<table>
<thead>
<tr>
<th>Title</th>
<th>Five year (consecutive) analysis of Japanese students' English proficiency using a large-scale in-house placement test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub Title</td>
<td>プレイスメントテストによる大学生の英語読解力測定 : 5年間の経年比較</td>
</tr>
<tr>
<td>Author</td>
<td>中村, 優治(Nakamura, Yuji)</td>
</tr>
<tr>
<td>Publisher</td>
<td>慶應義塾大学日吉紀要刊行委員会</td>
</tr>
<tr>
<td>Publication year</td>
<td>2011</td>
</tr>
<tr>
<td>Jtitle</td>
<td>慶應義塾大学日吉紀要, 言語・文化・コミュニケーション (Language, culture and communication). No.43 (2011.), p.37-49</td>
</tr>
<tr>
<td>Abstract</td>
<td></td>
</tr>
<tr>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>Genre</td>
<td>Departmental Bulletin Paper</td>
</tr>
</tbody>
</table>

The copyrights of content available on the Keio Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.
I. Rationale

Beginning in the spring semester of 2006 the Keio University Faculty of Letters has administered an in-house placement exam for incoming freshmen and new sophomore students. Students were given placement tests (hereafter PT) twice a year, once at the beginning of the academic year and again in the fall (confirmation test hereafter CT). The placement test aimed to measure students’ reading ability and overall English proficiency and to provide administrators with a quantifiable measure by which to give appropriate student instruction. Further, the test sought to optimize the student learning experience by better determining which areas of the multi-faceted English communicative skills needed to be strengthened.

Reading ability is thought to consist of grammar knowledge, vocabulary knowledge, long passage reading comprehension with full context (e.g. the text material does not have any deleted words or blanks intended for other questions), and passage understanding without sufficient context or information. In other words, a test of reading ability should be composed of grammar, vocabulary, reading comprehension and gap-filling items (cloze).

II. Purpose of the Study

The purpose of the present study is to report the findings from the five-year (consecutive) investigation of students’ English proficiency (i.e. grammar, vocabulary, cloze and reading comprehension).
III. Research Design and Test Method

a. Materials (Test Design and Content)

The test material, which was used for PT and CT, was based on the MELAB format. The test contained 15 grammar MC questions, 10 vocabulary MC questions, 10 gap-filling MC questions, and 3 long reading passages with 5 MC questions each. Applicants had 60 minutes to complete this test, which was scored by optical readers. The reading section consisted of one beginning level, one intermediate level and one advanced level passage; each passage consisting of ~400-500 words. Teachers rated test difficulty impressionistically in terms of content, topic, and vocabulary level.

The test format (for PT and CT) in Table 1 and the test content in Table 2 are as follows:

<table>
<thead>
<tr>
<th>Table 1 Test Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>No. of Items</td>
</tr>
<tr>
<td>Test format</td>
</tr>
<tr>
<td>Anchor items</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2 Test Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar</td>
</tr>
<tr>
<td>Vocabulary</td>
</tr>
<tr>
<td>Gap-filling</td>
</tr>
<tr>
<td>Reading Comprehension</td>
</tr>
</tbody>
</table>

Except PT1 (2006), each test was anchored by 12 items so that the five year (consecutive) information could be obtained. However, for the gap-filling section, because of its uniqueness, which requires the total integrative ability, it was difficult to utilize anchor items. Therefore, this section was not included in the five-year (consecutive) study. Also, the 2008 CT was not used for test equating and was excluded from the longitudinal analysis. Eventually, four pairs of test takers- (one: PT1 and CT1, another: PT2 and CT2, another: PT4 and CT4, and the last PT5 and CT5 )- were compared for the five-year (consecutive) analysis.
b. Procedures (Test construction, Administration, Timing)

Test Construction

The “reading ability” construct for this test was established mainly from specifications from the faculty of letters, which are as follows:
1) instructor teaching experience
2) reading sections of other existing tests
3) linguistic theories (e.g. Alderson, 2000; Grabe, 2000; Hughes, 2003)
4) academic and referencing needs of students’ areas of specialization when transferring to Mita campus (e.g. required reading ability at the Mita campus).
5) textbooks that were used in students’ university curricula.

The materials were selected using the following criteria:
1) grammar items were chosen from those grammar items that were introduced at the high school level. Ministry of Education authorized books, which were obtained from a bookstore were used to determine the appropriateness of the high school level classification. As items were not pre-tested to empirically determine difficulty, linguistic theory was relied on to create the varied ability levels. For example, vocabulary items were based on word frequency counts using the benchmark of English-Japanese dictionaries available at bookstores; the grammar items were based on developmental sequences and on written structures on textbook analysis. The textbooks that were used were authorized by the Ministry of Education, Sports and Science and were available at bookstores.

2) The reading passages were selected from three disciplines (i.e. humanities, social sciences and natural sciences), and appropriate vocabulary levels were taken into consideration. The text passages were analyzed using L1 Flesch Reading Ease (Readability Formula) together with the judgments of experienced instructors.

c. Subjects (Test takers)

Subjects were the entering freshmen from the years 2006 to 2011. Table 3 shows information about test takers and corresponding test forms used.
According to Table 3, for example, the test taker group of 2006 took both PT1 and CT1. For the intents of this study, in order to examine the change of the students’ reading ability between PT and CT, each group that took different tests were regarded as different test taker groups. Different test population data were provided accordingly.

The 2006 PT1 test taker group was operationally defined as the norm group in this study to investigate the students’ change in ability across four consecutive years.

d. Analyses

**Test Analysis**

The test data was analyzed using the Winsteps statistical program, the Xcalibre statistical program and the Bilog MG calibration program. The fit-misfit information was investigated to determine if the test results fit the model or not in the Rasch measurement analysis. The information about item difficulty and item discrimination was obtained in order to check each item in terms of classical test theory. The benchmark for the Cronbach alpha index of test reliability was set at 0.75 or over. In addition, the content validity is discussed by using questionnaire analysis. The Bilog MG was used to confirm that each item was functioning properly to obtain item information as well as test information.
IV. Results and Discussion

1. Descriptive statistics of each test

<table>
<thead>
<tr>
<th>Test Form</th>
<th>N</th>
<th>Mean/50</th>
<th>SD</th>
<th>Max</th>
<th>Mini</th>
<th>Grammar/15</th>
<th>Vocabulary/10</th>
<th>Gap-filling/10</th>
<th>Reading/15</th>
</tr>
</thead>
<tbody>
<tr>
<td>PT1</td>
<td>853</td>
<td>32.42</td>
<td>6.94</td>
<td>49</td>
<td>6</td>
<td>10.82</td>
<td>5.24</td>
<td>7.58</td>
<td>8.78</td>
</tr>
<tr>
<td>CT1</td>
<td>790</td>
<td>31.39</td>
<td>6.42</td>
<td>48</td>
<td>4</td>
<td>9.53</td>
<td>6.08</td>
<td>5.86</td>
<td>9.93</td>
</tr>
<tr>
<td>PT2</td>
<td>856</td>
<td>29.89</td>
<td>6.43</td>
<td>48</td>
<td>9</td>
<td>9.98</td>
<td>6.03</td>
<td>5.81</td>
<td>8.06</td>
</tr>
<tr>
<td>CT2</td>
<td>830</td>
<td>27.81</td>
<td>6.29</td>
<td>45</td>
<td>1</td>
<td>9.72</td>
<td>4.85</td>
<td>4.53</td>
<td>8.71</td>
</tr>
<tr>
<td>PT3</td>
<td>841</td>
<td>31.28</td>
<td>5.97</td>
<td>49</td>
<td>8</td>
<td>10.47</td>
<td>5.56</td>
<td>5.62</td>
<td>9.62</td>
</tr>
<tr>
<td>CT3*</td>
<td>794</td>
<td>31.72</td>
<td>6.14</td>
<td>47</td>
<td>11</td>
<td>11.07</td>
<td>5.61</td>
<td>6.22</td>
<td>8.81</td>
</tr>
<tr>
<td>PT4</td>
<td>830</td>
<td>32.19</td>
<td>6.82</td>
<td>47</td>
<td>10</td>
<td>10.24</td>
<td>6.61</td>
<td>6.03</td>
<td>9.31</td>
</tr>
<tr>
<td>CT4</td>
<td>768</td>
<td>28.71</td>
<td>6.57</td>
<td>47</td>
<td>10</td>
<td>10.29</td>
<td>5.13</td>
<td>4.61</td>
<td>8.67</td>
</tr>
<tr>
<td>PT5</td>
<td>816</td>
<td>33.20</td>
<td>6.84</td>
<td>49</td>
<td>9</td>
<td>11.62</td>
<td>7.25</td>
<td>6.41</td>
<td>7.90</td>
</tr>
<tr>
<td>CT5</td>
<td>764</td>
<td>29.1/49</td>
<td>6.51</td>
<td>46</td>
<td>8</td>
<td>8.38/14</td>
<td>6.35</td>
<td>4.94</td>
<td>9.48</td>
</tr>
<tr>
<td>RPT1</td>
<td>820</td>
<td>31.9</td>
<td>6.41</td>
<td>48</td>
<td>6</td>
<td>10.91</td>
<td>5.70</td>
<td>5.38</td>
<td>9.45</td>
</tr>
</tbody>
</table>

N.B. In CT5 one grammar item was invalid, so the subtotal number for grammar is 14, and the total number of the whole test was 49.

2. Examination of reliability

The reliability was investigated by the Cronbach Alpha. The benchmark for the acceptable boundary is over 0.75. The reliability of placement tests had scores over 0.75 in terms of the Conbach Alpha index. This suggests the items in each test were internally consistent. Table 5 shows the Cronbach Alpha index in each test form.

<table>
<thead>
<tr>
<th>Test form</th>
<th>PT1</th>
<th>CT1</th>
<th>PT2</th>
<th>CT2</th>
<th>PT3</th>
<th>PT4</th>
<th>CT4</th>
<th>PT5</th>
<th>CT5</th>
<th>RPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach α</td>
<td>0.83</td>
<td>0.78</td>
<td>0.77</td>
<td>0.75</td>
<td>0.76</td>
<td>0.81</td>
<td>0.78</td>
<td>0.82</td>
<td>0.78</td>
<td>0.78</td>
</tr>
</tbody>
</table>

3. Comparison of each subtest (PT and CT)

Figure A shows the 5 consecutive year analysis of the estimated mean and standard deviation (sd) of theta in three subcategories (i.e. grammar, vocabulary, and reading). Since the project started in April 2006, PT1 (2006/04) is always the benchmark for the data analysis.

One interesting finding was that when grammar and vocabulary knowledge did
not improve within the new curriculum, students’ reading ability did improve. There was a continual improvement between PT and CT. Improvements between PT2 and CT2 should also be noted. The current findings may also suggest that there has been a teaching effect between pre-teaching (at the PT time) and the post-teaching (at the CT time).

Let us now examine the longitudinal change in PT test results of entering students’ proficiency (grammar, vocabulary and reading) for 6 years (2006 through 2011)—(PT1, PT2, PT3, PT4, PT5, RPT 1). The grammar section indicates that with the exception of PT 4 students, all other entering students rated higher than the benchmark.

The vocabulary section is slightly different from the grammar section. PT4, PT5 and RPT1 show an improvement over the benchmark PT1; however, PT grammar remained relatively close to the benchmark, and interestingly, PT4 shows a negative improvement.

The reading part always showed a positive (i.e. improved) result when compared to the benchmark.

Secondly, we will examine paired results between PT and CT (PT1 and CT1, PT2 and CT2, PT4 and CT4, PT5 and CT5) to explore the teaching and learning effect of the same groups of students within a year.

In reference to PT1 and CT1, although there was no quantitative difference between the grammar sections, the vocabulary score decreased while the reading
score increased. In other words, reading ability showed quantifiable improvement.

In the comparison between PT2 and CT2, while vocabulary and grammar scores decreased, reading scores increased. Here again, students’ reading ability improved.

Between PT4 and CT4, grammar and vocabulary scores decreased whereas reading score increased. Once again the reading score showed an improvement.

In the comparison of PT5 and CT5, while the vocabulary score shows no difference, the grammar score decreased; however, the reading score increased. Reading ability exhibited a quantifiable improvement.

In summary, the general patterns of the paired analyses show that while the grammar and vocabulary scores decreased in most cases, reading scores consistently increased. One possible explanation for this phenomenon could be attributed to the new curricula’s’ reading focus.

Students were expected to perform English reading activities in all English classes that were taken, so opportunities for reading English were secured.

Another potential explanation could be that typically teachers expose students to reading in English even if the eventual goal is oral communication.

Still another possibility is that the reading-focused project (i.e. curriculum) has been successful in enhancing students’ reading ability.

Still another possibility is that CT tests played an important role in motivating students because the CT test results were used for the placement purposes for the following year.

As for the explanation of the deterioration of grammar and vocabulary skills, compared with their high school classes where students had vocabulary building classes or grammar-centered composition classes on a regular basis to prepare for entrance exams, there were no such classes in college. Therefore, their vocabulary storage and grammatical knowledge may have potentially decreased since the entrance exam, when their grammatical knowledge was at its peak.

4. Comparison of each subtest (consecutive four year analysis)

Note: When we compare the test results on a year basis, the benchmark basis is always the results of 2006 placement test results.

Figures G1, G2 and G3 show two things: 1. the difference between the placement test of grammar and the confirmation test of grammar in the 2006, 2007 and 2009 academic years; and 2. the change of the five year (consecutive) comparison (i.e. 2006-2011). This comparison is based on the scores from each year’s placement tests.
Figures G1, G2, G3: Grammar comparison

Figure G1

Figures G1, G2, and G3 show that there was little change in the students’ grammar ability between the PT and the CT in each academic year. Unlike high school education, there were no English grammar courses or classes in university education.
and as a result it is not surprising that students’ grammatical abilities would not improve significantly. It should be noted that freshman students’ grammatical abilities, peaked during university entrance examination season. University education makes a contribution to sustain basic grammatical abilities in student courses.

Although the findings did not positively suggest that placement test results for the last four years showed an increase for entering students’ grammatical abilities, the pre-2010 test results indicated a greater standard deviation of the students’ grammatical abilities. In other words, students’ grammatical ability level showed a wider range than in previous years.

Figures V1, V2 and V3 suggest that there was no distinct increase in the results of the vocabulary section when results were compared. Moreover, students’

---

Figures V1, V2, V3: Vocabulary section

Figure V1

![Figure V1](image)

Figure V2

![Figure V2](image)
vocabulary abilities actually exhibited annual declines for each individual academic year. This lead the researchers to believe that because there were no specific vocabulary building courses during university, students’ crammed vocabulary knowledge peaked for the purposes of taking the entrance examination, and after entering university the retention of the learned vocabulary gradually waned. Student vocabulary that focused on textbook reading improved even if the width of the vocabulary did not show a significant increase.

The data suggests that entering students' vocabulary knowledge improved for the last four years.

Figures R1, R2 and R3 show that there was a visible change not only in the teaching or learning effect of reading, but also in the entering students’ reading.
ability. In 2006 there was an improvement in students' reading ability between their placement test and confirmation test results. Also, in 2007 there was an increase in scores between the placement test and confirmation test. Furthermore, in 2009 there was another improvement in students' reading ability between their placement test and confirmation test. It can be said that the curriculum change, which focused primarily on enhancing students' reading ability, has been successful in this regard. One hypothesis is that freshmen students, after taking general education courses, increased their background knowledge (schemata) of subject matters in each course. As a result of students’ newfound knowledge the subject matter (in other words, test topics) in the reading tests could begin to resemble those in the general education courses. In short, the increase of the background knowledge in the education courses
can affect the students' improved reading ability.

Still another possible explanation could be that unlike high school education where test taking strategies are likely utilized, university English education stresses a more thorough understanding of the text, which focuses on student analytical reading ability.

Figures R1, R2 and R3 also indicate that entering freshmen students' reading ability improved. One possible explanation for this was that placement test information, which students have access to, could have a positive impact on those who wish to enter Keio University (Faculty of Letters).

V. Conclusions and Implications

Overall the general pattern of the paired analyses show that while the grammar and vocabulary scores decreased in most cases, the reading scores always increased. One possible explanation was that there was a learning and/or teaching effect on reading because the new curriculum focused on fostering students' reading ability. The reading-focused project (curriculum) has been successful in enhancing students' reading ability.

Considering McNamara's (2000, p.83) statement “The right balance of three basic critical dimensions of tests—validity, reliability and practicality—will depend on the test context and test purpose,” the present placement test should be regarded as acceptable, judging from the statistical analyses and the test context as well as the test purpose. For future improvement, predictive and concurrent validity should be measured. Further, multi-trait multi-method (MTMM) analysis could also offer valuable insights into the potential merits of test validation. Future studies should also explore the issue of face validity and practicality more systematically. For this purpose, an external benchmark such as the Common European Framework of Reference for languages teaching and learning assessment (CEFR) could be a guideline.

Acknowledgement

I would like to thank all the members of the Keio Placement Test Research Group for their cooperative work editing, as well as administering the tests. The members are as follows: Andrew Armour, Yuji Nakamura (head author of this paper), William Snell, Yoshiko Uzawa, Kenji Adachi, Hikaru Sakamoto, Yoko Hemmi, Nobuya Takahashi, Kyoko Yoshida, Satoko Tokunaga, and Yuichi Akae. I am also grateful to Mr. Haruhiko Mitsunaga for his statistical help.
References


