Complexity and Fluency Indicators of "Good" Speakers

Robert W. Long III
Kyushu Institute of Technology

Reference Data:
Long, R. W. (2014). Complexity and fluency indicators of "good" speakers. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings*. Tokyo: JALT.

Although researchers have tried different ways of explaining the concept of fluency, few have examined speakers with different TOEIC (Test of English for International Communication) scores and specifically described the similarities and differences in their fluency. An overreliance on TOEIC scores and other testing measures masks problems related to language use and production. This study was aimed at identifying similarities and differences in the fluency of speakers within one range of TOEIC scores (683-793) using four kinds of discourse tasks. Fluency indicators were consistent for both monologues and dialogues except for fluency rates and pause frequencies. Syntactic complexity was found to be twice as high in dialogues as in monologues. As for dysfluency, articulation rates were consistent for most speakers, whereas a significant variation in the percentage of silence was noted, ranging from 16.0% to 44.7%. Information regarding how to improve fluency in the classroom is given.

研究者は流暢さという概念の説明を様々に行ってきたが、TOEICのスコアと特に流暢さについての類似点と相違点で話し手を調べるという研究はほとんど行われたことはない。TOEICスコアやそれ以外のテストの方法に過度に依存することは言語の使用と発話に関する問題を隠すことになりかねない。今回の研究は、4種類の談話の課題を使用して、TOEICスコアの(683-793)の間の一つの領域内で、話し手の流暢さの類似点と相違点を追求しようとするものだ。得られたデータから、流暢さの指針(indicators)は流暢さの比率(fluency rates)とポーズの頻度を除いて、モノローグ（独白）と発話の両方のモードで矛盾がないことが示された。統語的な複雑度（syntactic complexity）は、モノローグに比べて、発話のほうが2倍も高いことが分かった。失流暢については、沈黙度(silence)が16%から44.7%にまで及ぶなど有意義な相違点が見られるものの、ろれつ度（音声明瞭度）(articulation rates)は大半の発話者について矛盾はなかった。後半部で教室において流暢さを向上させる為の情報が提供される。

It is all too common in almost every English language classroom for teachers to become bogged down by issues relating to grammar and usage. Year after year, in classrooms around the world, grammatical forms are endlessly explained and exemplified several times before being tested—and this drilling and testing is done many times over the school year. The underlying motivation for this approach to teaching is the belief that *form* is far more important than *fluency*. This neglect of fluency development is compounded by the fact that fluency is misunderstood and poorly defined. Thus, in both EFL composition and speaking classes, students tend to decrease the speed with which they speak as well as their lexical and syntactic complexity so as to limit the number of grammar-related errors. These decreases in speed and complexity, in turn, affect their fluency. To get passing marks, students stick to safe, slow, and short sentences, often avoiding long comments with dependent clauses, parentheti-

cal comments, complex noun and verb phrases, gerunds, and adverbial clauses, to name a few forms or structures.

Although fluency has been extensively researched (Ferreira, Lau, & Bailey, 2004; Foster, Tonkyn, & Wigglesworth, 2000; Griffiths, 1990; Lennon, 1990; Lewin, McNeil, & Lipson, 1996; Towell, Hawkins, & Bazergui, 1996; Yuan & Ellis, 2003), with studies focusing on various indicators such as articulation rates (ARs), fluency rates, pause duration and frequency, mean length runs (MLRs), and even volume (delivery), one issue that has remained unexamined is how fluency differs (if at all) among speakers having different TOEIC (Test of English for International Communication) scores. In short, does a higher proficiency in listening, grammar, and reading correlate with improved fluency?

Based on the data collected from videotaped interviews of EFL speakers having TOEIC scores between 683 and 793, the general aim of this study was to examine whether fluency differed between monologues and dialogues, and if so, on which particular fluency indicator(s). This range of scores was selected as the scores are well above what Japanese industry uses for selecting individuals to work in areas where English is the primary language. Secondly, the study was designed to discover if there were particular dysfluency indicators, such as false starts, repetitions, rephrasing, high percentages of silence, and filled pauses, that were common to individuals within this range of TOEIC scores. Third, as fluency also involves syntactic complexity, a further aim was to investigate if there were any significant differences between monologues and dialogues. The findings could help indicate whether or not EFL speakers have more problems in monologues (such as presentations) or in dialogues and examine the issues of dysfluency and complexity in a more meaningful way.

Review of Literature

Fluency

Fluency has been without a doubt one of the most difficult concepts to define and understand, much less teach. Fillmore (1979) distinguished four kinds of native-speaker fluency: (a) "the ability to fill time with talk," which was commonly known as "disc-jockey fluency," (b) "the ability to talk in coherent, reasoned, and semantically dense sentences," (c) "the ability to have appropriate things to say in a wide range of contexts," and (d) "the ability . . . to be creative and imaginative in language use" (p. 93). Brumfit (1984) drew attention to how fluency differs from other elements of oral proficiency in that it is a performance phenomenon, taking into account pragmatics, grammar, lexical range, syntactic complexity, and idiomatic expression.

Koopmans-van Beinum and van Donzel (1996) examined fluency in both prepared and spontaneous speeches and found that speakers used pausing strategies to structure the continuation of the discourse. Lewin et al. (1996) examined pauses and verbal dysfluencies as an indication of speaking anxiety. They investigated whether speech disruptions, periods of silence, and a slower rate of speech were more prevalent in high-anxiety subjects than in their low-anxiety counterparts. After examining categories of pauses, pause length, verbal errors (corrections, distortions, fragments, and repetitions), and delaying verbalizations, the researchers found that the measures of anxiety immediately before and during the speech task did not correlate with dysfluencies or pauses. They concluded that pausing may constitute a means of escape or relieving stress. The researchers also introduced the issues of anxiety, whether monologues or dialogues are more stressful, and which one produces more dysfluencies as a result of this stress.

Chambers (1997) took into consideration the direct link between strategic competence and fluency, noting that fluency in

speech production is influenced by factors well beyond grammatical knowledge. Chambers examined AR (defined as the number of syllables uttered per second) but noted that AR has less impact on the perception of fluency than the length, nature, and location of pauses in the utterance. Chambers also distinguished between *natural* and *unnatural* pauses, with natural pausing occurring at clause junctures or after a semantic unit. These pauses generate the listener's expectation about prospective utterances and signal emphasis. Unnatural pauses are those that appear at other places and could indicate lexical or morphological uncertainty. These pauses may be related to anxiety, which can then lead to ungrammatical English, or they may function as a way to *fill in* particular words or just to give the speaker time to think about how to proceed.

The main reasons why the pausing of foreign-language learners differs from that of native speakers are as follows: (a) the distribution of pauses as a result of L1 aspects of pausing, (b) the greater overall pausing time in nonnative production with more pausing, and (c) shorter word groups or MLRs between pauses. This aspect of fluency became a more important factor in the research of Wendel (1997) and Yuan and Ellis (2003); both studies used a fluency measure that took into account both the amount of speech and length of pauses.

Research on fluency has shown the importance of pauses and hesitation; however, there are several gaps in the research. First, there is little to no information on how fluency in spontaneous speech changes with proficiency, particularly as measured by standardized test scores that are based on listening, grammar, and reading, such as the TOEIC test. Second, it has not been noted which kinds of dysfluency are the most noticeable to listeners and how dysfluent speech changes with increased proficiency, specifically which indicators decrease with more ability. Third, researchers have not examined whether fluency indicators are similar among various speakers (having similar test scores) from various countries. Fourth, more research is needed on how fluency differs (if at all) in monologues and dialogues.

Syntactic Complexity

In considering what makes for a good speech or piece of writing, one of the underlying issues is that of syntactic complexity: how sentences are arranged and connected in order to create effective meaning. In the past, there was a dearth of precise definitions and convincing approaches to operationalize these concepts, so it seemed that researchers avoided or sidestepped the issue or used poorly designed contexts and constructs. Givon (1991), drawing on a sizable body of psycholinguistic studies, asserted that complexity studies should take into account that subordinate clause structures are more complex to process than conjoined main clause structures. Thus, by counting linguistic tokens considered telltale signs of increased grammatical subordination and embeddedness, one can understand the complexity of a particular narrative. Some examples of these linguistic tokens include: (a) subordinating conjunctions (for instance, *because, since, as, when, that*); (b) WH-pronouns (*who, whose, whom, which*); (c) verb forms, both finite and nonfinite; and (d) noun phrases.

Szmrecsányi (2004) pointed out that complexity (or scope) could be understood by either taking into account pure length, duration, and size of the unit or by appealing to notions that are not related to these constructs. The main issue was that length does not necessarily correlate as a measure of *syntactic relationships*. Szmrecányi used the following example:

1. I wasn't there because I had to fill out all this.
2. I didn't do it, and the reason for this was that . . .

Both 1 and 2 have a length of 11 words. The first contains a main clause and a dependent adverbial clause of reason, but the second has a compound clause with the word *that* which might

take on any reason (e.g., the reason for this was that I was sick) or might not (e.g., the reason for that was that guy) (Szmrecányi, 2004, p. 1032). Complexity becomes more of an issue due to the word *that*, which, in turn, introduces a syntactically dependent complement clause. As Szmrecányi explained, the measure of scope interferes with one's ability to assess syntactic complexity.

Norris and Ortega (2003) indicated that complexity, as measured by subordination ratio, might not always increase linearly, but that syntactic complexity may grow in other ways, for example, through phrasal and clausal complexification. Yuan and Ellis (2003) also agreed with this concept of equating complexity with phrasal and clausal complexification, stating that "measures of complexity are generally based on the extent to which subordination is evident" (p. 2) (i.e., number of clauses per T-unit or c-unit). In some studies, lexical complexity has been assessed by means of type-token ratio.

For many other scholars, reducing complexity to type-token ratios and to the number of clauses does not provide for an in-depth understanding of the term. Skehan (1996) noted that complexity "concerns the elaboration or ambition of the language that is produced" and that complexity should also take into consideration "learners' preparedness to take risks" (p. 22). By involving the concept of semantics, pragmatics, and meaning, Skehan took the issue of complexity a step further by asking what effect complexity had on the overall message. O'Loughlin (1995) found lexical density (a measure of the relationship between grammatical items and high- and low-frequency lexical items of oral performance) to be influenced by test formats (live or tape-recorded) and task types (describing a familiar setting or a role-play), as well as the interactions of the two. However, the issues of how (if at all) complexity changes and differs in various forms of communication and how complexity may differ between fluent and dysfluent speakers have yet to be explored.

Complexity is the "extent to which learners produce elaborated language" (Ellis & Barkhuizen, 2005, p. 139) and is often related to the syntactic and lexical aspects of narrative performance. Of course, complexity has little value if the speaker's fluency is so poor that it interferes with meaning or the overall impact of the narrative.

The Study

Rationale

Although the research has adequately illuminated issues of dysfluency and complexity in speech, a question remains as to whether the type of oral communication affects fluency and syntactic complexity. Unless a speaker is making a presentation, monologues often tend to have fewer pragmatic and situational factors than dialogues. Thus, does the type of interaction greatly affect fluency? Secondly, a better understanding and description of the fluency of individuals with various TOEIC scores is needed so that companies can know whether or not potential employees have the English-language ability to speak, respond, and negotiate with customers. In short, how can the fluency of speakers (in this study, with a TOEIC score between 683 and 793) be described in regard to both fluency and dysfluency indicators as well as syntactic complexity as measured by T-units? For the purposes of this study, fluency indicators include: (a) rate of speech, fluency rates A and B, as identified by Wendel (1997); (b) pause duration, placement, and frequency; (c) syntactic complexity; (d) ARs; and (e) MLRs. In addition, the prevalence of dysfluency indicators such as false starts, repetitions, reformulations, and percentages of silence was also examined.

Research Questions

1. Are there any statistically significant differences in the fluency indicators of fluency rates A and B, pause durations, pause frequency, and MLRs in the monologues and dialogues of the 10 participants?
2. Is the prevalence of dysfluency indicators such as false starts, repetitions, rephrasing, high percentages of silence, and filled pauses similar among the 10 participants?
3. Are there any statistically significant differences in syntactic complexity scores between monologues and dialogues among the 10 participants?

The hypotheses were as follows:

- (H1) There will be no significant differences in monologues and dialogues in regard to the stated fluency indicators;
- (H2) there will be no similarity in regard to dysfluency indicators; and
- (H3) there will be no significant differences for syntactic complexity scores for the two types of interactions.

Terminology

Speaking rate focused on fluency rates A and B as defined by Wendel (1997). *Fluency rate A* was calculated as the total number of syllables spoken divided by the number of seconds used to complete the task; this was then multiplied by 60. *Fluency rate B* (often referred to as the speaking rate) referred to the total number of *meaningful* syllables (words or sounds that had not been repeated, reformulated, or replaced) spoken, divided by the number of seconds used in the task, multiplied by 60.

Using the differential between the two rates provided a better understanding of an individual's dysfluency. *AR* was computed by dividing the number of syllables by the cumulative time talking (in seconds) after deducting the amount for pausing. The AR was calculated based on a group of runs with a minimum of 30 syllables (Miller, Grosjean, & Lomanto, 1984, pp. 218-219). *MLRs* involved the number of syllables uttered until the speaker stopped talking or paused. *Pauses* were defined as any silence lasting one second or more; those less than one second were counted as micropauses. In verbal encounters common to all language users (mainly conversations and discussions) pauses and hesitations are normal features of interaction. However, silence is often seen as a sign of dysfluency, especially in foreign-language speech where it may be perceived as signaling poor functioning of mental processes. The T-unit (T) is used to measure the overall syntactic complexity of both speech and writing samples; it consists of a main clause plus all subordinate clauses and nonclausal structures that are attached to it.

Participants

The participants included nine Japanese speakers and one Chinese speaker, all of whom had a TOEIC score within the range of 683-793. The average score of these participants was 751.8 (SD 40.6) and their ages ranged from 19 to 25 years. The participants were from two universities, a women's university and a national university for engineering students. Two of the 10 participants were graduate students. The others were undergraduates. All participants were videotaped. The participants all gave permission for the data from the tapes and transcripts to be used for the purposes of the study. See Table 1 for basic data concerning the participants.

Table 1. Participant Data

Participant	TOEIC score	Age	Gender
SS	780	22	F
YS	725	21	F
MO	810	21	F
TK	760	22	F
MO	785	19	F
SY	765	21	F
MJ	729	20	F
WM	790	24	M
HO	685	22	M
YK	690	25	F

Interviews

The videotaped interviews were based on four kinds of discourse tasks: monologues, dialogues, structured interviews, and summaries. The first section involved a self-introduction monologue to measure fluency during unprompted speech and the individual's ability to introduce and describe him or herself.

The second section, an open-ended dialogue, involved prompts concerning the speakers' families, friends, schedules, hobbies, and whether or not they liked school. Thus, the comments and questions in this section were worded so as to have the speaker expand on the information that was previously stated in the self-introduction. The purpose of this section was to measure fluency during interactive speech, and to see how quickly the speaker could ask and answer questions, provide opinions, and fill in details and information.

The third section was in the form of a structured interview consisting of five questions, some of which had two parts and some of which asked for complex or abstract information. These questions were chosen as they might prove slightly difficult to answer. The interview covered topics such as how the speakers had studied and were still studying English, their weekly schedules, their dreams, and how the current year differed from the previous year. This section was designed to measure fluency and syntactic complexity related to various verb tenses that could arise in an interview format. The grammar in this section involved present and present continuous tenses, past tense, past perfect tense, and future tense.

For the fourth part, the individuals were asked to summarize three fables (Long, 2008) that they had read just before the interview. This task focused on whether the participants' fluency was impacted by having to adequately summarize information.

Data Collection

Fifty-four interviews were conducted; there were five overall levels of TOEIC score ranges: (a) 350-460, (b) 461-571, (c) 572-682, (d) 683-793, and (e) 794-895+. A sixth level was native speakers of English. These categories were formed starting with the lowest relevant TOEIC score (from the TOEIC Bridge test) and using 110 points as the range for each category. The fourth level (score of 683-793) seemed to represent an average range of fluency and was selected for this study. Videotaped interviews were conducted from June 2012 to June 2013, each interview containing two sets of monologues and dialogues. Specifically, each interview involved four parts: (a) a self-introduction monologue, (b) an open-ended dialogue based on the student's background, (c) a structured interview, and (d) a reading recall (of previously read material that was given just before the interview). Students did not know of the contents or questions beforehand. Students were paid for their interviews. Coding of the transcripts reflected Conversational Analysis conventions (see Appendix A). For examples of two transcripts, see Appendix B.

Data Analysis

Data related to fluency and dysfluency were analyzed with Excel and the statistical software WINKS-SDA 7 (1996-2012). As for the third research question, in order to avoid reliability issues related to manual coding of syntactic complexity, a web-based L2 syntactic complexity analyzer (Lu, 2011) was used as it counts the frequency of nine grammatical structures and computes 14 indices of syntactic complexity. Results are then presented in 12 indices. For the purpose of this study, data focused on included word count, sentences, verb phrases, clauses, Ts, dependent clauses, complex Ts, coordinate phrases, mean length of sentences, clauses per T (CT), T per sentence (T/S), and complex T ratio (CT/T). One limitation of the software is that oral communication has to be transcribed, and so the placement of periods is often subjective. This can affect sentence and T lengths, thus generating variation and misleading results. In addition, in the tokenization process, the software separates contracted forms, such as *I'd*, *can't*, and *wasn't*, into two tokens and counts each as a word. A Mann-Whitney U test was used to analyze these data as it is the nonparametric statistical test most frequently used to assess whether two independent groups are significantly different from each other.

Results

Results indicated that fluency indicators were highest for the self-introductory monologues, followed by dialogues, structured interviews, and summaries.

Research Question 1

As for whether or not there were any significant differences in the fluency indicators in the monologues and dialogues of the participants, the data for these variables are shown in Table 2.

In testing for possible differences ($p = .05$) in the results for the two types of monologues and dialogues, T-tests indicated that the only variable with significant differences was fluency rate A ($p = .072$) that included the speaker's dysfluencies. There was a significant effect for pause frequencies, $t(16) = 2.01$, $p < .058$, with more pausing occurring in dialogues, most likely due to responding to unexpected questions and comments. These data show that fluency indicators are fairly consistent for both modes of verbal interaction except for both fluency rates A and B and pause frequencies. Both fluency rates were higher in dialogues, indicating that speakers feel the need to speak faster than they would in monologues. Pause frequencies that were also higher

Table 2. Descriptive Data for Two Types of Monologues and Dialogues, $N = 10$

Variables	Monologue 1		Monologue 2		Dialogue 1		Dialogue 2	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Time talking	271.35	84.5	219.0	80.5	197.59	85.7	271.35	84.5
Pause durations (amount of silence)	51.7	39.2	79.5	33.7	47.5	27.9	81.7	34.0
Fluency rate A	94.4	17.3	62.7	20.4	91.5	19.6	84.46	20.3
Fluency rate B	89.5	18.4	57.1	19.0	85.1	19.1	79.3	22.6
Mean length runs (MLRs)	13.55	5.04	8.67	3.8	10.47	5.7	11.2	6.9
Pause frequencies	24.3	17.3	27.2	10.1	31.3	11.7	36.5	11.0

in dialogues indicated that speakers needed more time to react to various cognitive, linguistic, or pragmatic factors. Thus, the first hypothesis was rejected for these two indicators.

Research Question 2

To investigate the issue of similar dysfluency indicators among the 10 participants, averages and totals were calculated for each participant. False starts and rephrasing were the two categories that had the fewest occurrences, with each participant having one or two false starts in the four interactions, and with rephrasing occurring only four times. This indicated that the participants at this level were not necessarily conscious of self-editing their speech. ARs showed the greatest consistency, ranging only from 1.2 to 2.2, whereas micropauses showed the most differentiation, ranging from a low of 3 to a high of 33.7, averaging 12.4 per transcript. Likewise, percentage of silence ranged from a low of 16.0% to a high of 44.7%, but most percentages tended to be in the mid-20s and 30s, the average being 30.7%, which was still rather high (see Table 3 for data on dysfluency indicators). Thus, the second hypothesis, that no similarity would be found in regard to dysfluency indicators, was also rejected.

Research Question 3

Based on the descriptive data in Table 4, it can be further concluded that in dialogues there was an increase in the number of different words, word count, sentences, clauses, Ts, and complex Ts; however, mean sentence length in open-ended dialogues was a little shorter than in other discourse formats.

Table 3. Dysfluency Indicators for the 10 Participants, Totals or Means

Participant	False starts [total]	Rephrasings [total]	Articulation rate [mean]	Meaningless syllables (filled pauses, repetitions) [mean]	Percent of silence [mean]	Micropauses [mean]
TK	0	0	1.9	7	34.8	8.7
YS	2	1	1.9	19.2	44.7	14.7
MO	0	0	1.5	12.2	19.4	3
SS	0	0	1.2	12	36.3	8.5
MO	1	0	1.5	12.2	37.6	10.7
SY	2	1	1.7	37.5	22.6	19.7
MJ	2	0	2.2	14.5	38.4	9.5
WM	2	0	2.2	15	32.8	12.2
HO	2	1	2.1	33	16.0	33.7
YK	1	4	1.7	45	25.1	14.5

Note. Each participant took part in four interactions, two dialogues and two monologues. Articulation rate = words per second.

Table 4. Syntactic Complexity Descriptive Data, Averages (N = 10)

Factor	Monologues		Dialogues	
	Self introduction	Summariz-ing	Open-ended	Structured interview
Word count	85.4	67.7	156.7	166.4
Sentences	6.8	5.4	19.2	14.1
Verb phrases	13.7	11.2	25.2	27.5
Clauses	11.1	9.3	21.2	21.4
T-units	9.5	4.4	18.8	14.5
Dependent clauses	1.7	3.7	5.3	6.6
Complex T-units	1.5	1.8	4.2	4
Coordinate phrases	2.5	3	4.2	4
Mean length of sentences	12.29	15.77	8.45	12.56
Mean length of T-units	8.91	12.04	8.42	11.72
Clauses per T-unit	1.18	2.07	1.118	1.32
T-units per sentence	1.47	0.985	0.868	1.04
Complex T-unit ratio	0.11	0.336	0.192	0.31

The data indicate that the dialogues provided enough lexical input for participants to immediately use, which then helped them sustain more complex structures in their responses. To determine if there was any significant variation in syntactic complexity scores for both monologues and dialogues, a nonparametric Mann-Whitney U test was used to determine differences in complexity; the result was $U = 1366.6$, $Z = 4.950$, and $p = .000$, sig $\leq .01$, 2-tailed. Thus, for syntactic complexity, it can be concluded that there was a statistically significant difference between monologues and dialogues. Dialogues seemed to provide EFL students a means to incorporate entire sentences, noun and verb phrases, and dependent clauses into their own speech, which resulted in longer MLRs. The third hypothesis concerning a lack of significant differences for these two types of interactions was also rejected. This indicates that in this TOEIC score range the complexity of speech was very similar.

Discussion

These data indicate that fluency did differ between the two types of interactions, dialogues and monologues. The other variables of time talking, pause durations, fluency rates A and B, and MLRs were similar. This indicates that the only important differences were in the number of pauses, which further indicates that fluency was slightly more of a challenge in dialogues than in monologues. However, the fact that over 25% of the time talking was spent in silence is problematic, yet the silences might go unnoticed if pauses do not exceed 3 seconds or are at the end of sentence boundaries. Although the speaking rate was rather slow (native speakers average 178.8), it was not a factor that impacted comprehension.

In considering the indicators of dysfluency, it was apparent that speakers in this TOEIC-score range should focus on decreasing the amount of pausing and increasing their AR. Syntactic complexity was shown to be far more evident in dialogues than in monologues, reflecting that speakers had to incorporate verb and noun phrases and clauses that were asked into their own answers along with new vocabulary. In short, although it is important that students have a large vocabulary and

a firm grasp on grammar, syntax, and pronunciation, people are also judged by how they express their ideas.

From these data, it was clear that in comparing monologues and dialogues, there were similar fluency indicators in regard to time talking, pause durations, and MLRs; however, dysfluency still seemed to be a serious issue as indicated by the low ARs (1.2 to 2.2) and high percentages of silence (30%). At issue is how to best describe an individual's fluency. Riggenbach (1991) found that frequent unfilled pauses and a slow speaking rate were two of the most salient features indicating low fluency, yet judgments regarding fluency did not hinge only on fewer pauses and a faster rate of speech. Towell et al. (1996) found that increased MLR was the most important factor in their learners' increased fluency. For native speakers there was a slightly higher MLR, a high percentage of silence, and a comparatively short amount of time talking. Thus, is a native speaker or EFL learner with high AR, MLRs, fluency rates, and complex syntactic complexity still deemed fluent if he or she has a high level of dysfluency indicators? The answer lies in the *effect* and *impression* that the speech has on the listener. It is clear, however, that it would be difficult for a speaker to be deemed fluent if most of the fluency indicators are low and there are several incidences of dysfluency.

Implications

For EFL teachers who are pressed for simplifying proficiency into particular scores, fluency seems to be almost too complex to reduce to one number. It seems clear that EFL teachers should identify one issue to address with their students: (a) production (time talking, MLRs), (b) speaking rate (AR, fluency differential [fluency rate A - fluency rate B]), or (c) lexical or syntactic complexity. Nonetheless, providing indicators as ratios for both fluency and dysfluency might be helpful. As for the process of transcription, most teachers are too busy to record entire interactions, but transcribing four or five random speech samples (of 60 seconds or more) provides enough data to make tentative recommendations. In improving fluency in the classroom, techniques like shadowing, timed readings (read aloud), and timed interactions (wherein EFL learners are given less time to express themselves) are effective. Timed-structured gambits (conversations in which a speaker's comments and questions are read aloud so that another student's replies can be recorded; see Long, 2012, 2013) are helpful in that pragmatic elements are often included in the scenario. Teachers should also help students focus on semantic-extension tasks, in which students read aloud a series of phrases that become increasingly longer and more lexically and syntactically complex. Videotaped recordings, which students can later watch to analyze their performances, are important as well.

Conclusion

The results of this study lead to more questions. With more interviews and data collected from participants with other TOEIC scores, what are the most important differences among the various TOEIC score ranges? At which particular range is a person really fluent enough to engage in real-world negotiation and interaction effectively? Should teachers at lower or higher levels focus more on ARs than on pausing? Further research should also explore how participant confidence factors into fluency. Do non-Japanese foreign students (in the same TOEIC score range) have similar issues in fluency? Lastly, cross-gender fluency needs to be further examined.

What is apparent from these data is that these participants with relatively high TOEIC scores (which would allow many a chance at international jobs or jobs in which English is used) are not really ready to engage in rapid and complicated interactions and negotiations. Although there are confounding variables such as personality, motivation, and study abroad experience

that could impact fluency, it is clear that TOEIC scores can often provide unreliable and misleading information concerning the student's productive skills in speaking and in writing.

Finally, the understanding and analyzing of fluency are difficult research avenues for teachers. Software for speech recognition, such as Dragon Dictate 2.0 and Macintosh's (Maverick) OS Dictate 10.5.8, are problematic, and manual transcription is still the most reliable method of recording verbal interactions. However, it is clear that the time has come for teachers to place less emphasis on test scores that reflect passive skills such as reading and listening and to focus on productive skills such as speaking and writing.

Bio Data

Robert Long is an Associate Professor at Kyushu Institute of Technology. Besides being on the *TLT* Editorial Advisory Board and active with JALT, he has been working in areas of fluency, educational curriculum, intercultural programs, pragmatics, and discourse analysis, and has published several EFL and ESL textbooks.

Acknowledgements

This research is supported by the Grant-in-Aid for Scientific Research (Kakenhi) of the Ministry of Education, Culture, Sports, Science, and Technology in Japan (No. 24520626).

References

Brumfit, C. J. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.

Chambers, F. (1997). What do we mean by fluency? *System, 25*, 535-544.

Dragon Dictate 2.0: Speech recognition for MAC OS X. [Computer software and manual]. Singapore: Nuance Communications.

Ellis, R., & Barkhuizen, G. (2005). Analysing accuracy, complexity and fluency. In R. Ellis & G. Barkhuizen (Eds.), *Analysing learner language* (pp. 139-164). Oxford: Oxford University Press.

Ferreira, F., Lau, E., & Bailey, K. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science, 28*, 721-749.

Fillmore, C. J. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S. Wang (Eds.), *Individual differences in language ability and language behavior* (pp. 85-101). New York: Academic Press.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*, 354-375.

Givon, T. (1991). Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language, 15*, 335-370.

Griffiths, R. (1990). Pausology and listening comprehension: Theory, research, and practice. *JALT Journal, 12*, 99-120.

Koopmans-van Beinum, F. J., & van Donzel, M. E. (1996). Discourse structure and its influence on local speech rate. *Proceedings of the Institute of Phonetic Sciences, 20*, 1-11. Amsterdam: Institute of Phonetic Sciences.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387-417. http://dx.doi.org/10.1111/j.1467-1770. 1990.tb00669.x

Lewin, M., McNeil, D., & Lipson, J. (1996). Enduring without avoiding: Pauses and verbal dysfluencies in public speaking fear. *Journal of Psychopathology and Behavioral Assessment, 18*, 387-402.

Long, R. (2008). *New Aesop fables for children: Vol. 1*. Raleigh, NC: Lulu Press.

Long, R. (2012). *Reaching out*. Nagoya, Japan: Perceptia Press.

Long, R. (2013). *Stepping out*. Nagoya, Japan: Perceptia Press.

Lu, X. (2011). The L2 syntactical complexity analyzer. [Computer software and manual]. University Park: Pennsylvania State University.

Miller, J., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A re-analysis and some implications. *Phonetica, 41*, 215-225.

Norris, J., & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 716-761). London: Blackwell.

Macintosh OSX 10.5.8. (2013). Cupertino, CA: Apple.

O'Loughlin, K. (1995). Lexical density in candidate output on direct and semi-direct versions of an oral proficiency test, *Language Testing, 12*, 217-237.

Riggenbach, H. (1991). Towards an understanding of fluency: A micro-analysis of nonnative speaker conversations. *Discourse Analysis, 14*, 423-443.

Szmrecsányi, B. (2004). On operationalizing syntactical complexity. *JADT 2004: 7es Journées Internationales d'Analyse statistique des Données Textuelles.* Retrieved from http://www.benszm.net/omnibuslit/Szmrecsanyi2004.pdf

Skehan, P. (1996). Second-language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17-30). Oxford: Heinemann.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*, 84-115.

Wendel, J. (1997). *Planning and second language narrative production* (Unpublished doctoral dissertation). Temple University, Japan: Tokyo.

WINKS-SDA 7 statistical software. (1996-2012). [Computer software]. Cedar Hill, TX: TexasSoft.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics, 24*, 1-27.

Appendix A

CA Transcription Symbols

Manner/Quality

smile quality	£
exhale / inhale	hhh
vocalism	(sniffle)
click	.t
laugh pulse	heh
laughing word	wo(h)rd
laughter	heh heh
low pitch	↓
high pitch	↑
pause, timed	(1.2)
pause, short	(.)
lag (prosodic length / elongated sound)	
unintelligible	( )
uncertain	(word)
emphatic tone	!
interviewer comment	[[ ]]

Appendix B

Examples of Self-Introduction Monologues of Participants with TOEIC Scores of 683-793

In the example below, one can note the problems of oral communication in the language of two young Japanese women with fairly high TOEIC scores of 780 and 785, (their names have been abbreviated) who are introducing themselves.

TOEIC Score 780	Start time: 00.04
My name is S. S. and I'm (.) nineteen years old. (1.6) And I'm a second year student at the university of Kitakyushu. (1.7) And I'm from Fukui and there five members in my family. (1.1) And I have, (1.4) one brother and one sister. (1.6) And my: (.) hobby is listening to music (.) and I like singing and I often go shopping in my free time.	End time: 0:49 Total time: (45 seconds) Amount of silence: 7.4 Percentage of silence: 15.1% Average mean length run: 14.5 (87 syllables) (87 meaningful syllables) Articulation rate: 2.3 Fluency rate A: 115.9 Fluency rate B: 115.9 Micropauses: 4

TOEIC Score 785	Start time: 0:17
My name is M. O. and I am from Amami island (.) Kagoshima prefecture and I like making cakes and: (1.8) uh I want to to become (.) pastry cook before and then: (.) I like studying English (1.8) mmm: (1.0) and I will study abroad next year so (.) I (1.5) I do my best now, heh and (.) mmm: (3.4) mmm: (2.5) mmm: (3.9) um: (.) heh (6.1) mmm: heh.	End time: 1:23 Total time of interviewee speaking: 66 seconds Amount of silence: 22.0 Percentage of silence: 33.3% Average mean length run: (78 syllables) (71 meaningful syllables) Articulation rate: 1.7 Fluency rate A: 70.9 Fluency rate B: 64.5 Micropauses: 6