【Research Report】

Evaluating the Validity of In-house Developed English Placement Tests for First-year Students at a Japanese University

Jeffrey TRAMBLEY and Jonathan CHERRY

【研究ノート】

大学の１年生向け英語プレースメントテストの評価

ジェフリー トランブリー、ジョナサン チェリー

【キーワード】Key Words： language assessment, placement test, test validity, language test construction, language proficiency test

Summary

Accurate placement of students into level-appropriate English classes is essential to ensure that students receive instruction aligned with their proficiency levels when beginning their university studies. In the past, the placement test given to the department's first-year students showed evidence of invalidity, with instructors frequently suspecting that students were placed in levels which did not match their abilities. This led the researchers to evaluate the currently used placement test for validity by comparing the results with those of the TOEIC Bridge test. At the same time, a new placement test was also developed and administered to students and evaluated for validity against the TOEIC Bridge test. After administering the original placement test to incoming students before the first semester began in April 2025, the TOEIC Bridge institutional test was given to students approximately halfway through the semester. This was followed by the new placement test a week later. Once results for all three tests were compiled, the results of the original placement test and the new placement test were compared to assess correlation to the TOEIC Bridge test. A total of 48 students who completed each of the tests were included in the analysis. The analysis showed that the TOEIC Bridge test results aligned most closely with the original placement test, with a Pearson Correlation of 0.810. The new placement test results showed less correlation to the TOEIC Bridge test, with a Pearson Correlation of 0.56. While the correlation was strongest for the original placement test, there were limitations. It was noted that using the TOEIC Bridge results to stream students, the number of students who would have remained in the same level was more similar to the streaming levels using the original placement test. Using the new placement result to stream students resulted in more students changing levels and fewer remaining in the same level. In the future, we plan to explore other aspects which may be affecting test results such as test-taking environment, test fatigue, and implementation in digital

versus paper-based formats.

1．Introduction

The importance of placing first-year students into appropriate levels for streamed English classes cannot be underestimated. Students placed in classes with others of similar abilities can make the classroom environment more conducive to social learning as well as facilitate more effective teaching for instructors. While many universities use commercially available tests such as the TOEIC test as a placement tool, budget and time constraints prevent many institutions from using them. This is the case for Musashino Gakuin University's Department of International Communication. Due to both budget and time constraints, each year, the department's English faculty is tasked with implementing and delivering English class placement assignments for incoming students to the academic affairs office within one day of students beginning orientation prior to the start of the semester. As a result, the faculty has opted to give an in-house produced placement test.

For the past decade, we have used a placement test developed by the researchers to fulfill this task. However, over the years, English instructors have often doubted the efficacy of the placement test, citing anecdotal evidence of students ending up in levels seemingly not matching their abilities. Evidence points to the fact that students have been placed in levels much higher than their abilities in many cases, while the opposite case, students being placed in levels too low for their abilities, has also been observed. In light of such concerns, the researchers undertook the task of evaluating the validity of the original placement test using statistical analysis to compare test results to the TOEIC Bridge test. At the same time, a new placement test was developed and tested for validity.

2．The Three Tests and the Testing Schedule

The online original placement test (OPT) was given to 62 first-year students using Google Forms as well as a handwritten, paper-based writing sample. First, students created writing samples in 10 minutes in English on the topic of their self-introduction. The handout for students can be found in the appendix. Students were not allowed to use any technological devices while writing and submissions were collected at the end of the allotted 10-minute time period.

Next, test-takers completed the online portion of the placement test in 30 minutes using Google Forms. They were shown a QR code to access the test. The contents of the online portion of the OPT is outlined below in Table 1.

Table 1. Contents of Original Placement Test (OPT) with Writing Sample

Section	Question Type (focus)	Number of Questions	Total Points	Percentage of Test
1	Question-Response (grammar & vocabulary)	10	12	20%
2	Sentence Completion (grammar & vocabulary)	12	18	30%
3	Reading Passages 1-4 (comprehension)	13	18	30%
4	Writing Sample (willingness to communicate & accuracy)	1	12	20%
	Total	36	60	100%

Regarding the test content, for each question in section 1, three answer options were given. In sections 2 and 3, four answer options were given.

To grade the writing sample, the two researchers read each submission and gave scores according to the following rubric: willingness to communicate, 6 points, and accuracy, 6 points. Willingness to communicate was assessed by the test-takers' effort to share information about themselves. For example, if a student included information on several topics in their self-introduction, though with limited success regarding correct grammar or spelling, they were awarded a higher score for willingness to communicate, but a lower one for accuracy. These scores were later added to the scores which students received from the online component.

Several weeks later, on May 14, 2025, a total of 57 students took the TOEIC Bridge institutional test on the university campus. The test included both a listening and reading section as usual, and was given in paper-based form. The official results were obtained by the university several weeks after the test was implemented.

Finally, the new placement test (NPT) was given to students a week after the TOEIC Bridge test, on May 20, 2025. Students took the test in their regular English classes which were held during the same period on the same day. Each instructor was sent a link with a QR code through which students accessed the test on Google Forms. A total of 48 students completed the NPT which took approximately 30 minutes.

The NPT contents was based loosely on a framework developed by Nihon University's Department of International Liberal Arts for a placement test during online learning caused by the COVID-19 pandemic (Irwin et al, 2022). Since the original framework included a listening component, it was replaced with the writing sample. The NPT contents are outlined in Table 2. For all sections of the test, four answer options were given.

Table 2. Contents of the New Placement Test (NPT) with Writing Sample

Section	Question Type (focus)	Number of Questions	Total Points	Percentage of Test
1	Sentence Completion (vocabulary)	14	80	32%
2	Sentence Completion (grammar)	28	120	48%
3	Reading Passage 1 (comprehension)	3	12	5%
4	Reading Passage 2 (comprehension)	4	28	10%
5	Writing Sample (willingness to communicate & accuracy)	1	12	5%
	Total	50	252	100%

Other than content differences, the scoring construct used for each test was different. For the OPT, questions were categorized as either basic, intermediate, or advanced. The levels were decided by the researchers thinking about the level of students which generally enter the university. For basic items, 1 point was given, for intermediate, 2 points, and for advanced items, 3 points.

For the NPT, each item was scored according to the same three levels of difficulty: basic, intermediate, and advanced. These levels corresponded to CEFR levels. Basic items were between CEFR A1-A2, intermediate items were between A2-B1, and advanced items were B1 or above. The scoring constructs for both the OPT and the NPT are outlined below in Table 3 and Table 4.

Table 3. Scoring Construct for OPT without Writing Sample

Item Level	Number of Questions	Points per Question	Total Points
Basic	23	1	23
Intermediate	11	2	22
Advanced	1	3	3
Total	34		48

Table 4. Scoring Construct for NPT without Writing Sample

Item Level (CEFR level)	Number of Questions	Points per Question	Total Points
Basic (A1-A2)	20	3	60
Basic (A1-A2)	3	4	12
Intermediate (A2-B1)	16	7	112
Advanced (B1+)	7	8	56
Total	46		240

3．Data Analysis

Once results of the three tests, OPT, NPT, and TOEIC Bridge, were compiled, correlation analysis was carried out. The number of students who had completed all three tests was 48, so only these students' data were included in the analysis.

Pearson correlation coefficients were calculated to determine the strength of the relationship between the OPT and NPT compared to the TOEIC Bridge scores. The results are summarized in Table 5.

Table 5. Correlations between Placement Test Scores and TOEIC Bridge Scores (n = 48)

Test Version	Pearson r	p-value
OPT with Writing Sample	0.810	<0.001
OPT without Writing Sample	0.780	<0.001
NPT without Writing Sample	0.553	<0.001
NPT with Writing Sample	0.586	<0.001
NPT without 8-point Questions, without Writing Sample	0.623	<0.001

The results indicated that the OPT with the writing sample showed the strongest correlation to the TOEIC Bridge scores (r = 0.810), followed closely by the OPT without the writing sample (r = 0.780). In contrast, the NPT variants showed weaker correlations, although improvements were seen when the writing sample was added (r = 0.586) and when higher-weighted 8-point questions were removed (r = 0.623).

A heatmap was generated to visualize the relationship between variations of the two placement tests and the TOEIC Bridge test scores. As shown in Figure 1, the OPT scores are closely grouped together, while the NPT scores show more variation. This supports the conclusion that the OPT better aligns with the TOEIC Bridge test and is a more reliable placement tool.