Seeking the Benefits of E-testing in Assessing Students' English Proficiency

Mutsuko NAGASAKI, Akira NAKAYAMA, Sunao ORIMOTO, Neil HEFFERNAN

Ehime University

1. Introduction

The English Education Center (EEC) at Ehime University started using an official standard English proficiency test in 2007. We used the Global Test of English Communication (GTEC) for Students from 2007 to 2009, and since then have been using the Test of English for International Communication (TOEIC) Bridge. Both tests are paper and pencil tests (PPT), but the EEC began to explore the possibilities of using e-testing as a means of measuring university students' English proficiency. This paper aims to explore the possible benefits of e-testing, focusing on two e-tests; the Computerized Assessment System for English Communication (CASEC) and the GTEC College Test Edition. A trial of these tests was conducted to examine the possibilities of implementing them in the EEC curriculum. The results of the trail are reported and some suggestions for their use are made at the end of the paper.

2. Why e-testing?

The EEC provides about 2,000 first year students with an official standard English proficiency test twice every year at the beginning of June and December. The aims of the usage of the standard test are first to establish a higher validity of students' grade evaluation, and secondly to obtain the appropriate objective data according to which the EEC streams its English classes for the second semester. Twenty percent of student grade evaluation is based upon the results of this test, and two tests (GTEC for Students from 2007 to 2009 and TOEIC Bridge from 2010 until now) have been used. These tests were chosen from the perspective of both their appropriate levels for our first year students and also for their cost.

Although both tests served well enough to attain the initial aims of the EEC's programs, they have caused some problems in terms of operation. The biggest problem is that they both require manpower. In order for the approximately 2,000 first year students to take the test at the same time, about 50 members of faculty, 35 office staff, 35 student assistances (SAs) and 5 part-time assistants are asked to work as proctors or assistant proctors and receptionists on the day of the test. It especially requires great manpower to collect question and answer sheets and carefully count them in order to avoid the leakage of exam questions. Careful and adequate preparation is necessary to secure these testing staff and enough classrooms, which demands considerable expense in terms of time and effort for the office staff.

Yet another difficulty is in scheduling the test itself. It takes GTEC for Students about 3 weeks and TOEIC Bridge about 10 days until the students' score reports are returned to the university. As previously mentioned, since 20% of students' final grade is decided by one of these official standard English proficiency tests, the test has to be conducted at least a few weeks before the end of the semester in order for the scores to be available to teachers. Consequently, it has been held in the middle of each semester since 2007. As a result, regular classes held at the same time as the test need to be canceled. The scheduling is always quite a difficult and controversial issue at the university.

Furthermore, since the test is conducted in the middle of the semester, the students' proficiency has to be evaluated much earlier than would be ultimately desirable. This is problematic in terms of giving a fair and accurate evaluation to students.

One way to solve these problems is the use of an e-testing system with which students take a test online or on a CD-ROM using a computer. Under this system,

they can take the test on their own, following directions shown on the computer screen. It also omits the whole processes of distributing and collecting test materials and counting them. This requires fewer proctors, which will result in less effort on the human resources side of the equation.

Another advantage of e-testing is that it provides test scores immediately after the test. It allows flexibility in making the test schedule and helps solve the scheduling problem. In addition, this flexibility will contribute to being able to more fairly evaluate students' English proficiency, since it is possible to have the test towards the end of the semester.

3. E-tests selected for trail

3. 1 Selection

In order to improve the above situation, the EEC has begun to consider ways to promote more efficiency in the process of carrying out our English proficiency test, and thus, in 2012, conducted some research in order to find an alternative official standard English test. Among several e-tests, we nominated CASEC, GTEC College Test Edition, Versant Speaking test, and the TOEIC Speaking and Writing (SW) test, as prospective tests, after reviewing the content, level and cost of each test. Both the Versant Speaking test and the TOEIC SW test were refined and well-organized tests, but they are very challenging and may not be appropriate for low level students; the approximate English proficiency level of most Ehime University students is considered to be low-to-high intermediate. Another problem is that the above-mentioned tests are designed to measure speaking proficiency only, while the current TOEIC Bridge is designed to measure listening and reading proficiencies. In addition, they cost more than the TOEIC Bridge, which is another issue that needs to be considered. CASEC and GTEC College Test Edition[1] seem to satisfy all the conditions that we require at the EEC, and also eliminate the problems mentioned above. Above all, they are computer adaptive tests (CAT) based on item response theory (IRT)[2], which can measure test takers' English proficiency efficiently within a limited time.

The use of CAT is expected to solve some problems inherent in such tests as the TOEIC Bridge. That is, the problem of all students having to take the test at the same time is avoided, therefore allowing more flexibility in terms of human resources and time. In CAT, the order of test items presented on the screen is decided by test taker's responses. If they answer incorrectly, a slightly easier question will be given next. If they give a correct answer, the next item will be slightly more difficult. That is, according to learners' responses, items are chosen from an item bank in order to accurately reflect their level of proficiency.

In sum, the use of an e-test will undoubtedly suit the EEC for its testing purposes in that we will no longer have to set up one specific test time and date on which all the students have to sit for a test; and the leakage of exam questions will become a moot point. If several test dates are set, students can take it during a specific period in a computer room following the directions on the screen by themselves. It will contribute to flexibility in test scheduling.

3. 2 Two Candidates: CASEC and GTEC College Test Edition

Table 1 and 2 summarize the content of the CASEC and GTEC College Test Edition. As Table 1 shows, CASEC is composed of four sections, with the first two parts assessing knowledge of vocabulary or set phrases which are frequently used in daily life. Sections 3 and 4 assess learners' listening ability to understand main ideas as well as specific information given in daily situations. Though Sections 1 to 3 are multiple choice questions, Section 4 requires test takers to physically type their answers. Hence, accuracy of spelling is necessary to give a correct answer.

Regarding the GTEC College Test Edition (Table 2), its components and contents are quite similar to those of the TOEIC Bridge test. It has 2 sections, listening and reading, and each is composed of 3 parts. It evaluates learners' receptive skills.

The other characteristics of CASEC and GTEC College Test Edition, such as computer hardware requirements, score reports, and the price are summarized in Table 3.

Table 1 : *Summary of the CASEC test*

	Purpose and Focus of Questions	Format of Questions and Answers	Number of Questions	Time	Scores
Section 1	To assess the knowledge of vocabulary which are frequently used in daily life, school life, and business situations	[Question] Fill in the blanks [Answers] Choose a correct answer from 4 written choices	16	Within 60 secs per question	250
Section 2	To assess the knowledge or usage of set phrases which are frequently used in daily life, school life, and business situations	[Question] Fill in the blanks [Answers] Choose a correct answer from 4 written choices	16	Within 90 secs per question	250
Section 3	To assess listening ability to understand a general idea of daily conversations, news, and announcements in an airplane	[Question] Listen to a conversation [Answers] Choose a correct answer from 4 written choices	17	Within 60 secs per question	250
Section 4	To assess listening ability to understand specific information given in daily life, school life, and business situations	[Question] Listen to a conversation [Answer] Dictate two to three words	11	Within 120 secs per question	250
Total			60	Within about 80 mins (on average, 40~50 mins)	1,000

Table 2 : *Summary of the GTEC College Test Edition*

		Purpose and Focus of Questions	Format of Questions and Answers	Number of Questions	Time	Scores
Listening	Part A	To assess listening ability to understand the description of a situation or an object, based on given information, such as a picture	[Question] Listen to the description of a picture [Answer] Listen to 3 choices and choose a correct answer	5	Within 2 mins 30 secs (Target: 30 secs per question)	250
	Part B	To assess ability to give a correct response to a sudden question in a conversation	[Question] Listen to a question in a conversation [Answer] Listen to 3 choices and choose a correct answer	8	Within 3 mins 30 secs (Target: 26 secs per question)	
	Part C	To assess ability to grasp and comprehend information in order to answer questions given beforehand	[Question] Listen to a conversation or an announcement [Answer] Choose a correct answer from 3 written choices	8	Within 5 mins (Target: 37 secs per question)	
Reading	Part A	To assess ability to understand the context of a sentence and select the most grammatical or appropriate word for that context	[Question] Read a sentence and fill in the blanks [Answers] Choose a correct answer from 4 written choices	10	Within 7 mins (Target: 42 secs per question)	250
	Part B	To assess ability to understand the content and a main idea of a short passage (150 words)	[Question] Read a short passage [Answer] Choose a correct answer from 4 written choices	5	Within 5 mins 20 secs (Target: 64 secs per question)	
	Part C	To assess ability to understand the content and a main idea of a passage (350 words) and grasp necessary information	[Question] Read a passage [Answer] Choose a correct answer from 4 written choices	1 (including 3 related questions)	Within 4 mins 30 secs	
Total				37	Within 27 mins 50 secs	500

Table 3 : *Other Characteristics of CASEC and GTEC College Test Edition*

	CASEC	GTEC College Test Edition
Operation	Web or CD-ROM	Web or CD-ROM
Hardware Requirements	(1) OS: Windows XP/Vista/7/8 Browser: Internet Explorer 6 or later. Google Chrome (2) OS: Macintosh OS X Browser: Safari5.1 or After Google Chrome.	OS: Windows 2000 Professional (except 2000 Server)/XP/Vista/7/8 Browser: Internet Explorer 6 or later.
Score Report	(1) Available right after the test on the Web. (2) Equivalent scores of TOEIC, TOEFL, Eiken are available.	(1) Available right after the test on the Web. (2) Equivalent scores of TOEIC, Eiken are available.
Price	¥ 3,500 per person *There will be a discount for groups	¥ 2,100 per person *There will be a discount for groups

4. The Trial

The trial tests of the CASEC and GTEC College Test Edition were conducted on July 25th, 2013 at Ehime University. In total, 157 students took the CASEC and 159 took the GTEC. Out of these students, 72 were from the department of Engineering and 38 were from Law and Letters. The former group of students was asked to volunteer to take the test by a supervisor in their department, while the latter students were asked to do so by their English teacher. The rest of the students were volunteers from all other faculties, with the exception of the Medical faculty. The students underwent the two tests in a row, and a half of them took CASEC first and GTEC next, while the other half did the opposite. There were 148 students who took both tests. Therefore, only the results of 148 students were used for correlation analysis with their TOEIC Bridge test scores.

On test day, three computer rooms which hold 156 computers in total were used, and the students were asked to take the both tests anytime between 9:00 am to 5:00 pm. Three faculty members from the EEC, 3 SAs, and an office staff member from each testing company were present to operate the trial.

5. Results

5. 1 Test Results

Table 4 shows the mean of both tests: the CASEC and GTEC College Test Edition. The total average score of the CASEC was 567. 78[3], and that of the GTEC was 233. 03[4].

Table 5 shows the results of the correlation analysis among the three different tests: the TOEIC Bridge, CASEC, and GTEC for College. Since the CASEC has four different sections, which are not compatible to the other two tests, we summed up the scores of Part 1 and Part 2 of CASEC in order to compare the scores in the Reading section. In the same manner, the scores of

Table 4 : *Mean average for each section*

	CASEC (*N*=157) Score Range: 250 each	GTEC (*N*=159) Score Range: 250 each
Reading	Part 1 (Vocabulary): 144.78 Part 2 (Expression): 143.13	113.34
Listening	Part 3 (Listening): 143.42 Part 4 (Dictation): 136.44	119.68
Total	567.78 (Score range: 0-1000)	233.03 (Score range: 0-500)

Table 5 : *Correlation Matrix among the three tests* (*N* = 148)

	CASEC	GTEC
TOEIC Bridge (Total)	.76	.66
TOEIC Bridge (Listening)	.42 (TBL × Part 3 + Part 4)	.58
TOEIC Bridge (Reading)	.51 (TBR × Part 1+ Part 2)	.57

Notes. TBL: TOEIC Bridge Listening section, TBR: TOEIC Bridge Reading section

Part 3 and Part 4 were summed up to compare the scores in the Listening section.

As Table 5 demonstrates, the correlation between the average of the total scores of the TOEIC Bridge and the CASEC are stronger ($r = .76$) than the correlation between those of the TOEIC Bridge and the GTEC ($r = .66$). On the other hand, in each sub section (e.g., the Listening and Reading section), the GTEC has stronger correlations with the TOEIC Bridge.

5. 2 Questionnaire

Table 6 and 7 illustrate the results of a questionnaire distributed to students regarding the CASEC and GTEC College Test Edition. The questionnaire was given in Japanese. As can be seen from these tables, in the left row of the tables, the questionnaire had nine statements and the students were told to choose their opinion from five items, "strongly disagree," "disagree," "neither," "agree," or "strongly agree." In total, 75 students answered the questionnaire: 38 answered it on the 26th, the day after the trial tests, and 36 on the 29th, four days after the tests, and one student who took the tests on the 30th answered it immediately following the tests. The numbers in the tables are the percentage of people who chose the corresponding answers. Graphs 1 and 2 demonstrate the same results using a line graph.

The results demonstrate two important elements. First, the students had similar opinions towards both tests. As can be seen in the tables and graphs, the results of the CASEC and GTEC College Test Edition were quite similar. From Statements 1 to 7, the majority of the students marked either "agree" or "strongly agree", which means they were in favor of the e-tests, although the percentages of CASEC were slightly higher than those of the GTEC. With regard to Statements 8 and 9, the majority of the students marked either "disagree" or "strongly disagree" in both tests, which means they were against e-testing. The percentages of the two tests were very similar; therefore, it is reasonable to conclude that there was almost no difference between the two types of e-tests from the students' points of view.

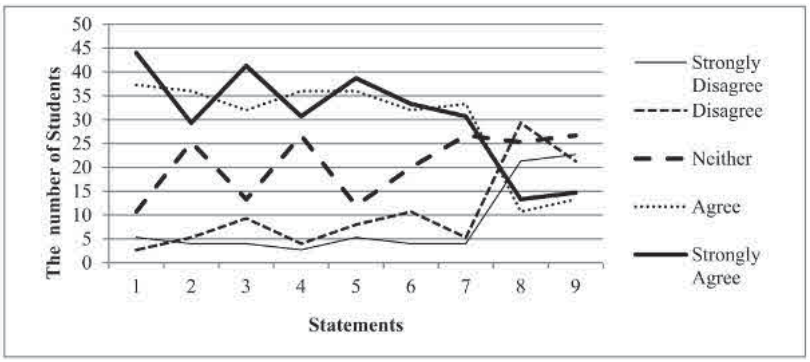
Second, the results showed that the students seemed to feel more comfortable in taking PPT than e-tests. Approximately 50% of the students who took the CASEC and 53% who took the GTEC felt that they

Table 6: *Results of Questionnaire for the CASEC*

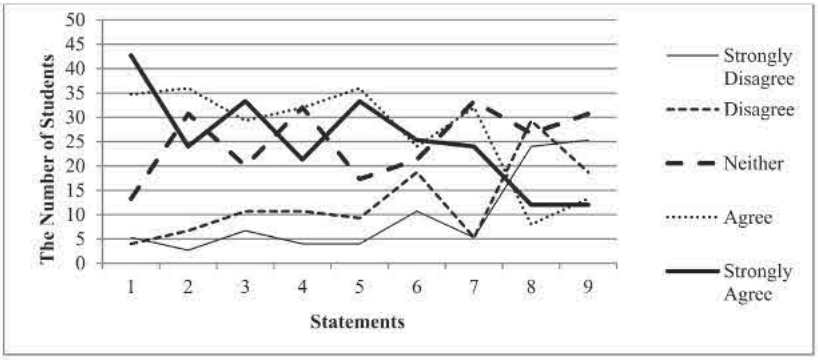
Statements	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
1. The procedure is easy to understand.	5.3%	2.7%	10.7%	37.3%	44.0%
2. The level of listening is appropriate.	4.0%	5.3%	25.3%	36.0%	29.3%
3. The quality of sound is good.	4.0%	9.3%	13.3%	32.0%	41.3%
4. The level of reading is appropriate.	2.7%	4.0%	26.7%	36.0%	30.7%
5. The size of the letters is appropriate.	5.3%	8.0%	12.0%	36.0%	38.7%
6. The time is appropriate.	4.0%	10.7%	20.0%	32.0%	33.3%
7. The score report is appropriate.	4.0%	5.3%	26.7%	33.3%	30.7%
8. I could concentrate on this test more than paper and pencil tests.	21.3%	29.3%	25.3%	10.7%	13.3%
9. I prefer this test to paper and pencil tests.	22.7%	21.3%	28.0%	13.3%	14.7%

Table 7: *Results of Questionnaire for the GTEC*

Statements	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
1. The procedure is easy to understand.	5.3%	4.0%	13.3%	34.7%	42.7%
2. The level of listening is appropriate.	2.7%	6.7%	30.7%	36%	24.0%
3. The quality of sound is good.	6.7%	10.7%	20.0%	29.3%	33.3%
4. The level of reading is appropriate.	4.0%	10.7%	32.0%	32.0%	21.3%
5. The size of the letters is appropriate.	4.0%	9.3%	17.3%	36.0%	33.3%
6. The time is appropriate.	10.7%	18.7%	21.3%	24.0%	25.3%
7. The score report is appropriate.	5.3%	5.3%	33.3%	32.0%	24.0%
8. I could concentrate on this test more than paper and pencil tests.	24.0%	29.3%	26.7%	8.0%	12.0%
9. I prefer this test to paper and pencil tests.	25.3%	18.7%	30.7%	13.3%	12.0%



Graph 1 : *Results of questionnaire for the CASEC*



Graph 2 : *Results of questionnaire for the GTEC College Test Edition*

could concentrate on PPT more than e-tests. Moreover, 44% of the students answered that they actually preferred PPT to e-tests, while about 25% of those who took the GTEC and 28% of those who took the CASEC claimed they preferred e-testing. In spite of the fact that the students were favorable to the contents of e-tests, as the answers of Statements 1 to 7 showed, it turned out that they liked PPT more than e-tests.

6. Discussion

In this section, first the results of the trial tests, and then those of the questionnaire are discussed. Third, the possibility of operating e-tests is considered from an administrative point of view.

6. 1 Test Results

The results indicate that there was a strong correlation between the average of the total scores of the TOEIC Bridge and the CASEC ($r = .76$). With regard to the correlation between the TOEIC Bridge and the GTEC ($r = .66$), there was a medium correlation. The reason why the CASEC had a stronger correlation with the TOEIC Bridge than the GTEC might be attributed to whether each test has a long reading passage in the reading section or not. While the GTEC has a passage of about 350 words, the TOEIC Bridge and CASEC do not include any long passages. The existence of a long reading passage of GTEC might affect the results seen in this research.

6. 2 Questionnaire

The results of the questionnaire indicate that the participants felt more comfortable in taking PPT than e-tests, although they showed positive attitudes towards e-tests. The reason for this might be that they are not accustomed to taking e-tests. In general, almost all the tests, including the National Center Test for university admissions which Japanese students must take before entering university, are PPT. Thus, it is no wonder that they are accustomed to paper-based tests rather than e-tests, which probably resulted in their preference for PPT.

One of the solutions for this problem is to encourage students to work on e-learning. Fortunately, the EEC has been using the e-learning program ALC Net Academy as a part of our curriculum. The students are required to complete five short sample tests of TOEIC-like questions in this program each semester as class assignments. The present purpose of using the program is to encourage students to become familiar with the content of the TOEIC test and practice taking

it, because we use the TOEIC Bridge as a proficiency test. However, now we can also use the program to have students get used to e-tests themselves.

Since we have already implemented an e-learning system in which students can practice taking TOEIC-like tests online, we should consider using it more effectively to promote their adaptability to e-tests. From an educational point of view, it is also important to facilitate it for their future because common aptitude tests for work such as SPI, CAB, and GAB[5] often require students to take them online. In an era of the internet and computer technology, more tests will be computerized; hence, it is imperative to motivate students to become accustomed to e-tests.

6. 3 Possibility for Operating E-tests

Finally, the possibilities for implementing and operating e-tests will be discussed. Table 8 shows the estimated number of students who will take an e-test each semester, which is made based on the enrollment of 2013, and the number of the students who repeated one of the required English courses in the EEC in the 1st semester, 2013 and 2nd semester, 2012.

Table 9 shows the number of the computers in five rooms in the Center for Information Technology building at Ehime University (CITE), which we might be able to use for e-testing purposes. They are generally available to faculties for classes if they are reserved in advance. Although there are more computer rooms in several other buildings on campus, using only the computers in one building has a strong advantage because of the administrative efficiency of operating the test. CITE has the most computers on campus.

Table 8: Estimated numbers of first year students taking e-tests

Dept. of Law and Letters		540
Dept. of Education		236
Dept. of Science		231
Dept. of Medicine		169
Dept. of Engineering		532
Dept. of Agriculture		179
Super Science Course		10
Repeaters	1st semester	167
	2nd semester	231
Total	1st semester	2,064
	2nd semester	2,128

Table 9: Estimated numbers of computers available for e-tests

Media Enshu Room, 2nd floor, CITE	57
The 2nd Enshu Room, 3rd floor, CITE	64
The 3rd Enshu Room, 3rd floor, CITE	42
The 4th Enshu Room, 4th floor, CITE	57
The 6th Enshu Room, 4th floor, CITE	57
Total	277

Regarding the total time required to take one of the e-tests outlined in this paper, about 45 minutes to 1 hour and 15 minutes should be necessary for one test round, depending on which e-test we use and how we conduct the procedures related to the administration of the test. It seems feasible to have seven to ten test rounds a day.

Considering the number of test-takers, at most 2, 128 students will take the test, with 277 computers being available. If we carry out seven to at most ten rounds of testing a day, all students will be able to sit for the test within one or two days.

This estimation shows the possibility of operating an e-test, but some administrative problems still remain. We are not sure yet whether it is possible to arrange the academic schedule to conduct the test on two days of the academic year. Even if we could, whether we can actually use these computer rooms simultaneously all day during the required periods is uncertain. In addition, we need to figure out how to certify each student. In the e-testing system, students take the test at different times, meaning it is crucial to ensure the accuracy of student identifications before each round of testing.

7. Conclusion

To summarize, the benefits of the use of e-testing in place of PPT are: (1) it can be operated with much less manpower; (2) it allows flexibility in creating the test schedule; and (3) this flexibility will make it possible to evaluate students' English proficiency more accurately because the test can be conducted towards the end of the semester. They will help solve problems that the EEC has faced in recent years.

From the results of the trial of the CASEC and the GTEC College Test Edition, and considering the simulation of carrying out these classes at Ehime University, it can be surmised that both tests are reasonable and feasible candidates as a proficiency test for the EEC. On the other hand, the results of the

questionnaire showed that the students prefer PPT to e-testing. As discussed, students in today's increasingly globalized world should be encouraged to become accustomed to taking e-tests because they will likely see more of these types of computerized tests as they move through both their academic and working careers. They will benefit from having good adaptability to e-testing in the future. Further research is necessary to determine how we can effectively promote students' adaptability to taking computer-based tests.

References

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.

Benesse Corporation. What is GTEC College Test Edition? Retrieved from http://www.benesse.co.jp/gtec/cte_college/index.html

Educational Testing Service (ETS). The format of TOEIC Bridge. Retrieved from http://www.toeic.or.jp/bridge/about/tests/

The Japan Institute for Educational Measurement (JIEM), Inc. The summary of CASEC. Retrieved from http://casec.evidus.com/test-summary/index.html

1) GTEC is a test which assesses the four skills of English: listening, reading, writing, an speaking. However, GTEC College Test Edition which is composed of two skills, listening and reading, was chosen for this trial considering the budget of the university.

2) IRT is a measurement theory to estimate the ability levels of test takers and the characteristics of test items such as difficulty and discrimination. IRT estimates are not based on specific samples; hence, they are stable across different groups of individuals and across different administrations. As a result, tests can be tailored to individual test takers' levels of ability and designed in the way they measure these abilities (Bachman, 1990).

3) According to the "CASEC Can-Do", the test takers can exchange ideas on general topics (e.g., daily conversation, not in the business situations and contexts) in English.

4) The "GTEC Can-do" indicates that the test takers can understand about 50 percent of the contents that the native speakers say in the general situation (e.g., sightseeing) and the test takers also can read and write in English only on their personal E-mails.

5) The SPI (Synthetic Personality Inventory) is produced by Recruit Career Co., Ltd., and the CAB and GAB are provided by SHL Japan Ltd. The SPI is the most widely used aptitude test in Japan, and GAB is the second most commonly used test. While these are general aptitude tests, CAB is for those who want to do computer-related work.