

Annual Review of “The Freshman English Placement Test” at Asia University 2019

Kim Pollard, Asia University

Abstract

This article examines whether the Freshman English Placement Test (FEPT) taken by students at Asia University in Tokyo, is still fit for purpose. FEPT test results were statistically analysed in three key areas: item difficulty, reliability and item discrimination. A key finding is that parts of the test continue to be too easy and fail to discriminate effectively between test takers. In the long term, a rewrite of the test based on the Common European Framework of Reference for Languages (CEFR) is proposed.

Introduction to the 2019 FEPT

The Freshman English Placement Test (FEPT) was used to stream 1127 Freshman English (FE) students from five faculties into compulsory English classes at Asia University in Tokyo. Following the format of previous articles (Pollard 2019; Bates, 2018; Mabe, 2017), the continued viability of the FEPT will be discussed by analysing test reliability, test item difficulty and test discrimination data from 2017-2019. Factors to be considered in future rewrites and modifications of the current test will be proposed.

The test, written by the Centre for English Language Education (CELE) teachers, is administered at Asia University at the beginning of each academic year. It comprises of 74 multiple choice questions which are divided into four sections: Listening, Vocabulary, Grammar and Reading. The test uses the Scranton format and is computer marked. The results for this paper have been analysed using SPSS.

Comparisons and Analysis of FEPT Results 2016-2018

Mean and Standard Deviation

As Table 1 shows, the standard deviation has continued increasing year on year (Bates 2018; Mabe, 2017; Carpenter, 2016). Standard deviation shows how scores are spread out from the mean so the higher the standard deviation, the higher the spread. If all scores are placed close to the mean, then the test is failing to differentiate students effectively (Alderson, Clapham & Wall, 1995). The 2019 standard deviation of 11.47 suggests that there is more spread from previous years but that students are similar in ability (Bates 2018; Mabe, 2017; Carpenter, 2016). This supports anecdotal evidence that most students at Asia University are Beginner, Elementary and Pre-intermediate on the Common European Framework of Reference (CEFR) scale.

Table 1
FEPT Mean and Standard Deviation

FEPT Year	Number of Items	No. Examinees	Mean	Std. Deviation
2019	74	1127	48.3	11.47
2018	74	1311	41.3	10.4
2017	74	1415	40.4	10.1
2016	74	1445	39.3	9.7

Test Reliability

One measure of test reliability is Cronbach's alpha which determines whether test items are consistently testing for the same thing. This alpha also shows whether the test performs consistently with different students each year. A high coefficient indicates greater likelihood that the test items are testing the same thing. If the theoretical alpha ranges from 0-1, a score over 0.7 is deemed acceptable for a homemade test (George & Mallery, 2003). Analysis of the 74 test items showed Cronbach's alpha to be $\alpha = 0.89$, a slight improvement on the previous year's $\alpha = 0.86$. The test can therefore be deemed reliable. It is important to note, however, that reliability does not mean that the test is valid (Hull 2012, p. 4).

Item Difficulty

The main aim of the FEPT is to discriminate between students' English language abilities. Analysing Item Difficulty is another way to assess the test's ability to do this. A test item is deemed too difficult if the mean score is below 0.3 as most of the test takers, including high scorers, got the answer wrong. Conversely, a score above 0.7 means the test item was too easy, even for low scoring students. Table 2 below compares the percentage of unsatisfactory performance in item difficulty for the last three years.

Table 2

	2019 (%)	2018 (%)	2017 (%)
Listening			
Part 1	50	50	50
Part 2	28.5	14.3	42.9
Part 3	30	40	50
Part 4	14.28	14.29	14.3
Vocabulary: Part 5	29.4	47.06	31.3
Grammar: Part 6 A (Gap fill)	42.8	42.9	28.6
Grammar: Part 6 B (Find the mistakes)	40	40	20
Reading: Part 7	33.3	33.3	0

Table 3 shows the percentage of questions deemed too easy and too difficult in each section of the 2019 test.

Table 3	2019 too easy %	2019 too difficult %
Listening		
Part 1	50	0
Part 2	28.5	0
Part 3	30	0
Part 4	7.1	7.1
Vocabulary: Part 5	29.4	0
Grammar: Part 6 A (Gap fill)	28.6	14.29
Grammar: Part 6 B (Find the mistakes)	40.0	0
Reading: Part 7	33.3	0

As in previous years, many questions are too easy for test takers which has implications for the ability of the test to differentiate between students. This is also reflected in the increase in mean score. Once again, Part 1 Listening stands out as 50% of this section is too easy. However, this may not be too problematic as easy questions are useful because

they can provide a good lead in for students and help put them at ease (Heaton 1989, p. 179). Questions should, however, become progressively harder (Carpenter, 2016).

As discussed in a previous paper (Pollard, 2019), anecdotal evidence suggests that most students come from a predominantly Grammar Translation pedagogy at Junior (JHS) and Senior High School (SHS). With this style of pedagogy, greater emphasis is placed on grammatical forms and accuracy over fluency in JHS and SHS examinations and tuition (McNamara & Rover, 2006; Sato, 2002). This may partly explain why students perform particularly well in Grammar sections A and B and why this part of the test consistently performs poorly at differentiating students. A third of the reading section is also too easy, continuing the trend from last year. Although there were some difficult items which fail to discriminate effectively amongst most students, they are still useful for distinguishing between good and very good students (Heaton 1989, p. 179).

Item Discrimination

Another way to determine if a test is functioning effectively is to analyse Item Discrimination. This index shows “the extent to which the item discriminates between the testees, separating the more able from the less able” (Heaton, 1989, p. 179). An index above 0.3 shows that the test item is discriminating effectively between higher and lower level students. An index below 0.3 indicates that the item discriminates poorly between test takers. A minus score shows that the lower level students answered the question correctly, but higher-level students did not. The table below shows the percentage of scores under 0.3 for each section of the FEPT over the last three years.

This year, only 53% of the final test discriminated effectively between students. In the long term, therefore, some sections of the test needs to be re-written so that they perform better. Based on the results in Table 4, it seems that the reading and vocabulary sections are performing well and do not need re-writing for the time being. However, Listening Parts 1, 3 and 4 discriminated between students poorly.

Table 4

Item discrimination % of questions
that performed poorly

	2019	2018	2017
Listening			
Part 1	50	50	75
Part 2	28.5	42.9	100
Part 3	80	70	80
Part 4	71.4	57.1	71.4
Vocabulary: Part 5	29.4	35.3	56.3
Grammar: Part 6 A (Gap fill)	57.1	57.1	57.1
Grammar: Part 6 B (Find the mistakes)	40	80	80
Reading: Part 7	0	33.3	16.6

Table 5

Sections that performed poorly in terms of difficulty and discrimination (%).

	Difficulty	Discrimination
Listening		
Part 1	50	50
Part 2	28.5	28.5
Part 3	30	80
Part 4	14.28	71.4
Vocabulary: Part 5	29.4	29.4
Grammar: Part 6 A (Gap fill)	42.8	57.1
Grammar: Part 6 B (Find the mistakes)	40	40
Reading: Part 7	33.3	0

Problematic Questions

Removing or changing problematic questions would improve the quality of the test and possibly improve Cronbach Alpha. Questions deemed problematic are those which fail in terms of difficulty and discrimination. Year on year the following questions have performed poorly and so should be altered or removed; 6, 8, 37 and 58.

As Heaton (1989, p. 182) states, it is important if replacing question to ask why students failed to answer these questions so that a suitable alternative can be found. As has been discussed extensively by Mabe (2017) question 6, a listening question, does not test for phonemes as it was intended to. In the example below, moat, vote and slope are difficult distractors, which low level students are unlikely to choose by default (Mabe, 2017).

Example:

Q6. You should put a _____ on. a) boat b) vote c) moat d) coat e) slope

More challenge could be added by using words that are more difficult for students to differentiate such as “won’t” and “want” but are likely to be familiar to them.

Listening question 37 is probably too difficult for most test takers due to the high speed of the dialogue delivery and the large amount of information students are expected to assimilate in a short time frame.

Part 5 Vocabulary continues to be a particularly problematic section. As most FE students are low level students it would be better to ask questions that can differentiate students at that level. This section, however, asks students to find the opposite of a word from a choice of four other low frequency, higher level words. In addition, Question 54 uses phrasal verbs that lower level students probably will not have studied yet (Mabe, 2017). In addition, this section does not test for understanding or ability to use said vocabulary.

Problematic Question Examples:

Q. 53. Find the opposite of practical a) accidental b) impractical c) intentional d) imperial.

Q 54. Find the opposite of continues a) break off b) carry on c) start up d) start off.

Test Validity

It is important to ask whether the test is valid, that is, whether it is measuring what it is supposed to measure. The test is designed to stream students into communicative English classes whose core focus is on speaking and listening skills. As there is no speaking component to the test the validity of the test can be called into question. The test also does not test language in a communicative way. A more communicative approach to teaching English in JHS and SHS has been encouraged by the Ministry of Education and is, based on anecdotal evidence, being more widely adopted, (e.g. the introduction of the nationwide

Leaders of English Education Project (LEEP) by British Council). The test should therefore aim to reflect this trend in education or risk becoming too easy and less valid (Nagata, 1995).

Secondly, the test does not take into consideration the language taught or curriculums of the FE course. As mentioned by Pollard (2019), Bates (2018) and Mabe (2017), the range and level of language structures tested is extremely limited. Many questions repeat the same language point, for example, prepositions of place and direction. Also, only receptive skills are being tested, unlike the FE course which teaches both receptive and productive skills (Mabe, 2017). Students are also unable to ascertain their progress when they repeat the test at the end of the year.

The current curriculums on the FE course are based around the *Four Corners* series which incorporates the CEFR into its design. This framework is also used by multiple language institutions such as British Council. It is an internationally recognised standard used around the world to demonstrate language ability. The CEFR-J, a Japanese version of the framework, is currently being rolled out in JHS by the Ministry of Education. Using this framework, students can monitor their progress against a list of “Can Do” statements. This can be used at every stage of their English language learning journey from Beginner to Advanced Plus. If the FEPT could be re-written, it would be beneficial to follow this framework as students could monitor their progress from the beginning of the year to when they take the test again.

Conclusion and Recommendations

As it stands, the test is still able to differentiate between students adequately. However, parts of the test continue to be too easy and fail to differentiate students as effectively as they could. In the short term, the four questions identified should be replaced. In the long term, however, CELE should consider replacing the test so that it fits the curriculum and wider changes in Japanese English education. Replacing it with a test that utilises the CEFR framework would enable students to monitor their progress and use their scores as a base level for further English study after university. A more communicative, standardised approach that tests both receptive and productive skills would enable students to track their progress more effectively.

References

- Alderson, J., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge Language Teaching Library. Cambridge: Cambridge University Press.
- Bates, D. (2018). An Analysis and Review of the 2017 Freshman English Placement Test at Asia University. *CELE Journal*, 26, 1-9.
- Brown, J. (2002). English Language Entrance Examinations: A Progress Report, Curriculum Innovation, Testing and Evaluation. *JALT 1st Annual PAN-SIG Conference*, 95-112.
- Carpenter, J. (2016). Past, Present and Future Placement Testing Practices at CELE: A View from 2015. *CELE Journal*, 24, 52-77.
- George, D., & Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Heaton, J.B. (1989). *Writing English Language Tests*. Longman Handbooks for Language Teachers. London: Longman.
- Hull, J. (2012). Modifying Asia University's Freshman English Placement Test, *CELE Journal*, 20, 1-11.
- Mabe, K. (2017). Review and Analysis of Asia University's 2016 Freshman English Placement Test: The Need for Major or Minor Change? *CELE Journal*, 25, 1-16.
- McNamara, T & Rover, C. (2006). *Language Testing: The Social Dimension*. Oxford: Blackwell.
- Nagata, H. (1995). Testing Oral Ability, ILR and ACTFL Oral Proficiency Interviews, Language Testing in Japan, *JALT Applied Materials*, 12, 108-118.
- Pollard, K (2019). The Freshman English Placement Test at Asia University 2018: Still Viable ? *CELE Journal*, 27, 141-149.
- Sato, K. (2002). Practical Understanding of Communicative Language Teaching and Teacher Development). In S.J. Savignon (Ed.), *Interpreting Communicative Language Teaching* (pp. 41-81). New Haven, CT: Yale University Press.