Development of a English Learners’ Reading Fluency Test with regard to Pre/Post Test Design

A Young Park

Abstract
To increase the benefit of a pre/post test design, the validity and reliability of the research should be carefully considered, since a pre/post test design require tests to be conducted multiple times. The current study developed a reading fluency test with a pre/post test design that aimed to measure the effect of the intervention on English learners. The developed reading fluency test was piloted twice and subsequently revised according to the pilot test results. According to the results of the pilot studies, the current study a) increased the test validity by measuring not only reading rate but also comprehension and b) strengthened the test reliability by using different reading texts that took into consideration the comparability of the readability of the text used across the pre and post tests.

Keywords: Reading fluency, Test, Reading rate, Reading comprehension, Pre/post test design, English learner

1. Introduction
Among the many experiment research designs that are used in language studies, the pre/post test design is widely used “to compare participant groups and measure the degree of change occurring as a result of the program or intervention” [1]. It is a useful design to determine whether a program or intervention has had the intended effect on participants based on the difference between the results of a pre test (before the intervention) and those of a post test (after the intervention) [2]. However, in order to maximize the benefit of a pre/post test design, it is essential that it be carefully designed with regard to increasing the reliability of the test. Most importantly, the different versions of the pre and post-test should be comparable; that is, there should be internal consistency reliability between two tests in order to strengthen the reliability of the test. Another important issue with regard to test design is ensuring that the test is truly measuring what it intends to measure, in order to increase the validity of the test. Test validity is an indicator of how much meaning can be placed upon a set of test results [3]. In language testing, where the importance and accuracy of tests is
paramount, test validity is crucial [4].

The current study developed a reading fluency test with regard to pre/post test design, strengthening the reliability through ensuring comparability between the pre and post tests and increasing its validity by measuring English learners’ reading fluency in terms of both reading rate and comprehension. The developed test was revised according to the pilot test results.

2. Literature Review

2.1 Validity and Reliability in Language Tests

Although there were several important issues to consider in measuring the language learners’ reading fluency, the current study focused on the issues of validity and reliability. Reliability refers to “the consistency of a given testing instrument: the extent to which participants would be expected to perform similarly across multiple administrations of the test under similar conditions” [4]. In a pre/post test design, reliability is a critical issue because of ‘practice effects’. Practice effects occur when participants who take the same test repeatedly show progress due to familiarity with the test items or the testing format [5][6]. One solution is to provide test instruments with at least two different versions, one for the pre test and another for the post test. At the same time, these different versions of the test should be comparable in terms of difficulty to avoid issues of reliability. The importance of reliability also applies to a pre/post test design in language acquisition research [7]. For instance, participants who are given the same reading texts in pre and post tests of reading fluency may remember the reading text and then be able to read it more rapidly in the post test. To improve the reliability of reading fluency tests, research studies employing a pre/post test design that uses more than one reading text in the pre/post test design should provide information about the similarity of participant performance on the different versions of the tests or the comparability of reading difficulty level across the tests [8].

Validity refers to “the extent to which a given test instrument is measuring what it claims to measure” [9]. Before applying the pre/post test design in the research, we should think “which area we are interested in measuring” and “does the measurer (test instrument) align with and cover this content area” [3]. Although there are many different types of validities, content validity is the central focus of the current study’s developed test with regard to pre/post test design. Content validity is “the extent to which the items or the tasks on a given test instrument sufficiently cover the area of interest” [8]. To maximize content validity, testing items or tasks should represent the “full range of skills connected with the purpose of the test. Skills not associated with this content area should not influence performance on the test” [8].
2.2 Reading Fluency Tests in Previous Studies

The current study reviewed reading fluency test in previous studies of English pedagogy particularly, related to this study’s research focus. Many quasi-experimental studies have attempted to measure English learners’ reading fluency using a pre/post test design. However, these studies have validity and reliability problems in terms of measuring reading fluency [10]. Firstly, some of the previous studies used identical reading texts for the pre and post tests to measure reading fluency; that is, the exact same passages were used for the pre and post reading fluency tests [10]. Considering that most of these studies were conducted over relatively short periods of time (i.e. a minimum of 2 weeks to a maximum of 17 weeks), there is a possibility that the participants became familiar with the reading text during the pre test or even memorized its content. This may have caused an artificial increase in their post test scores, threatening the test reliability. For example, in Iwahori’s [10] study, 37 Japanese who were English learners took a reading fluency test using the same text for the pre and post test. The result showed the experimental group’s reading fluency to have increased significantly after the intervention. However, since the interval between the pre test and the post test was only 7 weeks, the test takers may have remembered the content of the text. Thus, the pre test may have had a significant influence on the post test score. However, this potential influence was not identified in this study.

Secondly, although many of previous studies did use different texts for the pre and post fluency tests, there is no mention of whether the texts were of equivalent readability [11]. Readability means how easy a reading text is to read and understand [12]. That is, the comparability of the readability of the texts used in the pre and post tests was not examined in those studies, a factor that could weaken the reliability of these fluency tests and their results. For instance, in the research carried out by Kao [11] on 15 Chinese who were English learners, two different reading texts were used for the pre and post tests. However, Kao [11] doesn’t mention the readability of the two reading texts used. If the level of readability of these two tests differed, a comparison of the results of the pre and post tests is suggestive rather than conclusive. To increase the reliability of measuring reading fluency for data collection, within the process of data collection and analysis, the present study developed a fluency test that allows for different reading texts of equivalent readability to be used in the pre and post test. In brief, an overview of previous empirical studies provided the current study with valuable design criteria for a more valid reading fluency test. Keeping these criteria in mind, the current study selected different reading texts with equivalent readability for the pre and post tests in order to strengthen the reliability of the current study’s developed reading fluency test.

3. Development

Firstly, a base study was conducted to estimate the readability level of the reading fluency test. After
development of the reading fluency test, two pilot studies were administered to assure the reliability and validity of the test.

3.1 Base Study

For the base study, an English textbook used by participants was analysed using the Flesch-Kincaid Grade Level [13] method to estimate participants’ English reading levels for the test development. The readability of the reading texts used for the reading fluency test development should match the participants’ reading levels. Since it is difficult to measure participants’ reading levels directly, the participants’ English textbook, Finalizing English reading I [14], was analysed based on factors such as the number of words per sentence and the number of letters or syllables per word Flesch-Kincaid Grade Level [13] using the software VocabProfile V1.5 [15]. In the current study, the target population was 1st grade Korean high school students who took English courses. Therefore, this base study was conducted before the test development to establish the readability of the reading texts used in the developed reading fluency test. The results were Flesch-Kincaid Grade Level 5.1.

3.2 Development of the Reading Fluency Test

To measure the change in the participants’ reading fluency before and after the intervention, the current study developed reading a fluency test with a pre/post test design. However, to measure whether the current study’s intervention truly resulted in any long-term reading fluency development, a delayed post-test was added to the pre and post tests. As mentioned above, in this study, validity and reliability were the main concerns with regard to its pre/post test design.

To increase the reliability of the developed fluency test, the pre, post, and delayed-post tests required participants to read previously unseen passages. In other words, different reading texts had to be used across the three tests. And these three different versions of the test had to be comparable; that is, there needed to be internal consistency reliability between the three tests. To assure the readability of the reading rate tests, the Flesch-Kincaid Grade Level was adopted. This is one of the most popular reading measures in reading research. Next, six different passages were first chosen from New Year around the World [16]. New Year around the World is one of a series of graded readers from Oxford University Press. This graded reader was chosen because it was scaled as Flesch-Kincaid Grade Level 5.0, which was very close to the target populations’ readability (Flesch-Kincaid Grade Level 5.1). The target populations’ readability was estimated according to the English textbook analysis in the base study. Next, a detailed analysis of six chosen reading passages was carried out to assess their comparability in terms of text difficulty and length using the Flesch-Kincaid Grade Level formula. The reading passages for the pre, post, and delayed-post reading rate tests were also selected.
based on this estimated readability level (Flesch-Kincaid Grade Level 5.1), which resulted from the textbook analysis in the base study. Among them, three reading passages were chosen. The length of the three chosen texts were 242 (pre test), 244 (post test) and 242 (delayed-post test) respectively, and their Flesch-Kincaid Grade Levels were 5.2 (pre test), 5.1 (post test) and 5.3 (delayed-post test) respectively.

In order to improve content validity in the developed reading fluency test, all of the items or tasks on the test should be related to the skills associated with reading fluency [5]. As many L2 reading researchers such as Samuels [17] argue, rapid reading and accurate comprehension are the two main skills of reading fluency. Consequently, the developed reading fluency test covers two dimensions of fluency: (a) reading rate and (b) reading comprehension. Firstly, to measure the reading rate (the number of words per minute; WPM), Nuttall’s [13] reading rate assessment method was used. Also, in order to ensure the validity of the developed rate test, it needed to be designed to capture the actual skill levels of the participants. Secondly, to measure reading comprehension, follow-up questions were designed for each reading passage. According to Nuttall [13], “speed is worthless unless the reader understands what he reads”. After the reading rate passages were chosen, reading comprehension questions were developed for the reading comprehension. For each of chosen reading passages, 8 follow-up comprehension check-up questions in a multiple-choice format were developed. These questions were designed to ask participants to identify correct items such as the main idea or title of the reading passage. As Iwahori [10] points out, the main goal of these kinds of follow-up questions is to confirm that “participants read the tests with a certain degree of comprehension” (p. 78). These follow-up comprehension questions were used to encourage participants to read at a reasonable speed and to check whether they understood the text to an acceptable degree.

3.3 Pilot Studies

3.3.1 Participants

Thirty-six Korean high school students ranging from 15 to 16 years of age (M = 15.2 years) participated in pilot study 1 and pilot study 2. Half of these students, eighteen randomly selected students, participated in each pilot study. The average English proficiency of the participants was intermediate according to the mock College Scholastic Ability Test (CSAT) for English. The mock CSAT test is one of the most widely used placement tests in secondary schools in Korea [18] due to its practicality and accessibility. The majority of the participants had little opportunity to be exposed to English input other than during their English lessons at school since their school was located in a traditional rural area with relatively little access to English input.

3.3.2 Methodology

To check the internal reliability of the developed test, correlation coefficients were calculated. Three versions
Development of a English Learners’ Reading Fluency Test with regard to Pre/Post Test Design

(pre, post, and delayed-post test) of the reading fluency test were piloted on the same individuals and correlation coefficients calculated. In addition, to examine the validity of the developed test, the mean score of the reading comprehension test and reading rate test were calculated to see whether difficulties of reading texts and questions were appropriate for the participants.

3.3.3 Procedure

After development of a reading fluency test, the pilot study was conducted twice to examine the reliability and validity of instruments. In pilot study 1, the first 18 participants took the developed reading fluency test and their reading rates (WPM) were measured and their answers to the reading comprehension questions were graded. According to the results of pilot study 1, the developed test was adjusted. A week later, in pilot study 2, the other 18 participants took the adjusted reading fluency test to examine whether the adjustments made based on the results of pilot study 1 were appropriate for the target population.

4. Result

4.1 Pilot Study 1

In terms of the reliability of the developed test, the Cronbach’s Alpha reliability coefficient across the three developed tests (pre, post, and delayed-post tests) was .713. Detailed descriptions of the statistics of the developed fluency tests are presented in Table 1. Dörnyei [19] argues that internal reliability estimates for “well-developed scales ought to approach 0.8” (p. 207). However, he added that lower Cronbach Alpha coefficients are to be expected because of the complexity of second language reading. Although the estimated coefficient in pilot study 1 was not above 0.8, it can be said that these tests had relatively strong internal reliability, since items in each scale were alike in measuring over 0.70.

In the developed reading fluency test results, the reading comprehension section indicated that participants gained comparatively high scores as presented in Table 2 below. The test result was a mean of 88%. This revealed that the developed reading fluency tests (pre, post and delayed-post tests) were too easy for the participants.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Developed Reading Fluency Test Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>Cronbach’s Alpha</td>
</tr>
<tr>
<td>Developed Test</td>
<td>.713</td>
</tr>
</tbody>
</table>

Copyright © 2015 HSST
[Table 2] Mean Scores of Reading Comprehension

<table>
<thead>
<tr>
<th>Reading Comprehension</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed Pre_Test</td>
<td>80.6</td>
<td>18.70</td>
<td>18</td>
</tr>
<tr>
<td>Developed Post_Test</td>
<td>96.24</td>
<td>5.47</td>
<td>18</td>
</tr>
<tr>
<td>Developed Delayed-Post Test</td>
<td>88.13</td>
<td>10.95</td>
<td>18</td>
</tr>
<tr>
<td>Developed Test Total</td>
<td>88.32</td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

In the developed reading fluency test result the reading rate section showed that participants obtained the reading rates presented in Table 3. Nuttall [13] estimates that English learners may, on average, “read at 120 to 150 before training”. The mean reading rate in the pilot study 1 overpassed his suggested range. This result was aligned with the result of the reading comprehension section, which suggested that the developed reading fluency tests (pre, post and delayed-post tests) were too easy for the participants.

[Table 3] Mean Scores of Reading Rate

<table>
<thead>
<tr>
<th>Reading Rate</th>
<th>Mean (WPM)</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed Pre_Test</td>
<td>165.07</td>
<td>11.29</td>
<td>18</td>
</tr>
<tr>
<td>Developed Post_Test</td>
<td>177.63</td>
<td>38.23</td>
<td>18</td>
</tr>
<tr>
<td>Developed Delayed-Post Test</td>
<td>181.12</td>
<td>6.45</td>
<td>18</td>
</tr>
<tr>
<td>Developed Test Total</td>
<td>174.6</td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

Therefore, the current study increased the Flesch-Kincaid Grade Level of the reading passages in the tests from a mean of 5.2 to a mean of 7.5 to strengthen the validity of the test. Accordingly, passages with a slightly higher Flesch-Kincaid Grade Level were selected for the adjusted developed test. Based on the pilot study 1 result, the Flesch-Kincaid Grade Levels of the newly chosen reading passages were 7.5 (pre test), 7.6 (post test) and 7.5 (delayed-post test).

4.2 Pilot Study 2

For the reliability of the developed test, the Cronbach’s Alpha reliability coefficient of the developed test score was .802 as shown in Table 4. These results indicated that the adjusted reading fluency test had statistically significant internal consistency reliability among the pre, post, and delayed-post tests, since high internal reliability is considered to have been achieved when “[the] items … approach 0.8” [19].
Development of an English Learners’ Reading Fluency Test with regard to Pre/Post Test Design

[Table 4] Developed Reading Fluency Test Reliability

<table>
<thead>
<tr>
<th>Reliability</th>
<th>Cronbach's Alpha</th>
<th>Cronbach's Alpha Based on Standardized Items</th>
<th>N of Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed Test</td>
<td>.802</td>
<td>.854</td>
<td>3</td>
</tr>
</tbody>
</table>

With regard to validity, as shown in Table 5, the result of the reading comprehension section showed participants scores to be about 70%. Generally, a score of 70 percent indicates an appropriate readability level for participants in a fluency test [13]. This revealed that the developed reading fluency tests (pre, post, and delayed-post tests) were appropriate for the participants’ reading level.

[Table 5] Mean Scores of Reading Comprehension

<table>
<thead>
<tr>
<th>Score of Test</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed Pre_Test</td>
<td>69.0</td>
<td>25.34</td>
<td>18</td>
</tr>
<tr>
<td>Developed Post_Test</td>
<td>68.0</td>
<td>6.53</td>
<td>18</td>
</tr>
<tr>
<td>Developed Delayed-Post Test</td>
<td>72.0</td>
<td>22.45</td>
<td>18</td>
</tr>
<tr>
<td>Developed Test Total</td>
<td>69.66</td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

With further regard to validity, the result of the reading rate section showed that participants achieved a comparatively lower speed than in pilot study 1, as shown in Table 6. The test result fell within Nuttall’s [13] suggested reading rate, which is between 120 to 150 WPM. This revealed that the developed reading fluency tests (pre, post and delayed-post tests) were appropriate for the participants’ reading level.

[Table 6] Mean Scores of Reading Rate

<table>
<thead>
<tr>
<th>Reading Rate</th>
<th>Mean (WPM)</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed Pre_Test</td>
<td>144.11</td>
<td>11.29</td>
<td>18</td>
</tr>
<tr>
<td>Developed Post_Test</td>
<td>139.26</td>
<td>38.23</td>
<td>18</td>
</tr>
<tr>
<td>Developed Delayed-Post Test</td>
<td>129.79</td>
<td>6.45</td>
<td>18</td>
</tr>
<tr>
<td>Developed Test Total</td>
<td>137.72</td>
<td></td>
<td>18</td>
</tr>
</tbody>
</table>

5. Discussion and Conclusion

The current study developed a reading fluency test incorporating a pre/post test design to examine the effect of the intervention on English learners. The reliability and validity of the developed reading fluency test was then piloted. The result showed that the current study successfully increased the internal consistency of the test by ensuring the comparability of the reading difficulty of the texts used in the pre and post-tests. In addition, in order to assure the content validity of the developed test, the current study measured not only participants’
reading speed (reading rate) but also how much they understood the reading text (reading comprehension) in order to measure reading fluency.

The findings of the current study propose some useful suggestions in measuring reading fluency of English learners, particularly with relevance to pre/post test design. First, internal reliability should be considered of critical importance in pre/post test design since tests are conducted multiple times in this design. Accordingly, the different versions of the pre and post tests should be designed to be comparable in the pre/post test research design. Secondly, to test the reading fluency of English learners, both dimensions, reading rate and comprehension, should be measured. By measuring both aspects of reading fluency (i.e. reading rate and comprehension), researchers can obtain a more comprehensive understanding English learners’ reading fluency development. However, these findings need to be considered as tentative rather than conclusive because of the small sample size and limited research scope. Further studies should explore diverse factors of test reliability and validity such as face validity, construct validity, inter-rate reliability, and intra-rate reliability.
References


