, Positions, and Practices

edited by Matthew T. Prior, Yukiko Watanabe, and Sang-Ki Lee

Cascadilla Proceedings Project Somerville, MA 2010

ISBN 978-1-57473-439-3 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Iwashita, Noriko. 2010. Features of Oral Proficiency in Task Performance by EFL and JFL Learners. In *Selected Proceedings of the 2008 Second Language Research Forum*, ed. Matthew T. Prior et al., 32-47. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2383.