

Relationships between L2 Speakers' Development and Raters' Perception on Fluency in Group Oral Interaction

Junko Negishi
Waseda University

Negishi, J. (2012). Relationships between L2 speakers' development and raters' perception on fluency in group oral interaction. *Journal of Pan-Pacific Association of Applied Linguistics*, 15(2), 1-26.

This study aimed to investigate the characteristics and developmental phenomena apparent in Japanese learners of English with their raters' perspectives in terms of fluency which has been recognized as a major factor in judging non-native speakers' proficiency. The following temporal indices demonstrated strong relationships with the raters' scores on fluency: total number of syllables including/excluding dysfluency, total number of words, and total speaking time including pause time. The three indices can be useful when confined to Japanese students. With regard to pauses and hesitations, the university students exhibited results comparable to other participants; however, the locations and lengths of pauses at a phrase boundary alone gave evidence that they paused at grammatical junctures.

Key Words: speaking, fluency, group oral interaction, CEFR

1 Introduction

1.1 English ability

It is crucial to start to learn oral communication skills as soon as one begins learning English, namely, at junior high schools, when Japanese students begin learning English as a foreign language. Davies (1978) mentioned that a communicative approach should focus on oral skills before those of written. Canale and Swain (1980a, 1980b), Bachman (1990), and Bachman and Palmer (1996) brought various expanded notions of communicative competence and (communicative) language ability, which subsequently contributed to the Course of Study in Japan. The Ministry of Education, Culture, Sports, Science and Technology (MEXT, 2002) states as follows:

With the progress of globalization in the economy and in society, it is essential that our children acquire communication skills in English, which has become a common international language, in order for living in the 21st century..... we have formulated a

Junko Negishi

strategy to cultivate “Japanese with English abilities” in a concrete action plan with the aim of drastically improving the English education of Japanese people.

The author believes that English teachers in Japan should give their students more opportunities to exchange opinions freely in English to help realize this action. Investigating the characteristics and developmental phenomena apparent in Japanese learners of English with their raters’ perspectives will help us English teachers to realize the importance of and to naturalize speaking and interactional activities in classrooms. Therefore, the author investigated the speakers’ discourses in three types of educational institutions quantitatively and qualitatively in terms of their fluency. Fluency has been recognized as a major factor in judging non-native speakers’ proficiency (Riggenbach, 1991; Schmidt, 2000). In a second language (L2) classroom, accuracy has also been regarded as an important oral ability. In Japan, accuracy has long been considered more important; likewise, until recently, English curricula underestimated the importance of fluency. However, fluency has been gaining importance in Japan’s English education field in line with the globalization.

1.2 Group oral interaction

Paired and group oral test formats have recently been introduced to the range of oral performance tests because the assessment of L2 learners’ authentic conversational competence is considered important in the current era of globalization. Oral performance tests of the paired or small group (“oral interaction in a small group” will be termed “group oral” hereafter, following Bonk and Ockey [2003]) types are being administered, for example, in Cambridge First Certificate (paired), Cambridge Certificate of Proficiency in English (paired) and in the speaking test administered by the Council of Europe (paired and group oral). There are some local tests that utilize the group oral in Asia, as well. In comparison to interviews, the group oral is likely to produce natural and insightful conversation with peers, and it has been reported to be appropriate in certain test situations and in a battery of oral tests (Van Moere, 2006; Fulcher, 1996; Bonk and Ockey, 2003). Research dealing with the paired format has recently begun but only a few studies dealing with the group oral have been carried out as of today.

2 Fluency

What is fluency? Schmidt (2000) and Chambers (1997) assert the ambiguity of the term and the difficulty of specifying linguistic definitions of fluency. Many researchers have attempted to define fluency, and diverse definitions

L2 Speakers' Development and Raters' Perception on Fluency

do exist; nonetheless, in the words of Koponen and Riggensbach (2000), "[T]here can ultimately be no single all-purpose definition of fluency." They assert that fluency in language assessment is comparable to "continuity", "smoothness", or "evenness" of speech, without extreme breaks or hesitations (p. 8). Such aspects of fluency have been mentioned by other researchers, as well. For example, Ellis and Barkhuizen (2005) regard "the production of language in real time without undue pausing of hesitation" (p. 139) as fluency. According to previous studies, temporal variables, pausing, and hesitation can be regarded as indicators of fluency.

The principal temporal variable, speech rate, is normally measured by the number of syllables produced per minute or per second. Speech rate has been reported to be one of the best predictors of fluency, distinguishing non-native speakers from native speakers (Wiese, 1984). The second common temporal variable, mean length of runs, tells us how long, on average, a speaker can speak without pausing. Length of run may be influenced by the extent to which L2 speakers can access "ready-made chunks of language" (Ellis & Barkhuizen, 2005, p. 156) or automaticity.

In terms of pause-related studies, pause length in L2 speech has been demonstrated to be longer than that of L1, and it can be a key marker of fluency (Lennon, 1984; Mohle, 1984). The lower limit of pause length, the cut-off point, has been a controversial issue among scholars. Different researchers use different cut-off points, ranging from 0.1 seconds to 0.3 seconds; in other words, there is no consensus as to what constitutes a silent pause. As Towell, Hawkins, and Bazergui (1996) argue, if this cut-off point is too low, the pause may include plosive phase or voiceless stop phase. On the other hand, if it is too high, some hesitation time may be excluded from the analysis. With regard to the number of pauses, some studies concluded that the number of silent and filled pauses determines speakers' fluency (cf. Riggensbach, 1991; Freed, 1995, 2000), while van Gelderen's (1994) study did not find a relationship between the frequency of silent and filled pauses and fluency in Dutch students ages 11 and 12. Fluent speakers, who spend less time pausing, usually pause at clause boundaries or grammatical junctures between nonessential parts of a clause, while non-fluent speakers often pause within clauses (Freed, 1995, 2000; Riggensbach, 1991; Towell et al., 1996; Segalowitz & Freed, 2004). Chambers (1997) outlines features of temporal variables as follows:

Speech rate alone cannot be what contributes to the feeling that, as a listener, we are interacting with a foreigner. What appears significant from research in this area is:

- (1) the frequency of pauses rather than the length,
- (2) the length of run,
- (3) the placement of pauses in an utterance,
- (4) the transfer (or not) of pausing pattern from L1 to L2. (p. 541)

Junko Negishi

With respect to hesitation phenomena such as repetitions, false starts, and repairs, research findings have not reached a conclusion. For example, Riggenbach (1991) states that in native speakers' speech samples, a great deal of hesitation and repairs also occurs.

Circumlocution and paraphrase can be classified as communication strategies, in addition to temporal and hesitation phenomena. Canale and Swain (1980a, 1980b) assert that these strategies are utilized to compensate for breakdowns in communication; that is, learners apply circumlocution and paraphrase when they do not know an exact word or expression. It has been reported that the utilization of communication strategies decreases as learners' proficiency increases (Yoshida-Morise, 1998). Chen's (1990) study tells us that linguistic-based communication strategies (e.g., circumlocution) used by high-proficiency L2 speakers were more effective than the knowledge-based strategies (e.g., repetition) produced by low-proficiency speakers. Conversely, this may mean that lower-level speakers can use knowledge-based strategies but not linguistic-based strategies. If a speaker is capable of utilizing circumlocution or paraphrase to describe something concrete or abstract, the listener may feel that the speaker has a certain level of speaking ability.

3 Research Questions

As mentioned earlier, the group oral is still a new type of test format and a limited number of research has been carried out. The studies to date are mainly on interlocutor effects or differences apparent in interviews. Investigation of raters' behaviors/characteristics and rating criteria with respect to the group oral has not yet been conducted; nor the relationship between raters' ratings and learners' speaking developments. In response to such situation, this study was carried out by choosing 135 students from three types of educational institutions as representatives of English learners assessed during their group orals.

In order to accomplish the purpose of the study, the following questions were set for when Japanese learners of English took the group oral interactions:

- 1) What are the characteristics and developmental phenomena of the participants when analyzing their speaking samples based on their fluency?
- 2) What are the relationships of the participants' characteristics, the developmental phenomena in terms of fluency and the raters' scores?

L2 Speakers' Development and Raters' Perception on Fluency

4 Method

4.1 Participants

The participants in the study were 135 students from seven schools from among three kinds of educational institutions, that is, two junior high schools, two senior high schools, and three universities in and around Tokyo, Japan. They were divided into a total of 45 groups, each containing three students. The groups comprised of fifteen junior high school student groups, fifteen senior high school student groups, and fifteen university student groups. In order to ensure the appropriate balance of students from the various types of schools, about half of the participants were recruited from public schools, while the others from private schools. The university students belonged to a wide-range of faculties, none of them being English majors. From the questionnaire distributed at the time of the group oral, we knew that no students had received education abroad with English as the medium of instruction.

4.2 Data collection and transcription procedure

The data on the group oral were collected from each educational institution through the following process: (1) A questionnaire was distributed with questions on the participants' backgrounds; (2) The students were randomly allocated into groups of three; (3) Each group drew a card on which one of the seven interaction topics - School, Family, Friends, Hobbies, English, Dream, and Culture (the last being only for university students) - was written down, and they were asked to speak on the topic; (4) Five minutes were allotted to each member of the group to plan his/her speech without speaking to the other members of the group; (5) Each member of the group introduced themselves for about half a minute as a warm-up activity; (6) Finally, the three students interacted orally as a group for five minutes on the selected topic. They were encouraged to have a natural and casual conversation while sitting and looking at each other. The interaction was videotaped after acquiring the permission of the participants.

The sound and movie files were separated using DVD Decrypter Ver. 3.5.4.0 (free software); subsequently, wav. files were created by means of DVD2V Ver. 1.86 (free software). All conversation saved as wav. files was transcribed with the aid of Transcriber Ver. 1.5.1. Along with transcribing, the DVDs compensated for the deficiency of the information.

4.3 Rating criteria and rating procedure

For the purpose of evaluating participants' oral interaction skills adequately and sufficiently, it is of great importance to use reliable,

Junko Negishi

well-established rating criteria. The criteria have to be designed specifically for the group oral and appropriately evaluate novice learners, such as junior high school students. It was adopted from the Common European Framework of Reference for Languages (CEFR: Council of Europe, 2001). It was beneficial to use the CEFR criteria because the Council of Europe has disclosed all information such as its rationale and framework as well as rating scales and a training DVD for raters (North & Hughes, 2003). The video included concrete samples for standardization training for English as case studies.

Ten Japanese teachers of English rated the participants. Before rating, they received training using the training DVD for raters. They rated the students by applying both a holistic rating scale and analytic rating criteria of the CEFR. The latter consists of five subcategories; Range, Accuracy, Fluency, Interaction, and Coherence. The raters assessed the students while watching their performance DVDs by 7 scales: Below A1, A1, A2, B1, B2, C1, and C2.

The CEFR oral assessment criteria grid for Fluency includes temporal variables related to the speed of speech, including pauses, and hesitation phenomena related to dysfluency, such as false starts, reformulation, hesitation, and repair.

4.4 Multi-faceted Rasch analysis

Multi-faceted Rasch measurement analysis eliminated various kinds of bias from the raw scores as much as possible and calibrated the CEFR *measures* which were utilized for the subsequent analysis. Utilizing the results obtained from the analysis, the next aim was to explore the relationship between the discrete variables obtained by the analysis and the measures obtained from the CEFR criteria. This study, however, solely reports the results of analysis concerned Fluency, one of the subcategories in the CEFR.

4.5 Methods of analysis

The author mainly drew on the analysis of Towell et al. (1996), Kormos and Denes (2004), and Wiese (1984) for the temporal variables and the analysis of Skehan and Foster (1999) and Ellis and Barkhuizen (2005) for the hesitation phenomena. The discrete item for analysis was selected based on whether or not multiple studies have investigated on them, and some differences were found between fluent and non-fluent speakers. Since fluency is often regarded as a major key for learners' proficiency, a variety of items that may exhibit speakers' fluency were explored.

Contrary to the work of Kormos and Denes (2004) and Wiese (1984), this study classified pauses as temporal variables: Pause length is categorized as a temporal variable in the studies of Lennon (1990) and Ellis and Barkhuizen (2005), while Wiese (1984) and Riggensbach (1991) classify the

L2 Speakers' Development and Raters' Perception on Fluency

number of pauses as a hesitation variable. In this study, pause-related items were classified as a temporal phenomenon rather than a hesitation phenomenon, based on Skehan's (1998) consideration of temporal variables as breakdown fluency and hesitation phenomena as repair fluency.

Many studies have reported that fluent speakers pause at grammatical junctures while non-fluent speakers do so within a phrase or clause (Cambers, 1997; Freed, 1995, 2000; Riegenbach, 1991; Towell et al., 1996; Segalowitz & Freed, 2004). Based on these studies, four locations of pauses were explored. Some characteristics of pauses were elicited qualitatively. In addition, the frequency of circumlocutions and paraphrases, which are in the Fluency grid of the CEFR, was counted alongside the temporal and hesitation variables.

4.5.1 Temporal variables

Speech rate

Speech rate refers to the number of syllables produced per minute. The total number of pruned syllables (excluding dysfluencies) produced by a speaker in a five-minute interaction was divided by the total speech time, including pause time. Since the obtained figure was expressed in seconds, it was multiplied by 60 to express the rate in syllables per minute.

Articulation rate

Articulation rate refers to the mean of syllables produced per minute. The total number of syllables (including dysfluencies) produced by a speaker in a five-minute interaction was divided by the total speech time, excluding pause time. The resulting figure was multiplied by 60 to express the rate in syllables per minute.

Mean length of runs

Mean length of runs refers to the mean of syllables produced between two pauses that last 0.25 seconds or more. There has been a lot of debate about the cut-off point, and no standard has been established. This study chose 0.25 seconds because many studies related to the quantification of fluency phenomena have employed this length.

Number of silent pauses

The number of silent pauses over 0.25 seconds in length was counted. Because each speaker produced different lengths of speech, the total number of silent pauses produced in a five-minute interaction was divided by the total speech time, expressed in seconds. The resulting figure was multiplied by 60.

Number of filled pauses

The filled pauses such as "mm", "uh...", and "eh:" were counted. Because each speaker produced different lengths of speech, the total number of filled pauses produced in a five-minute interaction was divided by the total speech time, expressed in seconds. The resulting figure was multiplied by 60.

Junko Negishi

Mean length of pauses

Mean length of pauses was determined by dividing the total length of silent pauses over 0.25 seconds by the total number of silent pauses over 0.25 seconds. Mean length of pauses gives us information about silence in an interaction.

The following variables were also calculated because they were counted/measured while obtaining the data above. The number of syllables per minute and the number of words per minute could also be indicators of quantity of talk. Dysfluency includes filled pauses and hesitations (repetitions, false starts, reformulations, and replacements).

Total speaking time including pause time

Total number of syllables including dysfluency

Total number of syllables excluding dysfluency

Total number of words

4.5.2 Characteristics and locations of pauses

Characteristics of pauses

Characteristics of pauses were explored including the length of silent/filled pauses and the use of fillers and hesitations.

Locations of pauses

The length and the number of four types of pauses were explored:

- 1) a pause at a clause boundary;
- 2) a pause at a phrase boundary;
- 3) a pause with dysfluent utterances (e.g., a pause before/after a hesitation phenomenon); and
- 4) a pause located at an unexpected place within a phrase.

Length and number were divided by the total amount of time, expressed in seconds, and multiplied by 60.

4.5.3 Hesitation phenomena

Repetitions

Repetitions referred to the immediate repetition of words, phrases, or clauses without modification, divided by the total amount of time, expressed in seconds, and multiplied by 60.

False starts

False starts were utterances or sentences that were abandoned before completion, divided by the total amount of time, expressed in seconds, and multiplied by 60.

L2 Speakers' Development and Raters' Perception on Fluency

Reformulations

Reformulations were phrases or clauses that were repeated with some modification, syntactically, morphologically, or by changing word order. These were divided by the total amount of time, in seconds, and multiplied by 60.

Replacements

Replacements were lexical items that were instantly replaced by other lexical items, which were divided by the total time, in seconds, and multiplied by 60. (These can occur either within the same clause, within a subsequent clause if the repetition is otherwise verbatim, or within a following clause if the repetition is a reformulation [Skehan and Foster, 1999, p. 107].)

Use of first language, Japanese

The use of Japanese was regarded as a hesitation phenomenon because less fluent speakers showed a tendency to use Japanese when they could not find an English word or did not know what to say in English. The instances of first language use was divided by the total time, in seconds, and multiplied by 60.

4.5.4 Other strategies found in CEFR

Circumlocution

The use of many words where fewer would do, especially in a deliberate attempt to be vague or evasive (e.g., a book that lists the words for "dictionary").

Paraphrase

The rewording of something spoken, approximation, and word coinage, excluding circumlocution.

4.5.5 Procedure for measuring pauses

Pauses were measured using Audacity version 1.2.6, which showed speech waves, with a lower cut-off point of 0.25 seconds. The author looked at speech waves while listening to the sound. Before analysis, white noise was eliminated so that pauses were easier to discern.

5 Results and Discussion

5.1 Temporal variables

This section investigated 10 temporal variables with potential to explain the participants' fluency. Table 1 shows descriptive statistics of the temporal variables displayed by the three educational institutions. Some variables explained in Kormos and Denes (2004) will be used for comparison.

Table 1. Descriptive Statistics of Temporal Variables

Temporal variables	Mean			S. D.		
	JHS	SHS	U	JHS	SHS	U
Speech rate	88.97	80.53	92.07	36.35	25.72	26.07
Articulation rate	162.56	163.24	160.75	42.17	33.05	26.67
Mean length of runs	2.93	3.06	3.24	0.75	0.69	0.79
Number of silent pauses	16.39	21.28	20.28	5.95	4.96	6.04
Number of filled pauses	2.83	5.40	6.91	3.10	2.96	3.45
Mean length of pauses	1.55	1.23	0.93	0.99	0.49	0.35
Total speaking time	44.18	73.73	85.17	28.17	32.88	48.20
Total number of syllables including dysfluency	63.96	111.84	154.71	33.53	51.31	90.91
Total number of syllables excluding dysfluency	55.13	92.36	124.60	25.42	39.75	67.25
Total number of words	44.76	71.07	93.29	21.87	30.68	48.98
CEFR measures (Fluency)	-6.55	-2.43	0.19	1.44	1.79	2.02

Note. First six variables are expressed per minute; JHS: junior high school; SHS: senior high school; U: university

5.1.1 Speech rate and articulation rate

In terms of the speech rate, no significant difference is observed between the educational institutions, nor is any developmental phenomenon perceived. This result confirms the Chambers' (1997) study in that speech rate alone does not contribute to the feeling of fluency. This index differentiates non-native speakers from native speakers, as Wiese (1984) claims, but does not seem to make a distinction between the participants of this study. The senior high school students produce the fewest syllables, 80.53, followed by the junior high school students, 88.97. The university students produce the most: 92.07. The discrepancy between the junior high school students and students of the other two institutions can be attributed to the fact that the former tend to rehearse utterances in their head in the time between turns, which is not calculated into the speech rate. The number of syllables produced by these participants totals only about 80% and 50% of the number produced by low-intermediate and advanced participants of Kormos and Denes (2004), respectively.

With regard to the articulation rate, the values indicate no significant difference among the educational institutions. The senior high school demonstrates the highest articulation rate and the university the lowest (JHS:

L2 Speakers' Development and Raters' Perception on Fluency

162.56; SHS: 163.24; U: 160.75), which may be because the articulation rate excludes pause time and includes dysfluency, such as repetitions and false starts. In other words, the articulation rate is calculated irrespective of content. The junior high school shows the largest standard deviation and the university the least. Compared with the study of Kormos and Denes (2004), in which the low-intermediate participants produced 227.45 syllables per minute and advanced participants 241.99, the speakers in this study produced 162.18 syllables on average, which amounts to 71% and 67%, respectively.

5.1.2 Mean length of runs

The mean number of syllables is around 3 (JHS: 2.93; SHS: 3.06; U: 3.24), indicating that the speakers utter only three words between the cut-off points. When compared with the low-intermediate participants' mean length (3.49) in Kormos and Denes (2004), this study shows no substantial difference; however, the figure of 6.23 for the advanced speakers is an enormous difference. Unlike the results of Kormos and Denes and Chambers (1997), the participants of this study are not distinguished by the mean length of runs. It is likely that until speakers reach a certain stage, e.g., the advanced level, the mean length of runs is not an indicator of fluency.

5.1.3 Number of silent and filled pauses per minute

After measuring each pause, silent and filled pauses over 0.25 seconds were counted. In regard to the number of silent pauses, no major difference or developmental feature regarding fluency is observed (JHS: 16.39; SHS: 21.28; U: 20.28). These numbers are much smaller than those of Kormos and Denes (2004 [low-intermediate: 31.2; advanced: 30.3]), which may be due to different speech/articulation rates; in other words, the more the participants speak, the more pauses they produce. Their study also shows that the number of silent pauses does not depend on the level of speaker. In contrast, Chambers' (1997) claim that the quantity rather than length of pauses contributes to fluency contradicts the result of this study.

The number of filled pauses seems to be related to the number of words produced and the speaking time (JHS: 2.83; SHS: 5.40; U: 6.91). Nevertheless, in the study of Kormos and Denes (2004), the more fluent the participants, the fewer filled pauses they produced (low-intermediate: 16.30; advanced: 8.28). As discussed earlier, the junior high school speakers seem to compose a sentence in their mind before articulating it, while university speakers are likely to verbalize extemporaneously, which may have resulted in more pauses than expected. More research is necessary to explain this phenomenon.

5.1.4 Mean length of pauses

The junior high school speakers demonstrate the longest pauses (1.55 seconds), followed by the senior high school speakers (1.23 seconds). The university

Junko Negishi

speakers demonstrate the shortest pauses (0.93 seconds), equivalent to the low-intermediate participants of Kormos and Denes (0.96). It seems that the fluency level of the most fluent speakers of this study is similar to the lower-intermediate participants of Kormos and Denes (2004). Considering that their advanced participants' mean length of pauses is 0.62 seconds, mean length of pauses appears to be a good indicator of speaking development.

5.1.5 Total speaking time including pause time

The total speaking time each speaker produces in a five-minute interaction, including pause time, averages 44.18 seconds for the junior high school, 73.73 seconds for the senior high school, and 85.17 seconds for the university. This increase in length clearly shows the participants' speaking development. During interactions, the junior high school students speak during only 44.2% of the time allotted, the senior high school students 73.75%, and the university students for 85.2% ($44.18 \times 3 / 300 \text{sec}$, $73.73 \times 3 / 300$, $85.17 \times 3 / 300$, respectively). This means that the junior high school participants spend more than half of the total time in silence, while the university students spend their time more effectively, spending only 15% in silence. As mentioned earlier, the junior high school students rehearse what they are going to say, which may account for much of the silence. The university students show the largest (48.20) standard deviation. The speech rate or the articulation rate only indicates how long participants produce "sound", excluding the time between turns. Kormos and Denes (2004) did not mention the total speaking time in their study.

5.1.6 Total syllables, including/excluding dysfluency, and total words

The total number of syllables each speaker produces, including dysfluency, during an interaction is 63.96 for the junior high school, 111.84 for the senior high school, and 154.71 for the university. These numbers clearly show the development of the participants' speaking ability. These numbers also indicate the quantity of talk. High standard deviations at the university level (90.91) indicate that the quantity of talk varies from speaker to speaker.

The total number of syllables excluding dysfluency also demonstrates significant differences among the educational institutions: 55.13 for the junior high school, 92.36 for the senior high school, and 124.6 for the university.

The words uttered in a five-minute interaction are also counted. In this case, words include filled pauses and partial words which are recognizable as words, containing not only a first consonant but also a vowel, based on Riegenbach (1991). The numbers are 44.76 for junior high school, 71.07 for senior high school, and 93.29 for university, suggesting a developmental phenomenon. Words per minute are also calculated but the results do not correlate with fluency. The number of syllables per word is as follows: JHS 1.43, SHS 1.57, and U 1.66 (including dysfluency) and JHS 1.23, SHS 1.30,

L2 Speakers' Development and Raters' Perception on Fluency

and U 1.34 (excluding dysfluency). In other words, the number of syllables per word increases along with education level. The study of Kormos and Denes (2004) counted the total number of words produced in two–three minutes, which does not allow for comparison with this study.

5.1.7 Correlation between temporal variables and CEFR measures

Table 2. Correlation (Spearman's rho) between Temporal Variables and CEFR Measures

	SR	AR	MLR	NSP	NFP	MLP	TST	NSID	NSED	TNW	CEFR
SR	1	.633**	.515*	-.127	-.076	-.541*	-.383*	.002	.099	.077	.134
AR		1	.384*	.120	-.066	-.106	-.151	.087	.167	.140	.069
MLR			1	-.149	.059	-.220	.084	.301*	.363*	.311*	.334*
NSP				1	.299	-.358*	.423**	.436**	.414**	.412**	.427**
NFP					1	-.254*	.428**	.495**	.431**	.440**	.503**
MLP						1	-.018	-.305*	-.314*	-.317*	-.410**
TST							1	.892**	.846**	.847**	.626**
NSID								1	.983**	.974**	.772**
NSED									1	.986**	.772**
TNW										1	.752**
CEFR											1

Note. N=135; * $p < .05$, ** $p < .01$; SR stands for speech rate, AR for articulation rate, MLR for mean length of runs, NSP for number of silent pauses, NFP for number of filled pauses, MLP for mean length of pauses, TST for total speaking time, NSID for number of syllables including dysfluency, NSED for number of syllables excluding dysfluency, and TNW for total number of words.

Kolmogorov-Smirnov Test was carried out and the result showed that three variables—articulation rate, the number of silent/filled pauses—were normally distributed; however other seven variables were significantly non-normal, $D(135) \leq 0.020$, $p < .05$. Rank order statistics were carried out and Table 2 shows the correlation coefficients (Spearman's rho) between the temporal variables and the participants' CEFR measures. The factors with the highest correlation with the CEFR measures are: 1) the total number of syllables either including or excluding dysfluency (NSID: .772, NSED: .772); 2) the total number of words (TNW: .752); 3) the total speaking time (TST: .626); 4) the number of filled pauses (NFP: .503); 5) the number of silent pauses (NSP: .427); and 6) the mean length of pauses (MLP: -.410). Nonetheless, in terms of the number of filled and silent pauses denoted by 4) and 5), they are expected to show a negative correlation. It is likely that the number of filled and silent pauses correlates with the amount of talk, but this finding may be inconclusive. As was described earlier, the speech rate, the articulation rate, and the mean length of runs do not contribute to the CEFR measures.

5.2 Characteristics and locations of pauses

5.2.1 Characteristics of pauses

This section will explore some characteristics of pauses in detail. Excerpts 1 – 3 display parts of the discourse carried out by each of the three different educational institutions. Silent pause time over the threshold, 0.25 seconds, is indicated in parenthesis, e.g., (0.43), indicating 0.43 seconds, as seen in Turn 1. Filled pauses are shown with { }, e.g., {mm, 0.52}, demonstrating a sound, “mm”, that lasts 0.52 seconds, as seen in Turn 2. Hesitation phenomena such as repetitions, false starts, reformulations, and replacements are shown as follows: [who, 0.67], as seen in Turn 1, indicates a repetition that lasts 0.67 seconds. The leftmost and rightmost columns designate the starting and ending times of the utterance, e.g., 30’56.20, indicating 30 minutes 56.20 seconds. A period “.” denotes falling tone and a question mark “?” denotes rising tone, not always indicating the end of a sentence or a question. If this were a monologue, a pause between the sentences would be measured; however, it was an interaction between three speakers, and it was difficult to attribute pauses to a particular speaker between turns. For this reason, pauses between turns were not included in the analysis.

Excerpt 1. (Junior High School: Group 7)

Starting time	T	S	Utterance	Ending time
30’56.20	1	L:	[who, 0.67] (0.43) who do you (0.66) like your (0.36) family (0.40) [with, 0.35] (0.67) with me?	31’03.22
31’04.22	2	M:	I like (1.33) {mm, 0.52} [mama, 0.46]. mommy (1.53). She makes (1.41) hamburg (0.27) very good (1.15). {hh <i>mm-toh</i> , 1.95} (0.39) How about you?	31’17.68
31’19.22	3	R:	[I, 0.40] I (0.35) like (0.45) father.	31’21.28
31’22.12	4	L:	{mm mm uh..., 0.58} (0.27) why (0.43) please say (0.25) it again.	31’29.68
31’59.00	5	R:	oh I like (1.58) very kind.	32’01.86
32’03.39	6	L:	{oh: <i>mm-toh</i> mm ah mm, 3.62} (0.93) do you have a pet with you?	32’10.69
32’11.28	7	M:	yes I do (1.13). { <i>mm-toh</i> , 0.75} (1.95) my pet (0.51) names (0.67) cocoa.	32’18.73
32’19.40	8	L:	{mm, 0.76} (1.46) [what, 0.26]? (0.64) { <i>mm-toh</i> , 0.93} (4.84) [what kind, 0.99] (0.72) { <i>mm-toh</i> , 1.11} (3.44) [pet, 0.26] (1.30) what (0.26) kind (0.28) of (0.31) pet?	32’38.71

Note. T stands for Turn; S for Speaker; Numbers in (parentheses) indicate pauses between two words over 0.25 seconds. Words and numbers in {parentheses} and [parentheses] indicate the content and length of filled pauses and hesitation phenomena, respectively.

As Excerpt 1 shows, one of the junior high school groups tends to pause within a clause: for example, for 0.36 seconds between “your” and “family” in Turn 1. Speaker R, in Turn 3, pauses at every word: “I (0.35) like

L2 Speakers' Development and Raters' Perception on Fluency

(0.45) father.” This is what Freed (1995, 2000), Riggenbach (1991), Towell et al. (1996), and Segalowitz and Freed (2004) have reported. On the other hand, in Turn 6, Speaker *L* produces a set of utterances (“Do you have a pet with you?”) without any pauses after a long hesitation (“oh: *mm-toh* mm ah mm”) of 3.62 seconds and a subsequent 0.93-second silent pause. It is likely that the speaker is planning what to say during the hesitation and pause. In Turn 8, however, the same speaker, *L*, struggles before saying “What kind of pet?” and generates repetitions: [what, 0.26], [what kind, 0.99], and [pet, 0.26]. This utterance also contains many filled pauses: {mm, 0.76}, {*mm-toh*, 0.93}, and {*mm-toh*, 1.11}.

Excerpt 2. (Senior High School: Group 30)

12'16.61	1	<i>L:</i>	do you have any brothers (0.63) or sisters?	12'19.56
12'19.82	2	<i>M:</i>	{eh, 0.24} I have one brother. (0.24) and he go to junior high school. (2.00) How about you.	12'26.04
12'26.74	3	<i>R:</i>	{uh:, 0.63} my brother is {eh:, 0.21} (1.06) two. {eh, 0.19} my large brother is {eh:, 1.64} (1.58) [have, 0.50] (1.78) have been [to, 0.29] in G***** (0.51). {uh:, 0.41} but (0.43) [he, 0.28] (0.73) he have gone to (0.97) G***** for (0.41) four years. {uh:, 0.25} (2.28) my (0.53) {eh, 0.51} little brother (0.67) {uh:, 0.94} is (0.71) {eh, 0.31} [ten year, 0.68] ten year old. {uh:, 0.31} my little brother (1.02) is not same (0.53) me (1.43). {uh, 0.33} (0.65) [how many, 2.32] (1.43) {eh <i>mah</i> , 0.75} (0.25) [how many people, 0.92] {uh, 0.31} [how many how many how many, 4.57] how many family do you...	13'22.57
13'23.23	4	<i>L:</i>	[10.28] {uh eh, 0.60} [I, 0.23] I have (1.03) a father and (0.50) mother (0.66). I have no brothers and sisters (0.85). {um, 0.34} my father is a high school teacher (0.49). he teaches biology. {uh e:h uh, 2.05} (2.81) [my father my father oh no no no no, 4.81] [my mother, 0.75] (0.25) my mother (1.08) {uh, 0.26} (0.30) works for (0.83) {er, 0.45} (0.45) elementary school (0.99). {um, 0.50} (0.98) she works (1.01) {mm, 0.75} (0.63) about (1.66) {mm, 0.23} three (0.35) or (0.27) four days (0.28) or a week (3.33). {um, 0.38} and my mother (1.09) have a lot of housework (0.74) {oh, 0.17} [she, 0.54] (0.56) {oh, 0.24} I sometimes help her (1.97). {mmm uh, 0.88} (1.36) what do you think about (0.79) your father.	14'27.65

Excerpt 2 is an example of a private senior high school group. In Turn 2, Speaker *M* pauses at a grammatical juncture: “I have one brother. (0.24) and he go to junior high school [*sic*]. (2.00) How about you.” On the other hand, when Speaker *R* produces long utterances in Turn 3, he demonstrates many repetitions, self-corrections, and pauses: “[how many, 2.32] (1.43) {eh *mah*, 0.75} (0.25) [how many people, 0.92] {uh, 0.31} [how many, how many, how many, 4.57] how many family do you... [*sic*]”. The senior high

Junko Negishi

school groups' average number of silent pauses is the largest among the three educational institutions (see Table 1). The private senior high school students, as this excerpt indicates, tend to produce longer utterances compared with the students at other educational institutions. They do not easily give up producing talk. They ask questions when they yield the floor, as seen at the end of every utterance in the excerpt. They do not interrupt their classmate's talk even when the speaker hesitates or pauses a lot.

Excerpt 3 (University: Group 42)

20'38.53	1	<i>M:</i>	do you have any idea of [cul-, 0.30] Japanese culture?	20'41.60
20'42.08	2	<i>R:</i>	Japanese culture.	20'43.03
20'47.06	3	<i>L:</i>	Japanese culture.	20'48.12
20'55.67	4	<i>M:</i>	I think that (0.38) {uh, 0.24} through [my, 0.25] (0.52) my (1.38) bizmate (0.34) with Korean students (1.03)? my partner knows (0.26) lot of about (0.46) Japanese comic books (0.69) then (1.71) so (2.52) Korean people (0.53) knows about Japanese culture well a lot (0.81) but (0.50) Japanese (0.97) students don't (0.28) know (0.42) much about Korean culture?	21'25.21
21'27.27	5	<i>L:</i>	{uh::, 1.02} I agree with you. [I'm, 0.44] (0.45) when I was a high school student? I went to (0.65) Canada [to, 0.27] (0.59) to join the culture (0.56) exchange program (0.38) and [there are, 0.38] {uhm, 0.27} (0.26) there were a lot of (0.75) Korean students (0.42) and [they know, 0.70] they knew about Japanese culture a lot. but [I don't, 0.50] I didn't know about Korean culture (0.55) and [they, 0.71] (1.36) they loved to listen X Japan (0.68)? [but, 0.60] (0.89) but you know {uhm, 0.19} (0.62) [our, 0.31] {eh, 0.29} our generation [don't, didn't have a, 1.77] didn't know a lot about X Japan (0.50)? {uh, 0.62} so (0.56) they (0.53) really wanted to talk about X Japan with Japanese students but we don't know about that (0.43) so [we don't, 0.50] we don't know about that so it's interesting aren't they.	22'24.25

An example of a university group, Excerpt 3, reveals that these speakers also pause often, but the length of each pause is shorter. Speaker *M* in Turn 4 sometimes pauses at every single word: "I think that (0.38) {uh, 0.24} through [my, 0.25] (0.52) my (1.38) bizmate (0.34) with Korean students (1.03)?" Speaker *L* in Turn 5 demonstrates a number of self-corrections, specifically from present tense to past tense, as underlined in the following example: "[there are, 0.38] {uhm, 0.27} (0.26) there were a lot of (0.75) Korean students (0.42) and [they know, 0.70] they knew about Japanese culture a lot. but [I don't, 0.50] I didn't know about Korean culture". What we learn from this excerpt is that although the length of pauses becomes shorter as educational level increases, even the university students' interactions included many pauses and hesitations.

L2 Speakers' Development and Raters' Perception on Fluency

5.2.2 Locations of pauses

As reported by researchers, non-fluent speakers often pause within clauses (Freed, 1995, 2000; Riggensbach, 1991; Towell et al., 1996; Segalowitz & Freed, 2004), and this phenomenon is reinforced in this study, as seen in Excerpt 1 – 3. At each point of pause, four types of pauses were investigated in terms of length and number. The findings are expressed as the mean length and number of pauses observed per minute.

- 1) Pause at a clause boundary
- 2) Pause at a phrase boundary
- 3) Pause before/after a dysfluent utterance (hesitations/filled pauses)
- 4) Pause within a phrase (at an unexpected place)

Table 3. Mean Length of Pauses and Mean Number of Pauses per Minute per Participant, Sorted by Locations of Pauses

	Mean length of pauses			Mean number of pauses		
	JHS	SHS	U	JHS	SHS	U
Clause boundary 1)	3.42	4.36	2.26	1.58	2.99	2.19
Phrase boundary 2)	2.31	4.27	3.00	1.49	3.61	3.44
Around dysfluency 3)	7.49	7.64	6.13	4.33	5.40	6.14
Within a phrase 4)	9.59	8.74	6.58	8.97	9.22	8.26

Table 3 shows the mean length of pauses and the mean number of pauses per minute per participant, sorted by location. Not only fluent but also non-fluent speakers pause at a clause or phrase boundary (1 and 2), while non-fluent speakers often pause within a phrase at an unexpected place (4). Although there is no significant relationship between the pause location and development, the mean length of pauses within a phrase decreases as educational level increases. In terms of the mean number of pauses, no distinctive phenomena are observed except for the number of pauses that occur before/after dysfluencies. This is not always in accord with the hypothesis that the higher the level of education, the shorter/fewer the pauses.

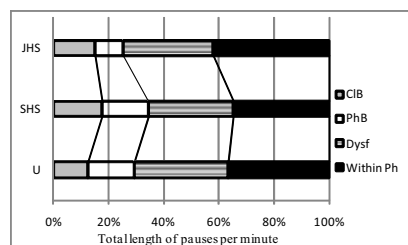


Figure 1. Proportion of four pause locations displayed by the length of pauses per minute
Note. CIB = clause boundary; PhB = phrase boundary; Dysf = before/after dysfluent utterance; Within Ph = within a phrase

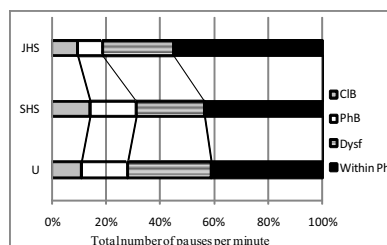


Figure 2. Proportion of four pause places displayed by the number of pauses per minute

Junko Negishi

Figures 1 and 2 display the proportion of pauses at each of the four locations. In both figures, the senior high school participants pause most frequently at a clause boundary or a phrase boundary, that is, at acceptable locations. In terms of the number of pauses per minute, the junior high school participants demonstrate the lowest proportion of pauses at acceptable locations. In other words, junior high school students pause at unexpected locations (within phrases) the most, followed by the senior high school and then the university students. This is the only phenomenon that varies according to educational level. Although various approaches were taken to elicit characteristics about the location of pauses, no other features were found.

5.2.3 Correlation between locations of pauses and CEFR measures

There were not many developmental features found amongst the educational institutions, but correlation coefficient towards the CEFR measures for Fluency showed some characteristics. Table 4 displays correlation coefficient (Spearman's rho) between the place of pauses and the CEFR measures. As can be seen, the correlation basically shows low values. Among the eight variables, the number of pauses at a phrase boundary shows the highest correlation, $.525, p < .01$. The second highest also relates to a phrase boundary in terms of the length of pauses, $.326, p < .01$. This result suggests that speakers pause at a phrase boundary tend to obtain higher measures.

Table 4. Correlation Coefficient (Spearman's rho) between Location of Pauses and CEFR Measures

	Mean length of pauses				Mean number of pauses				CEFR
	Within Ph	Dys	PhB	CIB	WithinPh	Dys	PhB	CIB	
WithinPh	1	-.105	-.138	-.117	.657**	-.175*	-.158	-.148	-.243**
Dys		1	.106	.107	-.190*	.682**	.053	.059	.046
PhB			1	.258**	-.157	.104	.825**	.259**	.326**
CIB				1	-.096	.027	.123	.848**	.163
WithinPh					1	-.062	-.068	-.040	-.031
Dys						1	.183*	.082	.315**
PhB							1	.192	.525**
CIB								1	.308**
CEFR									1

Note. N=135; * $p < .05$. ** $p < .01$

5.3 Hesitation phenomena and other strategies found in CEFR criteria

Hesitation phenomena, namely, repetitions, false starts, reformulations, replacements, and the use of Japanese, and other strategies found in the CEFR criteria such as paraphrase and circumlocution were explored next. Nevertheless, there were only 10 instances of paraphrase (9 participants out of 135 used paraphrasing) and no instances of circumlocution. This may be because, as Bialystok (1990) argues, these strategies are too demanding for

L2 Speakers' Development and Raters' Perception on Fluency

lower-level learners because of the heavy linguistic loads. It is likely that intermediate learners, at least, can make use of the strategies. Consequently, these strategies were eliminated from the analysis, and only the hesitation phenomena will be explained here. Table 5 shows descriptive statistics of the mean hesitation frequencies per minute sorted by educational institution.

Table 5. Descriptive Statistics of Hesitation Frequencies per Minute

Hesitation variables	Mean			S. D.		
	JHS	SHS	U	JHS	SHS	U
Repetitions	3.15	3.25	4.20	2.76	2.64	2.68
False starts	0.08	0.26	0.32	0.27	0.51	0.84
Reformulations	0.25	0.77	0.78	0.56	0.88	0.99
Replacements	0.56	0.77	0.72	0.94	1.00	0.77
Use of Japanese	3.59	0.75	0.85	4.55	1.38	1.39

5.3.1 Repetition

Repetitions are defined as the immediate repetition of words, phrases, or clauses without modification. For example:

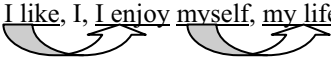
(Junior High School: Group 7, Speaker *L*, Turn 1)
Who, who do you like your family with, with me? [*sic*]

The mean frequency of repetitions per minute is highest among the university students, 4.20, followed by the senior high school students, 3.25, and the junior high school students, 3.15. The difference is greatest between the university and the other two institutions. The junior high school speakers tend to pause for a long time while planning what to say, which leads to fewer repetitions. On the other hand, university students seem to speak extemporaneously, which may cause more repetitions.

5.3.2 False start

False starts are utterances or sentences that are abandoned before they are completed.

(Senior High School: Group 29, Speaker *L*, Turn 1)
 ... I like, I, I enjoy myself, my life in school...



The participants do not regularly use false starts; the total number of false starts is less than 1/10 the total number of repetitions. As with repetitions, the most false starts are demonstrated by the university students, 0.32, followed

Junko Negishi

by the senior high school students, 0.26, and the junior high school students, 0.08. Here, the difference between the senior high school and the university is not great, but there is a significant difference between the junior high school and the other two institutions. This may be because the junior high school students rehearse what to speak.

5.3.3 Reformulation

Reformulations are phrases or clauses that are repeated with some modification. Reformulation is the second least observed hesitation phenomenon of the five.

(University: Group 36, Speaker *M*, Turn 8)
Did you <<Japanese words>> have you been to Kyoto?

The senior high school speakers reformulate about the same as the university speakers, while the number of reformulations among the junior high school students is significantly lower. Reformulations, which require speakers to modify their utterances syntactically or morphologically or to change word order, may be difficult for novice learners.

5.3.4 Replacement

Replacements are expressed as lexical items that are instantly replaced by other lexical items.

(Junior High School 7, Speaker *M*, Turn 2)
I like, mm, mama. Mommy.

There are fewer differences among the three institutions compared with other hesitation phenomena: 0.56 for the junior high school speakers, 0.77 for the senior high school speakers, and 0.72 for the university speakers.

5.3.5 Use of Japanese

Use of Japanese refers to the number of times the participants use their mother tongue.

(Senior High School: Group 20, Speaker *M*, Turn 52)
First <<Japanese: ichi>>, second <<ni>>. One <<ichi>>, two.

A large disparity is observed between the junior high school (3.59 times per minute) and the other two institutions (0.75 for the senior high school and 0.85 for the university). The use of their native language may stem from

L2 Speakers' Development and Raters' Perception on Fluency

attempts to avoid breakdowns. This phenomenon can be used to distinguish novice learners from other participants. Basically, the frequencies were small, and not every participant employed hesitations, so it may be difficult to draw a conclusion.

5.3.6 Correlation between hesitation variables and CEFR measures

Table 6. Correlation Coefficient (Kendall's tau) between Hesitation Variables and CEFR Measures

	Repetition	False start	Reformulation	Replacement	Japanese	CEFR
Repetitions	1	.050	.152*	.011	-.005	.144*
False starts		1	.091	.098	-.187*	.181**
Reformulations			1	.060	-.056	.265**
Replacements				1	.030	.162*
Use of Japanese					1	-.323**
CEFR measures						1

Table 6 shows a correlation coefficient (Kendall's tau) between the hesitation variables and the CEFR measures. All of the hesitation variables demonstrate some correlation with the CEFR measures to some degree, but the correlation coefficients are not high. Among them, the use of Japanese is likely to be the best indicator of hesitation phenomena, that is, the less usage of Japanese, the more fluent the speaker.

6 Conclusion

The study explored the participants' fluency with reference to the two major aspects: temporal variables and hesitation phenomena, including such features of pauses as characteristics and pause placing. To investigate the extent to which variables and phenomena influence fluency, various items that would have the potential to explain the speakers' development were analyzed.

In terms of the temporal variables, the mean length of pauses, the total speaking time including pause time, the total number of syllables including/excluding dysfluency, and the total number of words demonstrated significant differences among the three educational institutions, that is, the level of education was likely to impact the participants' fluency development. In contrast, the speech rate, the articulation rate, the mean length of runs, and the number of silent and filled pauses, contrary to other research, were not determined by the educational level. The results suggested that the participants of this study were at a similar or lower level than the lower-intermediate speakers of Kormos and Denes (2004). Wiese (1984) reported

Junko Negishi

that different speech rates were observed between native speakers and non-native speakers. Comparing the speech rates and articulation rates of low-fluency speakers seemed to be unproductive. The correlation between the temporal variables and the CEFR measures indicated that the total number of syllables including/excluding dysfluency and the total number of words could be indices for higher measures.

An analysis of pauses showed that the junior high school students tended to pause within clauses, the senior high school and university students were likely to pause at grammatical junctures, and the university students paused for a shorter time. Nevertheless, sorting the mean length of pauses and the mean number of pauses by the four pause locations—at a clause boundary, phrase boundary, before or after dysfluency, and within a phrase—suggested that only the unexpected placement of pauses was a distinguishing factor; the less fluent speakers paused at unexpected places (within phrases) the most and university students the least. Otherwise, the pause location did not seem to play an important role in determining speakers' fluency.

There were not many instances of hesitation variables, and it might be difficult to identify relationships between the hesitation phenomena and the students' development. The correlation between hesitation variables and the CEFR measures were not significant, either. The only item that indicated a negative correlation was the use of Japanese; less usage of Japanese implied more fluency.

Although fluency is said to be a major factor in judging L2 speakers' proficiency, this study did not clearly support the accepted notions. It seems that the results stemmed from the participants' low proficiency, whereas other studies investigated more advanced speakers. However, this study represents the current situation of students in Japan, with the exception of returnees. The three indices, the total number of syllables including/excluding dysfluency and the total number of words, can be useful when confined to Japanese students.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in Practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bialystok, E. (1990). *Communication strategies: A psychological analysis of second language use*. Oxford: Blackwell.
- Bonk, W. J., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing*,

L2 Speakers' Development and Raters' Perception on Fluency

- 20(1), 89–110.
- Canale, M., & Swain, M. (1980a). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1–47.
- _____. (1980b). A domain description for core FSL: Communication skills. In S. J. Savignon, & M. S. Berns (Eds.), *Initiatives in Communicative Language Teaching* (pp. 27–39). Massachusetts: Addison-Wesley Publishing Company.
- Chambers, F. (1997). What do we mean by fluency? *System*, 25, 535–544.
- Chen, S-Q. (1990). A study of communication strategies in interlanguage production by Chinese EFL learners. *Language Learning*, 40, 155–187.
- Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Davies, N. F. (1978). 'Putting receptive skills first -- an experiment in sequencing'. Paper read at the 5th AILA congress, Montreal, August 1978. Mimeo. In Canale, M. and Swain, M. 1980. "Theoretical Bases of Communicative approaches to Second Language Teaching and Testing" *Applied Linguistics*, 1, pp. 1-47.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Philadelphia: John Benjamins.
- _____. (2000). Is fluency, like beauty, the eyes, of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243–265). Ann Arbor: University of Michigan Press.
- Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing*, 13, 23–51.
- Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency*, (pp. 5–24). Ann Arbor: University of Michigan Press.
- Kormos, J., & Denes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145–164.
- Lennon, P. (1984). Retelling a story in English. In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions* (pp. 50–68). Tubingen, Germany: Funter Narr Verlag.
- _____. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387–412.
- Ministry of Education, Culture, Sports, Science and Technology (2002). *Developing a strategic plan to cultivate "Japanese with English abilities": Plan to improve English and Japanese abilities*. Retrieved January 3, 2010

Junko Negishi

- from <http://www.mext.go.jp/english/news/2002/07/020901.htm>
- Mohle, D. (1984). A comparison of the second language speech production of different native speakers. In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions* (pp. 26–49). Tübingen, Germany: Gunter Narr Verlag.
- North, B. & Hughes, G. (2003). *CEF illustrative performance samples: For relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment* (CEF). Eurocentres and Migros Club Schools.
- Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Processes*, 14(4), 423–441.
- Schmidt, R. (2000). Foreword. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. i–vii). Ann Arbor: University of Michigan Press.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173–189.
- Skehan, P. (1998). A cognitive approach to language learning. Oxford: Oxford University Press.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119.
- van Gelderen, A. (1994). Prediction of global ratings of fluency and delivery in narrative discourse by linguistic and phonetic measures: Oral performances of students aged 11–12 years. *Language Testing*, 11, 291–319.
- Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing*, 23(4), 411–440.
- Wiese, R. (1984). Language production in foreign and native languages: Same or different? In H. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions*. Tübingen: Gunter Narr.
- Yoshida-Morise, Y. (1998). The use of communication strategies in language proficiency interviews. In R. Young, & A. W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 205–238). Amsterdam: John Benjamins.

Junko Negishi
Waseda University
2-722-10, Mito, Ibaraki,
311-4151, Japan
Phone & Fax: +81 29 252 4866
Email: jnegishi@dream.com

L2 Speakers' Development and Raters' Perception on Fluency

Received: June 23, 2011

Revised: November 4, 2011

Accepted: December 7, 2011

Junko Negishi

Appendix

Transcript Notations

[Overlapping utterances
=	Latching that indicates no interval between adjacent utterances
.	Falling intonation: e.g., sentence final
?	Rising intonation (does not mean a question)
CAPITAL	Stressed syllable
:	A prolonged stretch
...	Unfinished utterance
<i>Italics</i>	Japanese words
(inaudible)	Inaudible or incomprehensible utterance
(laughter)	Laughter particle
/word/	Severely mispronounced word
<< >>	Author's description