*Multi-level components of L2 reading fluency: An exploration*

Andrew Atkins

1.0 Introduction

Fluency development and vocabulary instruction has been overlooked in favour of grammar teaching in secondary EFL education in many Asian countries, and as a result when students arrive in tertiary education many are unable to use English in a meaningful way. With the ubiquitous nature of the internet, being able to read efficiently is a skill that many educators see as important in their students’ futures. This paper reports on an attempt to explore the reading fluency of a cohort of first-year students at a Japanese university and tries to ascertain which components from a set of variables predicted to influence reading fluency has the greatest influence. The reason for interest in these variables is to advise curriculum and syllabus design, to provide a focus for remediation, and to assist informed placement decisions.

1.1 Background to the study

In recent years there have been some attempts in Asian EFL classrooms to promote reading fluency as perhaps a backlash to the traditional focus on intensive reading for academic purposes. The reason for this small revolution against the norm is that many teachers feel that English education is not serving the needs of students and that it results in students reading in a highly inefficient manner (Gorsuch & Taguchi, 2010). The most common tasks that have been used to facilitate the process, if the literature is an indication of what is being done in the classroom, have been extensive reading (Waring,
1997; Day & Bamford, 1998), timed reading (TR) (Chang, 2010; Underwood, Myskow & Hattori, 2012), and repeated reading (Samuels, 1979/1997; Gorsuch & Taguchi, 2008). In many cases a combination of two or more of these activities have been employed with varying success. This paper will not assess longitudinal increases in fluency but rather take a snapshot of the situation as it stands using TR to assess fluency at a point when participants have been performing timed reading for over six months.

1.2 The problem, purpose of the study, and significance

Research by Bernhardt (1991, 2005) suggests that the various aspects of vocabulary knowledge may account for 50 per cent of the variability in language ability, and it is interesting to explore to what extent fluency is a part of lexical knowledge, as suggested by Meara’s (1996) dimensions of lexical knowledge.

A TR course is proposed to be a means of increasing reading fluency (Nation, 2007). The effectiveness of timed reading practice though, as a tool for helping to increase reading fluency, is still being debated; however the texts used are roughly equivalent in terms of difficulty, and therefore should provide a measure of fluency at a given time of administration that can be used in this study.

Over a number of years of collecting data from weekly timed reading practice, I started to speculate on why students, even after extensive practice, varied greatly on performance (a between-individual fluency gap), and whether it was purely vocabulary knowledge that controlled the attainable speed and comprehension level, or it was more to do with a more static variable such as cognitive speed (Carver, 1992) or working memory span (Osaka & Osaka, 1992).

The present study builds on previous research into timed reading and reading fluency and addresses a gap in the literature by examining reading fluency, albeit on a
preliminary level, and which predicted components of reading fluency effect performance on timed reading passages.

The study contributes to the reading fluency literature in several important ways. Many previous studies of TR have looked at longitudinal growth in reading speed, as this is the stated purpose of the activity, and have tended towards small sample sizes (Chung & Nation, 2007, Utsu, 2004, 2005), as is often the case with longitudinal studies. In contrast, this study has a larger sample and utilises a cross-sectional approach to explore the relationships between variables and influences on the dependent variable. Many studies have also struggled to find dependant variables with sufficient face validity in that they have used words per minute as the dependent variable and failed to fully account for variation in reading comprehension. This means that readers may move their eyes quickly over a text without understanding and can still obtain a high score. This seems to be a major problem.

The importance of reading fluency cannot be stressed enough, as expounded by Segalowitz (2010, p. 2), who explains the concept of the fluency gaps. He suggests that L2 learners are often frustrated by their deficiency in the L2, as compared to their L1, especially where fluency is concerned. This within-individual fluency gap is of course of great concern to teachers whose job it is to narrow the gap and to learners who desire to become proficient in their L2. Segalowitz (ibid.) also discusses another type of fluency gap, a between-individual fluency gap, which is more the focus of this paper. This paper hopes to contribute to the field by providing information about the factors that make for more fluent readers, and why some students are more successful in achieving fluency than others.
1.3 Definition of Fluency

Reading fluency has been defined by many researchers and scholars, however; the amalgamated definition Wolf and Katzir-Cohen (2001) arrived at from reviewing a large number of other studies is concise and appropriate for the purposes of this research: “reading fluency refers to a level of accuracy and rate where decoding is relatively effortless; … and where attention can be allocated to comprehension” (p. 219). Thus when examining reading fluency, the two key components that must be taken into account are decoding, meaning the process of translating written symbols into meaningful units, and comprehension. It should be noted though that the studies that Wolf and Katzir-Cohen (2001) examined were L1 studies that used oral reading to measure reading fluency; however as oral reading could really be a proxy for silent reading (Fuchs and colleagues 1988, 2001), the definition still has validity.

1.4 Timed Reading

TR, according to Champeau de López (1993), and for the purposes of this research, is the reading of texts of equal length and equal lexical difficulty, regularly over a period of weeks or months with the purpose of increasing reading fluency, and this is achieved by increasing what Samuels (1994) termed automaticity, that is the ability to decode and comprehend text without much conscious effort. Texts are timed as they are read and reading is followed by comprehension questions, which are answered without referring back to the text. The pressure of timed reading is thought to be the catalyst for improvement in reading fluency, as this provides what Nation (2007) calls “encouragement to perform at a faster than usual speed” (p. 7). Research indicates that reading speed and comprehension generally increase with practice (Utsu, 2004, 2005).
Recent attempts to assess the effectiveness of L2 TR include studies by Chung and Nation (2006), Crawford (2008), Macalister (2010), and Utsu (2004, 2005). Chung and Nation (2006) suggested there was no established statistical way to measure gains for TR, although most fields outside of ESL offer a number of means, and explored three similar methods to assess gains using only percentage increase in reading rate. Most students improved their reading speed over the course. Macalister’s (2010) study also used increases in reading rate to measure fluency. The study was a small preliminary study, but reported increases in reading speed for most. Utsu’s (2004, 2005) studies measured changes in reading rate, adding separate comprehension scores as another variable in the research. The studies reported improvement in both reading rate and comprehension as percentage changes. Crawford’s (2008) study analyzed reading rate growth using a repeated measures design, and provided some positive support for the use of TR, although the author suggests there were issues of validity.

1.5 Word recognition and reaction time

To read a passage of text fluently, one must possess the ability to recognise the letters and words contained in the texts with automaticity. There are many levels to this skill, but first, certainly when reading, the word’s form must be known and recognised before comprehension can begin. Ideally this recognition and the processes that follow should be as close as possible to automaticity in order that attention can be given to meaning of text above the word level.

According to Koda (2005), “word recognition, despite its significance, has received scant attention in L2 research” (p. 37). Although there is more in L1 research, this cannot be generalised to L2 contexts as we are unsure the processes are the same. This research attempts to replicate the speeded number matching test (explained below)
used by Brown and Haynes (1985) in order to try to isolate the aspect of speeded decision task that is unrelated to alphabetic processing in order to try to parcel this variance.

1.6 Testing medium

For the purposes of this study, TR and reaction time tests (RT) were administered on a computer to facilitate accurate timing. The Vocabulary Size Test (VST) (Nation & Beglar, 2007) and the Speed of Thinking Test (STT) (Carver, 1992) were administered on paper as accurate timing was less important.

1.7 Research Purpose and Questions

The purpose of this research is to examine more closely the necessity of vocabulary knowledge and its effects, along with the effects of innate physiological variables, on reading fluency performance. The research questions examined in this study are:

(1) Which of the variables measured in this study (RT, VST and STT) have a relationship with reading fluency performance as measured by TR?

(2) Which of the measured predictor variables account for the most variation in reading fluency?

2.0 Methodology

2.1 Participants and Sampling

A total of 134 participants from 3 intact classes were recruited to take part in this study. The students were from a small national university in western Japan and were studying required English classes as part of their Economics related majors. The
participants came from the Economics faculty in the university and had an average age of 19 years. There was a gender bias with 33 female subjects (24.6%) and 101 male subjects (75.4%). Students at the university were not divided in ability-based streams, but rather assigned in a relatively random manner to groups. There were however some scheduling constraints that made assignment not entirely random.

2.2 Materials

2.2.1 Timed reading

The instrument at the backbone of this research was a test of timed reading. The treatment consisted of two equivalent reading passages, each with 6 comprehension questions. The passages were controlled for vocabulary load (Webb & Nation, 2008) by checking their vocabulary against that known by the learners, and questions were designed to be equivalent, with one overall gist question, 2 localised gist questions and 3 simple comprehension questions. The questions were designed to test understanding and not memory. The readings were a combination of specially created passages and those adapted from other freely available sources. The passages were administered randomly from a bank of 26 passages using a specially created web application (see Appendix B).

The TR application was a simple web application, built using PHP, Javascript and an MYSQL database. Students were able to log on over the internet. On the click of a button they were presented with a short reading passage and a timer, showing the time spent reading in seconds. Clicking a further button presented them with a set of six comprehension questions that they were required to answer from memory without recourse to the passage. Students’ reading time and score were then recorded on the database (Atkins & Cole, 2011).
2.2.2 Vocabulary Size Test

Nation and Beglar’s (2007) Vocabulary Size Test (VST) provided information required for analysis and is easy to administer in a short period of time. The VST provides data for receptive vocabulary knowledge by test sampling 10 words from each 1000-word level of vocabulary in the British National Corpus (BNC), and not just the total number of words known.

Knowing approximately the words that students are likely to know at a receptive level provided by the VST and comparing this with word profiles of texts from the RANGE program helps teachers to identify words that are likely to be unknown. The VST makes some assumptions that remain to be tested, in that the test uses the unit of word families and assumes that if students know a headword then they will have receptive knowledge of most of the other members of the word family. This assumption seems to hold for L1 situations (see Bertram, Laine, & Virrkala, 2000 for a discussion), but there is some doubt that it is applicable for L2 situations (Schmitt & Zimmerman, 2002). The validity and reliability figures for the test are also unknown.

Only the first five levels of the fourteen-level test were used as these were felt to be the most relevant for the purpose of this investigation. The test consists of ten test items for each one thousand-word level; therefore a fifty-item test was used for this study. Timed reading is designed to promote fluency and as a result the lexical load is intentionally low and seldom exceeds the most common five thousand words in English.

2.2.3 Speed of Thinking Test

Carver’s (1992) Speed of Thinking Test was used to discover information about participants’ cognitive speed at the letter level. The test consisted of 184 items administered over 2 minutes in total. On Paper, the test presented paired upper and lower
case combinations of the letters $a$, $b$, $A$, and $B$, and test takers were required to mark whether the letters were the same or different. See Appendix A for an excerpt from the test. The test was designed so that it could not be completed in the two minutes given, although originally the test had fewer questions but was increased when a student in a previous cohort finished it.

### 2.2.4 Word Recognition Reaction Time Test

Due to results attained from administering the VST, it was known that students were potentially weak on the second 1,000 words of English as defined by the BNC. It was decided to test students on the “Anglo-Saxon” words from the aforementioned second 1,000 words of English, as it was felt that due to the nature of English study at high-school level in Japan, these would produce the greatest variance between the students' performance. This test was designed to check students’ word recognition skills. Using the Compleat Lexical Tutor website (Cobb, 2003), I generated a list of Anglo-Saxon words from the second 1,000 words of English, and this list was then randomised. The first 40 words were selected for the reaction time test. In addition to these words 20 non-words were chosen in a similar manner. Using Paul Meara’s list of 100 plausible-non-words (available on Cobb’s website), the list was randomized and the first 20 chosen. The 40 Anglo-Saxon words and the 20 plausible-non-words were made into a reaction time test using the Compleat Reaction Timer v.4 (Cobb, 2003). Another test was also made as a practice test using 20 Anglo-Saxon sounding real words and 10 plausible-non-words, so that participants were familiar with the format before they did the actual test. Reaction times were recorded by the program and sent by email to the teacher.
2.2.5 Speeded Number Matching Test

A number matching reaction time test was also created using Compleat Reaction Timer v.4 (Cobb, 2003). The test was based on one described in Haynes and Carr (1990), but was delivered on computer instead of on paper. The test was designed to check whether number recognition was related to word or letter recognition, and which was most closely related to reading performance. It was also thought that the test score could function as a covariate in later analysis.

In order to make the test, a random list of 3-digit numbers from 100-999 were generated using the Compleat Randomizer v.2.4 (Cobb, 2003). The first 30 numbers were chosen to be used in the test. 20 of these numbers were used as correct answers and 10 as incorrect answers. The correct answers were presented as a matching pair of 3-digit numbers, e.g. 123_123, and the incorrect answers were presented as non-matching pairs, e.g. 456_654. Reaction times were recorded on the individual students computers and these were sent by email to the teacher.

2.3 Procedures

At the beginning of the school year the participants were asked for their cooperation in research over the year and none of them objected to being involved. TR started from the second week of the school year, and two randomly-assigned TR passages and corresponding questions were completed each week, at the start of every lesson. The TR data for this study was taken from week 20 of the school year. Instructions for each test were given in Japanese to ensure there were minimal mistakes. All of the tests were group-administered.

The first five levels of the VST (Nation & Beglar, 2007) were administered in week 4 of the school year, taking 40 minutes to administer. The STT (Carver, 1992) was
administered in week 5, and took only 10 minutes to explain and administer. In week 20, before the TR tests were completed, the *Word Recognition Reaction Time Practice Test* was administered, and this was followed by the *Word Recognition Reaction Time Test*. Finally, in week 21, the *Speeded Number Matching Test* was administered.

2.3.1 Analytical Approach and Variables

As mentioned earlier, a consensus on how to measure timed reading studies has yet to be reached; however in this study an attempt has been made to develop a variable with better face value than raw words per minute. For the purposes of this study a composite score for each reading is calculated from words-per-minute (WPM) multiplied by the raw percentage correct on the multiple choice comprehension test expressed as a decimal. Essentially this WPM adjusted for comprehension, e.g., if a reader reads a passage in 60 seconds and gets all of the comprehension items correct their reading rate of 300 WPM is not adjusted; however for a student reading in the same time of 60 seconds (300 WPM) who attains only 83.33% (5 out of 6) on the comprehension items, their reading rate is multiplied by 0.8333 to give an adjusted score of 249.9 WPM.

To further account for variation in performance and uncontrollable differences in the difficulty of the TR reading passages the mean adjusted WPM for the 2 passages administered in week 20 was used as the dependent variable. The higher the number the more proficient a reader is assumed to be.

The VST gave raw scores from 10 for each of the five levels tested and a raw total score for all five levels combined. A higher number suggests greater vocabulary knowledge. The STT provided a raw total of questions answered in the two minutes allowed, and a higher score suggests that a participant has faster cognitive speed. The *word recognition reaction time test* provides us with an average reaction time for
distinguishing whether a word is known or not, and it also provides us with an error score. The reaction time shows us how fast a participant can distinguish a word from a non-word, but the error score is ambiguous, as it can show genuine errors in processing, but can also include real words that are unknown. The *speeded number matching test* provides less ambiguous results, as it excludes any knowledge of English, and should be a pure measure of cognitive processing speed. Errors on this test are mistakes made by the participant and the score should be adjusted in some way to account for this, but I am unsure how to do this in a fair manner.

For the purposes of addressing research question one, multiple correlation was used to examine the relationship between the dependent TR variable and independent variables. Research question two was addressed using hierarchical linear regression. All analyses were carried out using SPSS 17.0 for Windows.

3.0 Results

Before addressing the research questions directly, it is useful to examine the descriptive statistics for the variables measured in this study. Table 1 shows descriptive statistics for participants generated using SPSS 17.0.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRmean</td>
<td>124</td>
<td>127.53</td>
<td>80.52</td>
</tr>
<tr>
<td>VST_k1</td>
<td>128</td>
<td>8.76</td>
<td>1.32</td>
</tr>
<tr>
<td>VST_k2</td>
<td>128</td>
<td>5.62</td>
<td>1.78</td>
</tr>
<tr>
<td>VST_k3</td>
<td>128</td>
<td>5.84</td>
<td>1.71</td>
</tr>
<tr>
<td>VST_k4</td>
<td>128</td>
<td>5.42</td>
<td>1.50</td>
</tr>
</tbody>
</table>
TRmean is the adjusted mean reading score for the two timed reading passages. The metric is based on words per minute, but is adjusted for comprehension. VST_k1 is the raw score on the VST at the first 1000-word level. VST_k2 is the raw score on the VST at the second 1000-word level. VST_k3 is the raw score on the VST at the third 1000-word level. VST_k4 is the raw score on the VST at the fourth 1000-word level. VST_k5 is the raw score on the VST at the fifth 1000-word level. VST_k1to5 is the raw score on the VST for the first five 1000-word levels. RTtest is the mean reaction time for the 60 items on the reaction time test based on Anglo-Saxon words from the second 1000 words of the BNC. RTerrors is the number of errors on the RTtest expressed as a percentage. STT is the total number of questions answered on the Speed of Thinking Test. RTnumTest is the mean reaction time for the 30 items on the Speeded Number Matching Test. RTnumERRORS is the number of errors made on the Speeded Number Matching Test expressed as a percentage.

It can be seen that the valid sample size for all of the variables is 100, and as alluded to earlier that longitudinal studies often have small sample sizes, this is due mainly to the instruments being administered over a range of occasions, and missing only
one test would exclude a participant from the total valid sample. The mean score for the VST_k2 is also of note, as this shows that vocabulary size as assessed by the VST is greater for the third 1,000 words of English than for the second.

The reaction time test results and errors are worthy of some description as it can be seen that there is little difference in the average reaction times for each test ($RT_{test} = 0.94$, $RT_{numTest} = 0.87$); however the cognitive demands of the tests are not equal, with the $RT_{numTest}$ being more complex as it is a comparison of two numbers. The $RT_{test}$ score appears to be rather slow, and perhaps double the time needed by an educated native speaker of English. The error percentages are more interesting. There are still errors on the number test, although they are much fewer than in the word test, but this shows that speed was being prioritised over accuracy. The much higher percentage for the word test will include these errors from prioritising speed, but also contains real words, which are unknown to the participant, and in this sense it appears to be conflated and difficult to use in further analysis.

Research question 1 asked which of the variables measured have the stronger relationship with reading fluency as measured by TR performance, and to assess this multiple correlation was used. Table 2 shows correlations between the variables.

Table 2. Intercorrelations between measured variables.

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. TRmean</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2. VST_k1</td>
<td>.245**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3. VST_k2</td>
<td>.386**</td>
<td>.430**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4. VST_k1to5</td>
<td>.456**</td>
<td>.641**</td>
<td>.780**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5. RT_test</td>
<td>-.326**</td>
<td>-.066</td>
<td>-.068</td>
<td>-.158*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6. RErrors</td>
<td>-.264**</td>
<td>-.318**</td>
<td>-.442**</td>
<td>-.517**</td>
<td>-.247**</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
We can see a number of interesting correlations in the table. Reading fluency (TRmean) correlates significantly with all of the other variables except the speeded number matching test (RTnumTest). Significant correlations however vary from small to medium-large. The speeded number matching variable comes from the only instrument that is non-alphabetic, whereas the letter-level speed of thinking test (STT) and the word-level reaction time test (RTtest) are both significant. The negative correlation of these two variables is due to the nature of the scales used; a higher score on the STT faster, whereas a lower score on the RTtest is faster. The vocabulary-based variables (VST_k1, VST_k2, and VST_k1to5) all correlate significantly with reading fluency, however it is surprising that the total vocabulary variable (VST_k1to5) correlates most strongly, because the reading passages are written to be within the first thousand words of English, not the first five thousand. Relating to research question 1, we can see that vocabulary as measured by the VST, and word-level reaction time as measured by a reaction time test, have the strongest relationships with reading fluency as measured by timed reading performance.

Research question 2 asked which of the variables measured account for the most variation in reading fluency. Bearing in mind the results from the multiple correlation data and to discover the answer it was necessary to perform a multiple regression analysis. The results of the analysis are shown in Table 3.

Table 3. Hierarchical Regression Analysis Summary for Measured Variables Predicting Reading Fluency (TRmean) (N = 109)
A hierarchical regression analysis was used to test if the most likely predictor variables significantly predicted participants' performance on the criterion variable (trAVEadj). From the results of the multiple correlations it was predicted that the total VST score (VST_1to5) would account for the most variation, and in Model 1 this was the only predictor included. All assumptions were met including the assumption that independent errors were tenable. The results of the regression for model 1 indicated VST_1to5 explained 20.8% of the variance ($R^2=.21$, $F(1,107)=28.164$, $p<.001$). Other vocabulary predictors could not be used as autocorrelation was present. In other words, they were measuring the same thing. For Model 2 it was thought that the performance on the reaction time test (RTtest) would account for the next portion of the variance, it was added to Model 2. The results of the regression for model 2 indicated that RTtest accounted for an extra 6.9% of the variance ($R^2=.28$, $F(2, 106)=20.316$, $p<.001$). Finally it was thought that performance on the speed of thinking test (STtest) could account for some of the variance, and it was added to model 3. STtest did not account for a significant

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE B</td>
<td>β</td>
<td>B</td>
<td>SE B</td>
<td>β</td>
</tr>
<tr>
<td>VST_1to5</td>
<td>6.076</td>
<td>1.145</td>
<td>.456**</td>
<td>5.485</td>
<td>1.115</td>
<td>.412**</td>
</tr>
<tr>
<td>RTtest</td>
<td>-90.332</td>
<td>28.455</td>
<td>-.266**</td>
<td>-86.001</td>
<td>29.365</td>
<td>-.253**</td>
</tr>
<tr>
<td>STtest</td>
<td></td>
<td></td>
<td>.205</td>
<td>.328</td>
<td>.054</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.208</td>
<td></td>
<td>.277</td>
<td></td>
<td>.280</td>
<td></td>
</tr>
<tr>
<td>F for change in $R^2$</td>
<td>28.164**</td>
<td></td>
<td>10.078**</td>
<td></td>
<td>.391</td>
<td></td>
</tr>
</tbody>
</table>
amount of variance in the model ($R^2 = .28$, $F(3, 105) = 13.596, p < .001$). It was found that vocabulary size as measured by the VST_1to5 significantly predicted reading fluency as measured by TRmean ($\beta = .41, p < .001$), as did word recognition reaction time as measured by RTtest ($\beta = - .27, p < .01$), and as explained above, the negative correlation between the variables is the cause of the negative $\beta$-value.

A further analysis to try to parcel out variance accounted for by the aspect of speeded decision tasks unrelated to alphabetic processing was performed, but the effect was not statistically significant.

4.0 Discussion

With regard to research question 1, many of the variables correlate with reading fluency, however the strength of the relationship also varies greatly. Many of the variables measured also correlate with each other, as in many cases they are measuring the same thing. For example, the VST variables along with the RTerrors variable are all full or part measures of vocabulary knowledge. They provide little unique insight into what, apart from vocabulary, is related to performance on the timed reading tests. This is in line though with Bernhardt (1991), that showed vocabulary to be the greatest predictor of English reading ability.

The TR passages and tests themselves are another issue that must be considered when discussing the results. The data obtained was from a much larger study of timed reading performance, and for the purposes of that study participants are randomly assigned passages from a pool of passages in order to eliminate an order effect in the repeated measures. This is then averaged out over six to seven readings, but in this study the metric is only the mean of performance on two passages, so we should expect greater variability in performance than when averaged over more passages. Even with the greatest
care taken to make the passages equivalent, there are so many factors that affect the difficulty that it is not possible to control for them all.

Research question 2 asked “(w)hich of the measured predictor variables account for the most variation in reading fluency?”, and the regression results suggested that receptive vocabulary, as measured by the first five levels of the VST account for the greatest amount of variance in the criterion variable for reading fluency, as measured by the mean words per minute for two 300-word, lexically controlled passages adjusted for comprehension. This accounted for 21% of the variance, far less than in Bernhardt (1991), although both studies found that vocabulary was the most influential predictor. The only other predictor that I was able to include in the regression that accounted for a small but significant amount of variance was the word recognition reaction time test. This accounted for 6.9% of the variance in the model, but this figure must be interpreted with discretion as in some cases fast reaction times came at the expense of a high number of errors, and vocabulary size had a large effect on errors made. In hindsight it would have been useful to have measured this variable with higher frequency words that participants would have been more likely to be certain of their knowledge.

5.0 Conclusions

Although the results of this study have suggested that vocabulary size and word recognition account for a significant proportion of reading fluency, the quality of the instruments used dictates the quality of the results. The dependent variable is perhaps not completely valid as it is really comparing apples with oranges as the participants actually read completely different passages on different subjects. This problem will be reduced in the parent study by using mean times for a series of readings. Reading fluency is a very personal thing and influenced by not only the text itself, but also what the reader brings to
the text on a particular day. It would be better to assess it over a longer period of time and with a variety of passage themes, and perhaps try to control vocabulary more. It is also conceivable that the controlled nature of the texts reduces the influence of vocabulary size on performance, and this is worthy of further investigation.

Although simplicity is often the best strategy in research design the complex nature of reading fluency makes simplicity untenable, and in this study more than 70% of the variance in the regression model is unaccounted for, which is comparable to Bernhardt’s (1991). Although, as discussed, the dependent criterion variable in this study may not be truly valid, it is also certain that other variables need to be measured, and these should include some measure of grammatical knowledge, some tests of a greater variety of individual differences, and some measure of cultural knowledge and schema for the passages used. Shiotsu (2009) used a more comprehensive battery of tests in his study and this study would benefit from some of those tests, especially the sentence-level reading ability test and a further word recognition exercise. Shiotsu (ibid) also makes the recommendation that “future research should employ standardized test scores or proficiency ratings” (p. 36), and although the VST is not standardized it has been used with thousands of students in Japan and around the world, and at least in part satisfies Shiotsu’s recommendation.

This study provides little in the way of pedagogical implications; however it has provided some valuable insight and knowledge that will be included in the parent study where some of the data used in this study was taken from. The parent study is a reading and writing fluency parallel growth study, and it is hoped it will provide some much needed information for L2, and especially EFL educators to help them prepare students to use language more fluently and effectively. Further examination of reading fluency and its
component parts can only lead to greater understanding of what is needed in the classroom to create students who can use English to make their futures brighter.

References


Appendix A

Items 1 to 20 of The Speed of Thinking Test (Carver, 1992)

<table>
<thead>
<tr>
<th>SAME</th>
<th>DIFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(</td>
</tr>
<tr>
<td>2</td>
<td>(</td>
</tr>
<tr>
<td>3</td>
<td>(</td>
</tr>
<tr>
<td>4</td>
<td>(</td>
</tr>
<tr>
<td>5</td>
<td>(</td>
</tr>
<tr>
<td>6</td>
<td>(</td>
</tr>
<tr>
<td>7</td>
<td>(</td>
</tr>
<tr>
<td>8</td>
<td>(</td>
</tr>
<tr>
<td>9</td>
<td>(</td>
</tr>
<tr>
<td>10</td>
<td>(</td>
</tr>
<tr>
<td>11</td>
<td>(</td>
</tr>
<tr>
<td>12</td>
<td>(</td>
</tr>
<tr>
<td>13</td>
<td>(</td>
</tr>
<tr>
<td>14</td>
<td>(</td>
</tr>
<tr>
<td>15</td>
<td>(</td>
</tr>
<tr>
<td>16</td>
<td>(</td>
</tr>
<tr>
<td>17</td>
<td>(</td>
</tr>
<tr>
<td>18</td>
<td>(</td>
</tr>
<tr>
<td>19</td>
<td>(</td>
</tr>
<tr>
<td>20</td>
<td>(</td>
</tr>
</tbody>
</table>

...  

Note. The complete test contains 184 items.
Appendix B

A sample reading passage used in the TR software.

Timed Reading

The Inuit

The Inuit live in Canada, Greenland and the north of America. They first arrived there about 5,000 years ago. They live in a land of ice and snow, a land without trees. For nine long months of the year it is cold, dark winter, while summer is only three short months.

Today, many Inuit live modern lives in towns and cities, but not so long ago the people lived the way their parents and their parents' parents had lived for thousands of years.

Like many people all over the world, the Inuit do not want to lose their languages. There are about 90,000 people who speak Inuktitut-Ainus and because there is not much difference between the languages, people from different places can understand each other. They want to remember their culture and teach their children the old ways.

Most Inuit lived in small groups of only a few families. Many of these groups lived near the sea because most of their food came from there. Because the weather was too cold to grow anything, their food was the fish and animals that they caught. Other Inuit moved from place to place following the animals called caribou, which made up the biggest part of their food. Sometimes the Inuit baked their food but often it was not cooked at all.

Because it was always very cold, the Inuit had to wear thick, warm clothes. The women made clothes from the skins of animals. Each piece of clothing could take weeks or even months to make and they could be very beautiful. Mothers wore special clothes so they could carry their babies on their backs against their skin to keep them warm and safe.

Now the Inuit way of life is changing but the old ways are remembered and valued.