Modern English Education, Vol.12, No. 4, Winter 2011

Looking at Read Out Loud Rate as an Indicator of English Speaking Proficiency for EFL Students

Youngsoo Song
Korea University

Song, Youngsoo. (2011). Looking at read out loud rate as an indicator of speaking English proficiency for EFL Students. *Modern English Education, 12*(4), 149-170.

It is important for instructors and raters to maintain objectivity and accuracy when assigning scores for EFL students. With this in mind, this study aims to determine if read out loud rates in words per minute, or WPM, spontaneous speech rates and degrees of chunking are significant indicators of speaking proficiency. Participants were Korean EFL undergraduate students who were instructed to read out loud three passages with the addition of a one-minute spontaneous speech. Intentional grammatical errors were inserted into one of the read out loud texts to discourage chunking. The spontaneous speeches were rated by three native English speakers. Quantitative analysis was performed by using CohMetrix, WordSmith, correlation, and regression to examine the relationship between participants' speaking proficiency scores, speech rates, and the employment of chunking. The results revealed that there was a significantly positive correlation between WPM value of spontaneous speech and spontaneous speech scores. Further analysis revealed that Adjusted R Square was .17 with a beta value of .41 when the WPM of spontaneous speech was regressed on the ratings of spontaneous speech. The study's implication is that raters and instructors should take the WPM of spontaneous speech into consideration before assigning a grade.

**[speech rate/language testing/read out loud/objective rating/chunking/
말하는속도/언어평가/소리내읽기/객관적평가/뭉칫말화]**

I. INTRODUCTION

1. Introduction to Testing, Speech Rate and Speaking Proficiency

Testing has historically been an intricate part of human society for many purposes, from measuring physique to language proficiency. Nowadays, tests are common in all fields of

studies and in the workforce to select the qualified and weed out the rest. In fact, academic institutions, workplaces and other formal and informal organizations use testing to obtain diagnostic performance and placement information (Bachman & Palmer, 1996). Consequently, test results can be used to determine an individual's work position, academic placement or social status. Given the historical importance of testing, it is unsurprising that many prospective students and candidates study relentlessly to obtain the highest scores possible. In settings where English is a foreign language such as Korea, students and employees are focused on obtaining high score on English proficiency exams such as TOEFL, TOEIC, and IELTS. With a constant overflow of applicants and few positions available, it is important for administrators to employ valid and reliable tests to accurately measure the appropriateness of an applicant for the relevant field. In addition, it is crucial to have well trained raters to provide accurate and reliable scores based on test responses; it is a two-part requirement that must be fulfilled for any test to be a useful and predictive tool. Therefore, it is imperative to seek and fine-tune new test methods and if necessary, modify already established tests and rubrics.

The goal of this study is to determine whether speech rate, also known as words per minute (WPM), will carry any significance when determining speaking proficiency. This study uses a speed-based indicator such as WPM as a key variable in determining speaking proficiency due, in part, to the following factors. First, EFL students at lower proficiency levels tend to be slower speakers due to smaller chunking. In other words, without the ability to chunk multiple words, phrases or clauses, speech rate is bound to be relatively low. Second, EFL students at lower proficiency levels tend to have undeveloped self-monitoring skills, resulting in slower speech. This is because undeveloped self-monitoring skills result in slower production of utterances and slower speed of self-correction. Ultimately, these factors will result in slower speech rates. Third, when it comes to reading out loud, EFL students who possess higher English proficiency also tend to possess faster reading speed. This allows them to quickly speak what they read given that their psychomotor skills are equally developed as well.

Then, it is possible to view speech rate as an important factor of speaking proficiency because speech rate is influenced by chunking, level of cognition, psychomotor skills, and monitoring. In other words, one can assume that a student who can speak relatively quickly can do so because he/she is presumably not hindered by any of the aforementioned factors (Bachman & Palmer, 1996). If so, if test takers were asked to read out loud a relatively easy passage, the rate of WPM should be higher because there would be no difficult vocabulary or complex chunking to process. In contrast, if the same test takers were to read out loud a difficult passage, the rate of WPM should be lower due to difficult vocabularies and inhibited chunking. Because similar principles apply to spontaneous speeches, an individual's rate of reading out loud and spontaneously speaking is believed to be similar

as well. The reason for using read out loud is due to its use and practicality. Although read out loud is not as common as spontaneous speeches, it is used in the workforce such as by announcers, public speakers, news anchors and more. In terms of practicality, evaluating spontaneous speeches can be more time consuming, costly and subjective because test takers' responses are open-ended, unrestricted and often unstandardized. Based on Bachman and Palmer's model (1996) which indicates speed as fluency, and therefore speaking proficiency, it would be a good idea to explore WPM in terms of both read out loud and spontaneous speech to determine their relationship to one another and to speaking proficiency. It is important to note that a person with fast speech rate does not necessarily make him/her a better communicator (high speaking proficiency). The study is merely exploring to see if those with high speaking proficiencies tend to speak faster than those with lower speaking proficiencies. It is also suspected that WPM is a good indicative variable, not an absolute determining variable.

This research subject carries a great value in the field of English language education because of its implications. If WPM has a significant relationship with speaking proficiency, raters can use WPM as a reference to give more objective grades and English educators can use WPM as an index to track EFL students' speaking proficiencies over time.

2. Purpose of Study

Ultimately, the purpose of this study is to see if there is a significant relationship between WPM of read out loud and spontaneous speech, and speaking proficiency. In addition, the effect of interrupted chunking on WPM, followed by speaking proficiency, will be investigated as well.

3. Research Question

Given the importance of WPM and speaking proficiency, the research questions are as follows:

1. For EFL students, does speaking proficiency, based on ratings, have any significant correlations with the rate of read out loud tasks and/or the rate of spontaneous speech?
2. For EFL students, do intentional grammatical errors affect chunking by affecting the value of WPM?

II. LITERATURE REVIEW

One of the goals for many English learners is to strive for excellence and reach a certain level of mastery in spontaneous speech (Leather, 2000). In order to measure this, it is necessary to develop proper test items and consider its test method facet, usefulness and characteristics (Bachman & Palmer, 1996). Although read out loud may seem like a simple task, it can still provide insightful information about the test takers' speaking competence. Not only does the task involve speaking, it involves reading as well, which has significant utility in integrated skills testing but not as significant for discrete skills testing. Testing the usefulness or the efficacy of reading out loud as well as comparing the results of it to spontaneous speaking should reveal if they are indeed good candidates to be used in EFL English testing for speaking proficiency.

In this study, Task 1 is a read out loud task with an easy level passage. Task 2 is the as Task 1, but with a relatively more difficult passage. Task 3 is similar to Task 1, but with a different passage with intentional grammatical and discourse errors to break chunking. Task 4 is a one-minute spontaneous speech without any passages.

1. Speaking Proficiency in Terms of Testing

According to Bachman and Palmer (1996), language knowledge can be broken down to organizational knowledge and pragmatic knowledge. The former can be further broken down to grammatical knowledge (how individual utterances are organized) such as vocabulary, syntax and phonology, and textual knowledge (how utterances are formed to make sense) such as cohesion and conversational organization. The latter can be further broken down to functional knowledge (how utterances are related to the communicative goals) such as creativity and control, and sociolinguistic knowledge (how utterances are appropriate and fitting to a given situation) such as dialects, cultural references and idiomatic expressions. Based on this model, spontaneous speech (Task 4) is a task that covers all attributes of language knowledge. Spontaneous speech is also a task that is suitable as a proficiency test because it meets the criteria of what a speaking test should evaluate: topical knowledge, psychomotor skills and strategic competence (Brown, 2007).

Read out loud (Tasks 1 and 2) may not cover all aspects of language knowledge, but by looking at the speed of spontaneous speech, if highly correlated with human rating and the speed of reading out loud, it may be an indicative task of speaking proficiency. Since WPM rate of read out loud and spontaneous speech fulfills the basic requirements of Bachman's framework of language competence, it would be good candidates to be used in language testing to help determine speaking competence.

However, there are other components to evaluate a person's speaking proficiency. For example, intelligibility is important when it comes to oral communication. In most cases, speakers with poor intelligibility would be considered to have low speaking proficiency. However, there is no universally accepted definition of intelligibility, nor is there a consensus on the best method to measure it (Isaacs, 2008). A similar problem persists in other components such as one's level of discourse, socio-pragmatics, organization, and creativity; all of these are subjective and no standards have been set. This is why something more objective and operationalized as speech rate (Tasks 1 through 4) and its relationship to speaking proficiency should be studied in order to determine what kind of relationship they have and if speech rate is a good indicative variable to use for assessing one's speaking proficiency.

According to Bachman (1990), the range of natural speech rate is from 160 to 190 words per minute and 160 to 190 (with alpha) for reading. In other words, test takers who read out loud or spontaneously speak should reach a WPM rate of 160-190 to be considered competent English speakers (Biancarosa, 2005; Gates, 1921).

2. Read Out Loud and Other Speaking Proficiency Tests

The use of read out loud in a language test is not uncommon. The Pearson company currently employs read out loud in Pearson Test of English Academic (PTE-A) where test takers are required to read out loud in a short time. This test item is reported to measure the test taker's pronunciation and fluency (Longman, 2009). There are many tests with speaking components such as TOEFL, TOEIC, and IELTS. However, these tests use different methods such as mock interviews, dialogues, listen and speak, read and speak, and more to measure one's English speaking proficiency (Educational Testing Service, 2008; IELTS, 2010). Ideally, real interviews and dialogues may allow evaluators to accurate gauge the test taker's speaking proficiency, but due to practicality issues such as time, cost and logistics, high-stakes tests simply cannot use one-on-one methods.

In terms of practicality, read out loud and spontaneous speech are good candidates for language testing, especially with the continuous advancement of natural language processing (NLP) and automated speech recognition (ASR) technology. It is possible to implement the technology to automatically grade a test taker's spoken output for read out loud and spontaneous speech (Chapelle & Chung, 2010). If this is established, time and cost for evaluation can be greatly reduced. However, ASR must be advanced enough to possess high validity, measuring what it purports to measure, and high reliability (low disparity between multiple measurements). In time, many researchers believe that the same technology can be used to accurately assess spontaneous speech for accuracy, discourse, fluency, socio-pragmatics and task completion to accurately reflect speaking proficiency.

3. Similar Studies and Chunking

Studies relevant to read out loud and speaking proficiency have been conducted in the past. One study was performed by Coniam (1990) where the relationship of fluency and read out loud was explored. The author stated that there was a common sense notion that generally, higher fluency resulted in faster rate of speech and higher proficiency. The result of the study supported Coniam's statement, but did not apply to writing and reading.

Another similar study was done where participants' read out loud task and story retelling (similar to spontaneous speech) were analyzed for speech rate, pauses, length, and frequency. The results showed that in general, increased speech rate and phrasal length were the results of higher English proficiency (Kowal, O'Connell, O'Brien & Bryant, 1975). Therefore, WPM, as well as the length of chunks, will be examined in this study to determine what significant relationship they have with speaking proficiency. It is important to note that chunking is an important factor that may affect WPM of both tasks of read out loud and spontaneous speech. Language consists of thousands of "chunks" which serve particular functions such as extended discourse, fluidity and speed (Ellis, 1998, 2003; Pawley & Syder, 1983; Sinclair, 1991). One of the reasons for investigating chunking and speed for speaking competence is that memorized chunks serve as a basis for the creative construction of discourse (Taguchi, 2007). Also, students who practice chunking, and thus have better chunking, may perform better in reading which is a required skill for reading out loud (Kyunghee Choi, 2011). In other words, test takers who possess short or small chunking are less likely to produce lower WPM values and lower ratings for speaking competence.

Based on the similar studies, it is not unwarranted to explore the second research question. If test takers are given read out loud tasks with different difficulties, their WPM's should be different due to chunking. Advanced students will likely to have higher WPM regardless of difficulty if they are proficient speakers whereas novice students may perform well on the easy read out loud task but struggle with the difficult one. Based on this premise, the study hypothesizes that if grammatical errors are introduced to an easy read out loud task, novice students are likely to suffer less in terms of WPM because their ability to chunk is not fully developed in the first place (Task 3). However, for advanced students who possess better chunking abilities, the disparity between the WPM values obtained from each task is expected to be greater than those of novice students. Furthermore, the study also aims to examine if the magnitude of WPM disparity is an indicator of speaking proficiency by performing a correlation analysis (WPM disparity and ratings). If the result of this study supports that there is a connection between chunking and speed, not only would the finding contribute to language testing, but may yield pedagogical implications to increase English proficiency for EFL students and teachers (Nahk-Bohk

Kim, 2011).

III. METHODS

1. Participants

Participants were 30 Korean EFL students, 9 males and 21 females, who were either enrolled in English conversation courses at a Korean university or completed their bachelor's degree. Each participant was informed about the purpose of this research. As shown in Table 1, demographic information such as gender, age, fluent language(s) they speak, years of experience in an English speaking country, major and the number of years spent learning English, and the number of years in university was collected (refer to Appendix A).

Table 1

Demographic Information Collected from EFL Participants

	Minimum	Maximum	Mean	S.D.
Age	20	36	24.07	3.90
Number of Years Spent Learning English	7	25	13.53	4.31
Number of Years Spent in University	2	8	3.40	1.81
Experience Abroad in Years	0	15	1.63	3.56

Twenty-six participants were majoring in English language education and the remaining 4 participants were majoring in math, psychology, general education and home economics. Twenty-six participants spoke English and Korean and the remaining 4 participants spoke additional languages such as French, Chinese and Japanese.

This research classified the participants as EFL leaners from the fact that this study was carried out in Korea where English not a second language (official language), but rather a foreign language. Also, the demographic information of the participants revealed that they would be better classified as EFL students rather than ESL because of their exposure level to English. It is noteworthy to mention that one participant (15 for Number of Years Overseas) claimed to have lived abroad for many years. After a brief dialogue, the participant notified the researcher that she worked overseas in a Korean speaking environment and socialized with other Koreans during her leisure time. In addition, she

reported that she has travelled back to Korea on a regular basis and therefore felt like an EFL student rather than an ESL student. After a detailed interview, the researcher concluded that she would better fit as an EFL student rather than an ESL student. Due to these circumstances, the study will only focus on EFL students rather than ESL students, and the outcome of the study will only be representative of EFL students.

2. Tasks 1 to 4

The excerpts for the tasks were adopted from various sources. Task 1 was extracted from an online essay[1], Task 2 was extracted from an excerpt from a sample GRE exam on the internet[2] and the material for Task 3 was an excerpt taken from a sample English 10 Provincial Exam[3].

Task 1 had an easy-to-read excerpt whereas Task 2 was relatively difficult. The difficulty of Task 3 was similar to that of Task 1, but intentional grammatical errors were added to Task 3. Since the focus of Task 3 was chunking, the researcher, a native English speaker, selectively inserted discourse and mechanical errors to break the natural flow of chunking. For control, lexical density (TTR), sentence complexity (Words per Sentence), length (Number of Words) and readability (Flesch Reading Ease) were used to determine each passage's level of difficulty by using Coh-Metrix 2.1 and WordSmith 5.0. The indices are shown in Table 2 below. For Task 4, the participants were presented with several topic choices to produce a spontaneous one minute speech. The participants were given approximately a minute to brainstorm and outline their ideas on a sheet of paper. They were encouraged to refer back to their outline to facilitate spontaneous speech. In addition to the researcher's judgment of text difficulty of the three tasks, the aforementioned indices were taken into consideration as well.

Table 2
Indicators of Lexical Density of Tasks 1 through 3

Task	Flesch Reading Ease	TTR	Words per Sentence	Number of Words
1	86.77	55.45	12.94	219
2	34.71	67.31	29.71	209
3	78.28	62.50	14.47	213

Note: For Flesch Reading Ease, lower value indicates higher difficulty of text.

[1] http://www.exampleessays.com/viewpaper/22557.html
[2] http://www.majortests.com/gre/reading_comprehension_test.php?testid=C105
[3] http://www.bced.gov.bc.ca/exams/search/grade10/english/release/exam/0708EN_p.pdf

3. Test Procedure

As shown in Appendix B, the participants were given a set of tasks. The researcher explained the purpose of the study and instructed the participants to read out loud as quickly as possible without compromising intelligibility and naturalness. A digital recorder (Olympus Digital Voice Recorder DS-50) was used to capture the participants' spoken output for each task. For Task 4, a digital watch was used to cue the participants to stop. The participants were also asked to fill out a survey of demographic information. After gathering the data, the audio files from the digital recorder were uploaded to the researcher's computer for analysis. First, any silence or irrelevant materials were edited out. The new audio playback times after editing were used for the calculation of WPM. After inspecting every file, the lengths of time of the audio files (in seconds) were recorded, along with their respective demographic information. As shown in Figure 2, a formula was used to calculate the values of WPM. As indicated in the formula, the number of words was counted using Microsoft Word 2010's Word Count. Next, the obtained value was multiplied by 60 (seconds) and the product was divided by the number of seconds taken to read. Since this study was specifically focused on WPM, the number of syllables and characters was omitted.

*WPM = (Number of Words * 60 seconds) / (Number of Seconds Spent Reading Out Loud)*

Figure 2. Formula used to obtain the value of words per minute (WPM).

4. Analysis Procedure

This study measured the participants' WPM from 4 the different tasks where the first three were read out loud tasks and the last one was a spontaneous speech; Task 1 was to read an low level passage, Task 2 was to read a difficult passage and Task 3 was to read a passage filled with grammatical errors. The difficulty of the passages was determined by several factors such as type-token ratio (TTR), the researcher's judgment, the origin of the excerpts, and readability indices. After gathering the data, the spontaneous speeches were assessed by properly trained raters to evaluate the test takers' speaking proficiency based on TOEFL's speaking test rubrics (refer to Appendix C). With the participants' scores and WPM's, statistical analysis methods such as Repeated Measures ANOVA (to determine whether there were significant differences of WPM among the read out loud tasks), correlation (to determine whether there was a significant correlation between WPM and scores), and regression (to determine whether Task 4 WPM had any predictive power of determining ratings of Task 4) were carried out.

5. Evaluation of One-Minute Speeches (Task 4)

Participants' responses of Task 4 were rated by 3 native English speakers. The ratings were based on the rubric that was adapted from ETS's iBT TOEFL Independent Speaking Rubric. The researcher explained the purpose of the study and the rubric in details to the raters. All questions by the raters were answered by the researcher via e-mail and telephone. A single holistic score was produced for each participant's performance in Task 4. The means of the 3 raters' scores were obtained for the purpose of further data analysis.

IV. ANALYSIS AND RESULTS

1. Descriptive Statistics

As shown in Table 3, the means of WPM obtained from Task 1, 2, 3, and 4 were 175.77 (SD: 28.32), 126.50 (SD: 21.29), 142.35 (SD: 24.35), and 131.85 (SD: 25.66), respectively. As expected, the difficult task (Task 2) yielded the lowest WPM value, with the error-inserted task (Task 3) yielding the second lowest and the easy task (Task 1) yielding the highest. Their standard deviations ranged from approximately 21 to 28. This meant that all students produced similar values of WPM for each task with only a relatively small deviation. Also, because the mean value of Task 3 was in between Task 1 and Task 2, Task 3 could be interpreted as a medium difficulty task. Task 4 could not be generalized in terms of difficulty because the nature of the task was different from the others.

According to Bachman and Palmer (1996), the natural rate of speech ranges from 160 to 190 WPM for fluent English speakers. Noting that the average WPM of spontaneous speaking was 131.85, the result suggests that the participants as a whole fell below the natural speaking rate though some individuals may have achieved that rate. However, this was expected because the participants were not native speakers but EFL students who were still learning English.

Table 3
Descriptive Statistics of WPM for Tasks 1 – 4

Task (WPM)	Minimum	Maximum	Mean	S.D.
1	125.14	238.91	175.77	28.32
2	90.22	169.46	126.50	21.29
3	98.54	198.69	142.35	24.35
4	87.46	181.36	131.85	25.66

The ratings of 3 raters were analyzed as shown in Table 4. The raters' mean values were

3.03, 3.30 and 2.70 ranging from 2 to 4. The relatively low standard deviations showed that the raters gave similar ratings to all participants. It was concluded that the reliability of the ratings was acceptable (Cronbach's Alpha = 0.86).

Table 4

Descriptive Statistics of Ratings for Task 4

Rater	Minimum	Maximum	Mean	S.D.
1	2	4	3.03	.72
2	2	4	3.30	.54
3	2	4	2.70	.65

Note. Cronbach's Alpha = 0.86

2. WPM Values Obtained from Tasks 1 through 3

For Table 5 and 6, Repeated Measures ANOVA was used to determine if there were any significant differences of WPM between Tasks 1, 2, and 3. This statistica l method was employed because the same participants were measured to three diffe rent conditions (Tasks 1 through 3).

In Table 5, Mauchly's test indicated that the assumption of sphericity had been violated, ($\chi^2(30) = 11.75$, $p < .00$), and therefore, degrees of freedom were corrected using Huyh-Feldt estimate of sphericity ($\varepsilon = .78$). Since the violation meant that the variances of the difference among the tasks were significantly different, the average WPM of each task was significantly different from one another.

Table 5

Mauchly's Test of Sphericity

					Epsilon		
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Huynh-Feldt	Lower-bound
WPM	.66	11.75	2	.00	.75	.78	.50

The results of Table 6 showed that there was a significant effect of each task on WPM ($F(1.55, 45.00) = 202.35$, $p = .00$). These results reinforced the previous claim that WPM of each task by the test takers was significantly different from one another.

Table 6

Test of Within-Subject Effects for Tasks 1 – 3 on WPM

Huynh-Feldt	Type III Sum of Squares	df	Mean Square	F	Sig.
WPM	37948.40	1.55	24459.11	202.35	.00
Error	5438.51	44.99	120.87		

A post-hoc test using Bonferroni was conducted for the WPM values of Tasks 1 through 3. As shown in Table 7, the sig. values ranged from .00 to .05. This implied that each participant produced a different set of WPM's across the three tasks with their corresponding lexical difficulties.

Table 7

Post-Hoc Test of Tasks 1 through 3 Using Bonferroni

Task	Mean Difference	Std Error	Sig.
Task 1&2	49.26	6.41	0.00
Task 3&2	15.85	6.41	0.05
Task 1&3	33.41	6.41	0.00

3. Correlation Analysis Results between WPM of Tasks 1-4 and Rating of Task 4

A correlation matrix is shown in Table 8 between Rating of Task 4 and each task's WPM value. As presented below, there were positive significant correlations between each task. For Rating of Task 4, there was one significant correlation: Task 4 WPM ($r = 0.41$, $p < 0.01$). This meant that although all tasks were significantly correlated to one another, only Task 4 (WPM) carried a significant correlation with Rating of Task 4. In other words, only the raters' rating of the spontaneous speech task and none of the read out loud tasks carried any significant relationship with the ratings.

Table 8

Correlation Matrix of Participants' WPM of Tasks 1-4 and Rating of Task 4

Pearson Correlation	Rating of Task 4	Task 1 WPM	Task 2 WPM	Task 3 WPM	Task 4 WPM
Rating of Task 4	1.00				
Task 1 (WPM)	.21	1.00			
Task 2 (WPM)	.26	.83**	1.00		
Task 3 (WPM)	.20	.84**	.93**	1.00	
Task 4 (WPM)	.41**	.62**	.48**	.50**	1.00

Note. *$p < .05$; **$p < .01$

4. Regression of WPM of Task 4 on Rating of Task 4

As shown in Table 9, a regression model summary is presented where the independent variable was Task 4 WPM and the dependent variable Rating of Task 4. No additional tasks were entered in the regression model because previous results showed that none of the tasks were significantly correlated to Rating of Task 4. Since the value of Adjusted R Square was .14 (sig. F change of .03), it can be concluded that the accountability of Task 4 (WPM) for Rating of Task 4 is low.

Table 9

Regression Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df 1	df 2	Sig. F Change
1	.41[a]	.17	.14	.52	.17	5.64	1	28	0.03

a. Predictors: (Constant) and Task 4 (WPM).

In Table 10, it is shown that the WPM value of Task 4 has a beta value of .41 (sig. of .03). The results indicate that in the given regression model, Task 4 (WPM) had a moderate predictive power of Rating of Task 4.

Table 10

Regression Model Coefficients

	Unstandardized Coefficients		Standardized Coefficients		
Model	B	Std. Error	Beta	T	Sig
1 (Constant)	1.83	.51		3.59	.00
Task 4 WPM	.01	.00	.41	2.37	.03

Note. Dependent Variable is Rating of Task 4.

5. Correlation Analysis Results of Rating of Task 4 and WPM Difference between Task 1 and Task 3

Table 11 shows the correlation between Rating of Task 4 and WPM Difference between Tasks 1 and 3. Since the difficulty levels of Task 1 and Task 3 are similar, the difference of WPM should be minimal and similar across all participants. However, with the intentional grammatical errors imposed on Task 3, participants with higher speaking proficiencies—those who take advantage of chunking—should yield much lower WPM values than those of Task 1. Conversely, those with lower speaking proficiency—those who cannot take

advantage of chunking—should have minimal difference of WPM between Task 1 and Task 3. As shown in the table, there was no significant correlation between the two variables. This suggests that the degree of difference between Task 1 and Task 3 is not related to speaking proficiency.

Table 11

Correlation Matrix of Rating of Task 4 and WPM Difference between Tasks 1 and 3

Pearson Correlation	WPM Difference between Tasks 1 and 3	Rating of Task 4
WPM Difference between Task 1 and 3	1.00	
Rating of Task 4	.08	1.00

Note. *$p < .05$; **$p < .01$

V. DISCUSSION AND CONCLUSION

1. Answering the Research Questions

Addressing the first research question, it was found that the WPM values of all tasks except Task 4 were significantly correlated with each other from .48 to .93 at the .01 level; only the WPM value of Task 4 was positively correlated with Rating of Task 4 ($r = .41$, $p < .01$). The implication from this result is that in a testing environment, it may be a good idea to consider WPM of spontaneous speaking as a moderate factor before assigning a grade for speaking proficiency; this is possible with the aid of technology such as a computer equipped a microphone. In fact, it is shown in Table 10 that Task 4 WPM had a beta weight of .41 (sig. of .03) when predicting Rating of Task 4. However, one should not solely rely on WPM of spontaneous speech to determine speaking proficiency since the adjusted R squared value was only .14. In detail, this meant that the model only had 14% accountability for speaking proficiency when Task 4 (WPM) was the only independent variable. As for the read out loud tasks, they should not be used to determine speaking proficiency (ratings) because no significant correlations were shown between the WPM of read out loud tasks and the ratings. It is important to note that speech slowness could have been a result from continuous cognitive processing (thinking about the topic) in Task 4, whereas the slowness in Tasks 1 through 3 may have been from difficult vocabulary. Therefore, a further study should be carried out to distinguish the two factors: difficulty of vocabulary and cognitive processing. Only then would it be safer and more definitive to claim that the WPM values of the read out loud tasks are unrelated to speaking proficiency.

Addressing the second research question, according to the findings in the Analysis and

Results section, interruption of chunking did seem to affect WPM of read out loud tasks. This is supported by the fact that Task 3's WPM had a significantly lower value than Task 1's WPM. However, although the results from Tables 5, 6, and 7 show that test takers produced different WPM's, whether the lower WPM's were truly the results of chunking or other factors cannot be determined from this study alone due to insufficient information. Further study geared towards qualitative research is recommended to find other causes, if any, and to what degree they affect speech rates. Lastly, the result from Table 11 shows that the magnitude of WPM disparity due to chunking is not related to the ratings. However, the result is not unexpected because WPM's of Tasks 1 and 3 were not related to speaking proficiency.

2. Conclusion

The results from various analyses were able to provide some answers to the research questions. WPM of spontaneous speech was the only variable that was significantly correlated to speaking proficiency. In addition, although chunking impacted WPM, it did not have a significant role in indicating speaking proficiency. In conclusion, WPM of spontaneous speaking was the only variable that carried some predictive power and was correlated to speaking proficiency.

1) Implications

To maintain objectivity, raters should consider WPM (a part of fluency) as well as accuracy, discourse, and socio-pragmatics when assessing language competence. The read out loud tasks did not prove to be predictive of one's speaking competence. Given the range of WPM for EFL students' spontaneous speech, test designers can assign appropriate time allotment for speaking tasks; low proficiency test takers are less likely to complete the task and thus receive a lower rating whereas the opposite would be true for high proficiency test takers. Another implication from this study is that in the classroom environment, instructors should keep a longitudinal record of their students' WPM. If the students are improving, their recent WPM should be higher than their initial WPM for spontaneous speaking. Time can be allocated to help students overcome whatever difficulties they may face to improve their WPM. With sufficient practice, students may greatly improve their speaking proficiency and no longer feel apprehensive about having a casual dialogue in English, allowing the students to interact with more English speakers. This implication however, may be controversial and circumstantial. For example, well-funded schools where there are fewer students per classroom, competent teachers and technological equipment, can certainly benefit from the previously mentioned implication;

for under-funded schools, this is not a possibility. In addition to pedagogical implications, future research ideas can be taken from this study as well. Since there is a relationship between WPM of spontaneous speech and proficiency, other variables related to proficiency such as pronunciation, grammar, sentence structure, lexical complexity, discourse, socio-pragmatics and creativity should be analyzed to see their how indicative they are. With an extensive list of variables, a model can be made using multiple regression and/or structural equation modeling. This model should give insight to researcher and educators to know what to look for when evaluating one's speaking proficiency.

2) Limitations

One of the limitations of this study was the low statistical power probably due to the small number of participants. With more participants, the statistical results would potentially carry a stronger impact and thus strengthen the arguments of this study. In addition, if the researcher had given the four tasks multiple times to the same participants over a longer period of time (longitudinal study), more reliable (i.e. converging) results might have been obtained.

Another limitation of the study was the methods. The researcher noticed that the participants attempted to read faster as they progressed through the texts. A "dummy" task should have been given to the participants to act as a buffer to prevent this effect from contaminating the study's data. In addition, there was no way to ensure and verify that the participants were speaking in a natural and speeded manner. It is possible that some may have unconsciously read the tasks faster or slower than they intended to, thus creating the problem of observer's paradox.

Also, the researcher was told by the raters that it was sometimes difficult to give accurate ratings based on the participants' one-minute speeches, especially when many of responses were truncated at the one minute mark. The tasks can be lengthened to gather more data, but there is also the risk of fatigue by the participants which will likely to contaminate the data.

Lastly, it is noteworthy to mention that the rating scheme was mainly focused on grammar, fluency, and discourse but not on phonological correctness. This may have produced biased or inaccurate ratings. For example, raters may have been biased against test takers with a heavy accent or poor pronunciation even if they produced a decent coherent speech in a speeded manner.

REFERENCES

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L., & Palmer, A., (1996). *Language testing in practice*. Oxford: Oxford University Press.

Biancarosa, G. (2005). Speed and time, texts and sentences choosing the best metric for relating reading rate to comprehension. *Written Language & Literacy, 8*(2), 79-100.

Brown, D. (2007). *Principles of language learning and teaching* (5th ed.). White Plains, NY: Pearson Education.

Chapelle, A., & Chung, Y. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing, 27*(3), 301-315.

Choi, Kunghee. (2011). Effects of learning chunks of words on reading comprehension. *Modern English Education, 12*(2), 1-21.

Coniam, D. (1990). Reading aloud speed as a factor in oral fluency and general language proficiency? *Hong Kong Papers in Linguistics and Language Teaching, 13*, 47- 53.

Educational Testing Service. (2008). *Speaking and writing sample tests*. Retrieved November 28, 2011, from http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_sw_sample_tests.pdf

Educational Testing Service. (2008). *TOEFL iBT tips*. Retrieved November 28, 2011, from http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf

Ellis, N. (1998). Emergentism, connectionism, and language learning. *Language Learning, 48*, 631–664.

Ellis, N. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M. Long (Eds.), *Handbook in second language acquisition* (pp. 63-103). Oxford: Blackwell Publishing.

Gates, A. I. (1921). An experimental and statistical study of reading and reading tests. *Journal of Educational Psychology, 12*, 303–314.

IELTS. (2010). *Information for candidates*. Retrieved November 28, 2011, from http://www.ielts.org/pdf/Information_for_Candidates_booklet.pdf

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review, 64*(4), 555-580.

Kim, Nahk-Bohk. (2011). Teaching in chunks: Facilitating English proficiency. *Modern English Education, 9*(1), 30-51.

Kowal, S., O'Connell, D.C., O'Brien, E.A., & Bryant, E.T. (1975). Temporal aspects of reading aloud and speaking: Three experiments. *American Journal of Psychology, 88*(4), 549-569.

Leather, J. (2000). *Second-language speech research: an introduction*. Amsterdam: University of Amsterdam.

Longman, P. (2009). *Official guide to Parson test of English academic*. London: Longman.

Pawley, A., & Syder, H. (1983). Two puzzles for linguistic theory: native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication*(pp. 191-226). London: Longman.

Sinclair, J. (1991). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research, 11*(4), 433-457.

http://www.exampleessays.com/viewpaper/22557.html

http://www.majortests.com/gre/reading_comprehension_test.php?testid=C105

http://www.bced.gov.bc.ca/exams/search/grade10/english/release/exam/0708EN_p.pdf

APPENDIX A

Demographic Information Form

Age: _____

Sex: _____

Major: _____

Language: _____

Number of years living overseas in an English speaking country: _____

Number of years spent learning English: _____

APPENDIX B

Read Out Loud Tasks 1-4

The purpose of this experiment is to measure how fast you can read out loud in a comfortable and natural manner. Please read the following passages in a speeded manner while properly enunciating every word. Task 1-3 are read out loud tasks and Task 4 is a spontaneous speaking task.

TASK 1

The ways my cat shows he is king of the house

I have three animals, one dog and two cats. The one I would like to tell you about is Dink. He is a cat who believes he is the king of the house. I will explain how he shows this to us.

First, you need to know about his physical characteristics. He is a solid black cat, with three white hairs on his back. Currently he has more white on him, due to the fact that he tried to help paint the living room wall. Dink weighs approximately fourteen pounds. Dink can stretch out to be as long as three and some half feet without his tail included. I believe his physical characteristics have quite a bit to do with his feelings about being king of the house.

The beginning of Dink's day begins about ten minutes before the alarm clock goes off.

He does not like waiting to be fed in the morning. He will come up on the bed and begin to meow. If simple meowing does not wake up the humans, he will get louder. It is as if he is beginning to yell. Once he realizes that yelling is not getting us out of bed, Dink will then begin to paw at our faces.

TASK 2

In regard to propaganda the early advocates of universal literacy and a free press envisaged only two possibilities: the propaganda might be true, or it might be false. They did not foresee what in fact has happened, above all in our Western capitalist democracies - the development of a vast mass communications industry, concerned in the main neither with the true nor the false, but with the unreal, the more or less totally irrelevant. In a word, they failed to take into account man's almost infinite appetite for distractions. In the past most people never got a chance of fully satisfying this appetite. They might long for distractions, but the distractions were not provided. Christmas came but once a year, feasts were "solemn and rare," there were few readers and very little to read, and the nearest approach to a neighborhood movie theater was the parish church, where the performances, though infrequent, were somewhat monotonous. For conditions even remotely comparable to those now prevailing we must return to imperial Rome, where the populace was kept in good humor by frequent, gratuitous doses of many kinds of entertainment - from poetical dramas to gladiatorial fights, from recitations of Virgil to all-out boxing, from concerts to military reviews and public executions.

TASK 3

Their no star at eleven o'clock that september nights, and many moons. It was overcastted and windy, and very, very black. Taken a deepening breathes I driven in so began the night that would change my life forever. When I oxygenated and became alerted, I shall see not where the lake ended and the sky began. I blind and I swam.

I say goodbye to my coach, Gus, earlier, I was very worried about finding him on top the escort boat. She Disneyland not gone home. I was nervous about my test score is too low. so low... Getting lost in the Dark and not nearly as brave as I've been portrayed. Gus just look me at 2 or 3 eyes and say, "When you dive in the water, keep your eyes open, to forward, you swim north." I believe in him.

gus ryder been my coach or mentor ever since I joined Toronto's Lakeshore Swimming Club. Although, Id been swimming, since I was nine—and always put my heart into it, I was never very fast and never very good, but that never stopped me so I persevered until I could no longer keep up, but I was so determined nevertheless! Strings got need cutting by sword after a race is finished.

TASK 4

Please answer ONE of the following topic questions.

**You have 1 minute for your response. If you'd like, feel free to brainstorm or outline your ideas. The goal is to talk continuously!

1) If you had a billion dollars, what would you do? How would you spend it?
2) Discuss any political, social or economic issues that are in your mind.
3) Talk about your past traveling experiences. If you don't have any, talk about the places you'd like to visit.
4) Talk about anything.

APPENDIX C

iBT TOEFL Independent Speaking Rubric

Score	General Description	Delivery/Fluency	Language Use	Topic Development
4	The response fulfills the task, with at most minor lapses in completeness. It is highly intelligible and exhibits sustained coherent discourse. A response at this level is characterized by all of the following:	Generally well-paced flow (fluid expression). Speech is clear. It may include minor lapses, or minor difficulties with pronunciation or intonation patterns, which do not affect overall intelligibility	The response demonstrates effective use of grammar and vocabulary. It exhibits a fairly high degree of automaticity with good control of basic and complex structures (as appropriate). Some minor (or systematic) errors are noticeable but do not obscure meaning.	Response is sustained and sufficient to the task. It is generally well developed and coherent; relationships between ideas are clear (or clear progression of ideas).
3	The response addresses the task appropriately, but may fall short of being fully developed. It is generally	Speech is generally clear, with some fluidity of expression, though minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener	The response demonstrates fairly automatic and effective use of grammar and vocabulary, and fairly coherent expression of relevant ideas. Response may exhibit	Response is mostly coherent and sustained and conveys relevant ideas/information. Overall development is somewhat limited, usually lacks elaboration or

	intelligible and coherent, with some fluidity of expression though it exhibits some noticeable lapses in the expression of ideas. A response at this level is characterized by at least two of the following:	effort at times (though overall intelligibility is not significantly affected).	some imprecise or inaccurate use of vocabulary or grammatical structure or be somewhat limited in the range of structures used. This may affect overall fluency, but it does not seriously interfere with the communication of the message.	specificity. Relationships between ideas may at times not be immediately clear.
2	The response addresses the task, but development of the topic is limited. It contains intelligible speech, although problems with delivery and/or overall coherence occur; meaning may be obscured in places. A response at this level is characterized by two of the following:	Speech is clear at times, though it exhibits problems with pronunciation, intonation, or pacing and so may require significant listener effort. Speech may not be sustained at a consistent level throughout. Problems with intelligibility may obscure meaning in places (but not through-out).	The response is limited in range and control of vocabulary and grammar demonstrated (some complex structures may be used, but typically contain errors). This results in limited or vague expressions of relevant ideas and imprecise or accurate connections. Automaticity of expression may only be evident at the phrasal level.	The response conveys some relevant information but is clearly incomplete or inaccurate. It is incomplete if it omits key ideas, makes vague reference to key ideas, or demonstrates limited development of important information. An inaccurate response demonstrates misunderstanding of key ideas from the stimulus. Typically, ideas expressed may not be well connected or cohesive so that familiarity with the stimulus is necessary to follow what is being discussed.
1	The response is very limited in content or coherence or is only minimally connected to the task. Speech may be largely unintelligible. A response at this level is	Consistent pronunciation and intonation problems cause considerable listener effort and frequently obscure meaning. Delivery is choppy, fragmented, or telegraphic. Speech contains frequent pauses and	Range and control of grammar and vocabulary (or prevent) expression of ideas and connections among ideas. Some very low level responses may rely on isolated words or short utterances to communication ideas.	The response fails to provide much relevant content. Ideas that are expressed are often inaccurate, limited to vague utterances, or repetitions (including repetition of prompt).

	characterized by at least two of the following:	hesitations.
0	No attempt is made.	

*Note: Task duration is 1 minute. A truncated response DOES NOT necessarily signify incompletion.

Youngsoo Song
Dept. of English Language Education
Korea University
College of Education
5-1 Anam-dong, Seongbuk-gu, Seoul, 136-701, Korea
Tel: 1-778-707-6076
Email: young.song86@gmail.com

Received 20 October 2011
Revised 1 December 2011
Accepted 13 December 2011