

報告

Exploring an Adequate Placement Test with Face Validity and Learners' Sense of Reality:

A Consideration of the Vulnerabilities in Existing English Assessment Models
and an Attempt at a Solution

YAMANAKA Tsukasa / KONDO Yukie

Abstract

In response to the problem that most existing English placement tests designed to measure learners' proficiency do not always adequately assess English communicative competence and its growth, this paper reports our development, improvement and implementation of the English placement exam New e-TAC, which was uniquely developed at the Colleges of Life Sciences and Pharmaceutical Sciences. New e-TAC is a thoroughly revamped version of the English Test in Academic Context (e-TAC), which was developed and implemented beginning in FY 2009 as a placement test for freshmen who took our original shared curriculum called the "Project-based English Program" at the Colleges of Life Sciences and Pharmaceutical Sciences. We created New e-TAC as a new evaluation model and began using it in FY 2015. This report describes the test concept, question construction, and results in order to examine New e-TAC in detail, especially in relation to the problems of existing evaluation models, and to discuss future challenges.

Keywords

English Education at the University Level, Evaluation Theory, New e-TAC, Placement Test, Project-based English Program

1. Introduction

In the context of rapid globalization taking place within recent years, Japanese universities face demands for aggressive reforms that achieve results. For Japanese universities to recover from the weak performance identified in the "QS University Rankings"¹⁾, various political incentives have been offered to boost university reforms, representing more direct interventions than those that have been made in the past²⁾. The importance of having English competence in a global society is undeniable, as not only English native speakers but also non-natives communicate with one another in English.

Regardless of one's field, having English expertise is highly desirable and even considered a necessary skill. For this reason, the globalization of universities should be accompanied by functional English educational reform; therefore, universities in Japan are engaging in fierce competition to create the ideal reform model³⁾.

However, regarding English education, "evaluation" is consistently problematic. While concrete and substantial outcomes are demanded in the actual practice of English education, evaluation studies on effectively assessing individual linguistic competence and its growth remain insufficient, including the development of references⁴⁾. Furthermore, along with the methods of assessment, it is vital to understand how the people making the assessment will guarantee the learners' competence with validity. If the test results for English competency are called "output" and other elements that cannot be scored are called "outcomes," desirably, both factors should be evaluated sufficiently. Moreover, those who are assessed supposedly have face validity to each given evaluation, which means that they must have a sense of real consent and of its adequacy. Ideally, the evaluation will also serve to provide convenient information for future studies.

In response to the problem that most existing English placement tests designed to measure learners' proficiency do not always adequately assess their English communicative competence and its growth, this paper reports our development, improvement, and implementation of the English placement exam called New e-TAC, which was uniquely developed at the Colleges of Life Sciences and Pharmaceutical Sciences. New e-TAC is a thoroughly revamped version of the English Test in Academic Context (e-TAC), which was made and implemented beginning in FY 2009 as a placement test for freshmen who took our original shared curriculum called the "Project-based English Program" at the Colleges of Life Sciences and Pharmaceutical Sciences. We created New e-TAC as a new evaluation model and began using it in FY 2015. This report describes the test concept, question construction, and results in order to examine New e-TAC in detail, particularly in relation to the problems of existing evaluation models, and to discuss future challenges.

2. Case Study: English Competency Assessment and the Gap between Theory and Reality

There are many types of assessment models for evaluating English competencies in Japan and abroad, the most famous of which are TOEFL, TOEIC, and IELTS, and new English evaluation models are being developed actively. By improving many aspects every year, the existing models allow us to draw closer to comprehending English performance and each learner's competence. However, there remains a huge gap between what human evaluators might judge to be a good response in actual communication and test results. An incident that occurred in FY 2014 made us acutely aware of this gap.

Table 1 shows a pre-post comparison of TOEIC-IP scores⁵⁾ for all students who participated in the English for Science & Technology Program, which has been uniquely developed as an overseas course by the Colleges of Life Sciences, Pharmaceutical Sciences, and Sports & Health Science.

During their stay of about one month in the U.S., they had many opportunities to use English through various extracurricular activities as well as intensive English classes. Additionally, because they chose homestay as their housing option, there was no shortage of chances to use English. Overall, it was one of the places where students could enjoy English practice in the context of personal human connections and could simultaneously enhance their ability.

Table 1 A Pre-Post Comparison of TOEIC-IP Scores (The example of the English for Science & Technology Program at University of California, Davis, which has been uniquely developed as an overseas course by the Colleges of Life Sciences, Pharmaceutical Sciences, and Sports & Health Science)

	N	Mean (990)	SD	<i>t</i>
Highest TOEIC-IP scores before the program	36	613.75	122.10	6.08*
TOEIC-IP scores, from the last day of the program at UC Davis	36	563.47	128.36	

* $p < .00$

The table shows that the English proficiency of students in the four-week overseas program did not improve as a whole if only the test results of pre/post TOEIC-IP scores are considered. A repeated measures t-test revealed that their post average scores were significantly lower in comparison with their highest scores before departure. The decreasing scores by and large came as a shock to the people who applied significant efforts into making the program a success, such as school faculty, secretariats, and English teachers, including us. The result was completely beyond the scope of assumption. The Colleges of Life Sciences, Pharmaceutical Sciences, and Sports & Health Science have confidently recommended the overseas program to students, and in fact, the students' responses to a post-questionnaire verified that the contents of the program are dense, fulfilling, and meaningful. Ultimately, the students' lower scores revealed our simplistic assumption that the intensive English environment of the overseas program would increase students' TOEIC scores.

Then, did students' English ability actually decrease as a result of their participation in the overseas program? The answer is "no." When I supervised the students in the U.S., for most of the program period, I observed that their English competency was improved markedly. With rich experiences such as the intensive English classes where the input was fairly authentic, and many extracurricular activities with local students who were native speakers of English, the growth of their communicative competence in English was nothing short of eye opening. Moreover, they themselves seemed to realize they were gradually becoming able to communicate in English. As proof of their feelings about the program, I present the following questionnaire responses written by the participants of the program.

Selected responses to the questions "By participating in the English for Science & Technology Program, including pre-lectures, did you experience any change?" and "Would you write down your impression, opinion, or suggestion for the English for Science & Technology Program as a

whole?”

- I think I improved my ability in English performance and motivation. But I found it hard to study specialized fields in English.
- I could do a lot of what I wanted owing to much free time, such as lunch and dinner with local students, trips, etc. In that regard, it was good that all the classes were in the morning. I built a good relationship with my host family, too. Chatting with them was fun and led me to improve my language ability and multi-cultural understanding.
- I think I gained a positive attitude toward English. Though I think I always had a positive attitude in exchanges in Japanese, I increased my positive feeling in using English. I used to be too embarrassed to address native speakers or tell them my opinion because of my limited English ability; therefore, I often pretended to listen without understanding. However, through staying in the U.S. this time, I keenly sensed that native speakers try very hard to grasp my poor English. They wait till I finish thinking of how to say something, and they speak many times whenever I ask. So, my “wall” disappeared. Now, I can convey my opinion with confidence and tell people “I don’t understand” when I need to.
- I came to be able to ask people for help when I come across things that I don’t understand. I could act out what I wanted to realize.
- I shed my discomfort about communicating in English. I gained abilities to act by myself. I improved my English competency.
- I raised my interest in English. I became interested in America. I got rid of my hesitation in communicating or writing in English.
- I definitely changed. Through the Hot Topic and guest lectures, my interest spread beyond my specialties to other fields. Also, through the interviews in the Intercultural Research Project (IRP), I felt less hesitation in taking initiative in English and less anxious about expressing my opinions in English.
- I gained strength. I felt curious about everything. My motivation for classes in general and English study has been increased.
- I no longer hesitate to address foreigners.
- I increased my motivation toward not only English study but also study in itself. Moreover, I felt courage in studying abroad.
- My eagerness to learn English has increased. I think I want to go abroad again. I feel confident I will keep studying English after this.

These comments show that, in spite of recognizing their limited English proficiency, the students pushed the boundaries of their capabilities by attempting to communicate in English. From the view of English learning, it is well presumed that the students improved their English ability by engaging in a significant degree of “listening” and “speaking” using English in this process, although we have to admit it is fairly “subjective”. We couldn’t find any “objective” proof that can guarantee their language

improvement at our end. We are at a loss for words about what led to their lower TOEIC-IP scores.

To be clear, this report does not intend to object formally to the English measurement methodology of TOEIC-IP. What this paper states, using this case as a typical example, is that the results from reliable, widespread, and time-proven evaluation models of English competence often diverge from the feelings and conditions of each learner and institution, and this aspect may be sharable among actors in English education. Additionally, it may leave people with lingering suspicions of the reliability of assessment, which is a serious problem that could affect learners' motivation and performance.

Ideally, the evaluation of English ability should visualize each learner's explicit capability and his or her progress from a comprehensive perspective. The learner ought to be able to reasonably accept the assessment results without needing to be persuaded. Furthermore, from the viewpoint of expertise, it should equip a high level of diagnostic function that can explain which language skills are developed to what extent in a way that is easy to understand. We must pursue such models because they might be strongly linked with the cultivation of learners' self-affirmation, correct analysis of their present state, and their target recognition for the next step. Of course, currently, there is no such model with these evaluation capacities, and the feasibility of developing one that embodies all these traits may be quite doubtful. However, for professionals dealing with English evaluation in education, striving for the realization of this ideal or, at least, the continuous refinement of practical evaluation models is imperative. New e-TAC, which is described, is one such attempt. While admitting that many improvements remain to be made, we regard it as an effective practice that is helpful in advancing the model closer to the ideal.

3. Problems in English Placement

By referring to the case of students' test scores following an overseas program, the previous chapter explained that the existing widespread evaluation models of English proficiency do not always reflect learners' actual abilities. Fortunately, in that case, the students had many qualitative opportunities not reflected in the TOEIC-IP score, such as their final presentation after returning and performance in daily classes after the program. Consequently, not only English teachers but also other people involved in English assessments, including the professors and staff of the three colleges, need to consider TOEIC-IP scores not as absolute indicators of students' abilities but as one of several elements related to the program. However, if the same type of problem occurs again with an English placement exam, we would have to say that the situation is serious. This report has materialized from an awareness of this problem.

The term "placement" used here means the class assignments made right after freshmen enter university, and all modules of Skill Workshops in the Project-based English Program apply a placement system that assigns students according to their English proficiency. The reason the classes are determined by ability grouping is that, when learners who have similar competence in English

communication are placed in the same section, it is not only highly expected that they will gain a stronger motivation to learn and achieve a better performance through friendly competition with one another but also that it will be easier for them to share the challenges they face in language communication. Moreover, the Skill Workshop, which is required to enhance learners' English communicative competence for a limited period, also functions to enhance the efficiency of class management by enabling instructors to easily create a strategy for the administration of each class by combining similar challenges into a particular section.

When all these perspectives are taken into account, it becomes a major issue that the placement system as a whole does not work well. If class placement is based on a heavily biased test and the test results do not reflect the actual English communicative competence of each learner, the learners will have to attend classes that may be very easy or difficult for them, making them feel alienated, and instructors will find it difficult to prepare appropriate lesson plans. Therefore, we introduced New e-TAC and CASEC as a new attempt at determining students' appropriate placement in English classes. To cope with the problems described above, increasing the face validity for learners was highly necessary by provisionally combining multiple evaluation references. In other words, we explored the practice of English placement more alongside actual English communicative competence.

4. What New e-TAC is aimed at

The e-TAC is a test to assess students' English proficiency level in an academic context. In FY 2015 attempt, New e-TAC had the following three objectives:

1. Efficient operation

To operate the test efficiently for about 430 students and score it efficiently in a fairly limited period of time, which is up to 2-3 days between the orientation day and the first day of class.

2. Effective placement

To form classes for Skill Workshop, which is one module of the Project-based English program along with Project, according to students' English proficiency level. This test will specifically place the highest and lowest level students appropriately to provide them with the feasible attention.

3. Assessing students' development

To set a yardstick in assessing students' development levels in various phases for the future.

The major purpose of the former e-TAC was to supplement the standardized test, TOEIC Bridge, enabling us to place upper and lower level students in an appropriate manner by solving the ceiling effect and floor effect that TOEIC Bridge had. The former e-TAC was originally used only as a sub-test, but through careful revision of the test and its implementation method, the former e-TAC was fully established as a complementary test in FY 2011. However, the analysis by Kondo and Yamanaka (2014) discovered two drawbacks: firstly, the former e-TAC did not contain a high enough number of question items to achieve sufficient reliability, and also, the questions were relatively easy. New e-TAC was developed to solve these issues, while achieving the above mentioned three objectives.

5. Changes to New e-TAC

5.1 New e-TAC to serve as a complementary test along with a standardized test, CASEC

Changes were made to all the major qualities of the test, such as the contents of the test material, the number of questions and the duration of the test, whilst keeping within the original framework of the established system. Before reporting the changes in detail, we should mention that the standardized test used along with New e-TAC was changed from TOEIC Bridge to CASEC. The test formats of TOEIC Bridge and CASEC are indicated in Table 2. The current developer of CASEC, the Japan Institute for Educational Measurement (JIEM)⁶⁾ describes CASEC as an English communication ability assessment test using a computerized adaptive testing (CAT) system (JIEM, 2015). As of 2015, CASEC has been adopted by 1178 institutions, including companies, universities and schools, and the number of test-takers has amounted to more than 1.5 million. The reliability of CASEC is reported to be high, with Spearman's rank correlation coefficient ranging from 0.92 to 0.96 (van der Linden 2010). That is to say, CASEC is now a widely acknowledged examination.

Table 2 The Test Format of TOEIC Bridge and CASEC

	Section	Question type	Number of items	Possible points	Duration
TOEIC Bridge	1. Listening	Multiple choice	50	90	25 min.
	2. Reading	Multiple choice	50	90	35 min.
CASEC	1. Vocabulary	Multiple choice	16	250	40-50 min.
	2. Expression	Multiple choice	16	250	
	3. Listening (Comprehension)	Multiple choice	17	250	
	4. Listening (Dictation)	Dictation/Typing	11	250	

The major reason why the standardized test was changed was to avoid the floor effect and ceiling effect that existed with TOEIC Bridge. In the previous chapter, it was mentioned that the former e-TAC was developed to solve these issues; however, changing the standardized test itself enabled us to use the New e-TAC more effectively for the placement. TOEIC Bridge is developed for introductory to intermediate level English learners to serve as a "Bridge" to TOEIC, containing much easier questions but in the similar format to TOEIC. On the other hand, CASEC is intended for learners with greater range of English proficiency. The difference between the target ranges of English proficiency levels of these tests can be seen in the reference score comparison that both test developers have published. (See Table 3) ETS, the developer of TOEIC and TOEIC Bridge, says that TOEIC does not have questions which are lower than the TOEIC Bridge 90 point level. They also recommend that learners should take TOEIC instead of TOEIC Bridge by the time they are able to score 160 or higher in TOEIC Bridge. In contrast, the CASEC's score range is much wider, corresponding to TOEIC's

range 10 to 990 points. As a matter of fact, we had 44 students out of 407 who scored 160 or higher in FY 2014's TOEIC Bridge, which clearly showed the limitation of TOEIC Bridge as a placement test. It is true that TOEIC or CASEC's highest and lowest scores do not mean the highest and lowest English proficiency level, but at least CASEC addresses the limitation that TOEIC Bridge has by providing the wider scale to assess learners English proficiencies more effectively for the placement.

Table 3 Reference Score Comparison of TOEIC Bridge/TOEIC and CASEC/TOEIC

TOEIC	TOEIC Bridge *1	CASEC *2
990		1000
570	160	
230	90	
10		0

*1 Education Testing Service (2007)

*2 JIEM (2015)

Another advantage of using CASEC as a placement test was that, with the above mentioned wider range of questions, it still required a shorter test time of 40-50 minutes compared with major standardized tests such as TOEIC, 120 minutes, TOEIC Bridge, 60 minutes, TOEFL ITP 120 minutes. As the CAT system changed question items depending on the test-taker's previous answer, it required less time compared to the paper-based test where every test-taker is supposed to answer all the questions regardless of their proficiency level. In addition to these advantages, since the answers were collected using a computer, the scores were available shortly after the test.

5.2 The Structure and Contents of New e-TAC

As a result of the reduced test time, we were able to allocate more time to New e-TAC. The former e-TAC was 45 minutes with 14 items, New e-TAC being 70 minutes with 40 items. The format of New e-TAC is indicated in Table 4.

Table 4 The Test Format of New e-TAC

Section	Question type	Number of items (40)	Possible Points (63)	Duration (70 min.)
1. Grammar & Structure	Ordering	20	40	60 min.
2. Reading	Multiple choice	15	15	
3. Listening	Multiple choice	2	2	10 min.
	Dictation	2	4	
	Free description	1	2	

Section 1. Grammar & Structure

The Grammar and Structure section, consisting of 20 multiple choice questions, was added to New e-TAC, as CASEC did not contain this section. Assessing students' basic grammatical ability was meaningful for our program, as the lower four classes were designed to be special-attention classes, whose main objective was to build students' basic grammar and structuring abilities necessary for English communication. This section was designed to gain a higher percentage of correct answers, containing fairly easy questions of a junior high school level. The results of this section were used to select students for special-attention lower classes as well as to identify students at higher levels who can speak and listen but have difficulty in constructing sentences when it comes to writing. These students otherwise would have been assigned to upper classes based on the results of CASEC only, which might have led to less appropriate attention of their language weaknesses.

Section 2. Reading

For the Reading section, an article about "spider webs", titled "Why spider webs glisten with dew: Two driving forces acting on wet spider silk help it to capture water" from Nature News was used. Reading texts were 503 words long, 80% of which consisted of vocabulary from junior high school to high school levels⁷⁾. As for some difficult words, Japanese annotations were given. There were fifteen multiple choice questions; eight were comprehension questions, seven were fill-in-the-blank questions. This section was expected to gain about the same percentage of correct answers as the former e-TAC, which was approximately 64%.

Section 3. Listening

For the lecture in the Listening section, a TED Talk about "spider webs", titled "The magnificence of spider silk", was used. This section included two questions for comprehension as well as two dictation questions and one free description question on the contents of the lecture. Students answered the questions, while watching the TED Talk video. After they answered each question, the video was repeated again. This section was conducted as follows. For the first question of this section, students watched the video for 40 seconds and they were given 30 seconds to answer the question. After that, the same scene of the video was repeated, and stopped for 30 seconds again. The second question in this section was dictation. Students watched the next part of the TED Talk for 10 seconds and wrote out what they had heard for 10 seconds. The video was repeated, for this question as well. Students proceeded to the third and fourth questions following the same sequence. The final question was free description. They were asked what spiders use their silk for. They were required to give one purpose. In the video, the speaker mentioned three purposes and also the keywords for them were written in the slideshow in the video (See Figure 1). This section was expected to be more challenging than the former e-TAC; therefore to gain a lower percentage of correct answers.



Figure 1 An Example of the slideshow from the TED Talk used in the Listening Section

The reason why we used Nature News and a TED Talk was that their contents were academic but at the same time they were directed to a general audience rather than specialists or researchers. Therefore, we judged that these materials were appropriate for freshman students compared to research papers or lectures in a specific field. In addition, Nature News has been introduced into the textbook that the senior students in our departments use. That may give us the opportunity to compare the comprehension levels of the article between freshmen and sophomores later on.

The changes made from the former e-TAC to New e-TAC are shown in Table 5.

Table 5 Comparison of the Former e-TAC and New e-TAC

		Former e-TAC	New e-TAC
Grammar & Structure	Number of items	N/A	20
	Question Type		Multiple choice
Reading	Number of items	8	15
	Question Type	Multiple choice	Multiple choice
	Material	BC Science 10	Nature News
Sections	Number of items	5	5
	Question Type	Multiple choice	Multiple choice Dictation Free description
Listening	Material	Lecture based on the reading article, originally recorded	TED Talk
	Number of items	1	N/A
Writing	Question Type	Free description	
	Duration	45 minutes	70 minutes

5.3 Operating Process of New e-TAC

The computer-scored answer sheets were used for question items except for dictation and free description in the Listening section. All the necessary directions for the test as well as the presentation video in the Listening section were formatted in a movie file. In other words, the movie file for New

e-TAC contained directions for the whole test and each section, how to answer questions, and when to start and finish the test. Therefore, all that the teachers, or supervisors, had to do was simply play the movie. After the test, teachers who were in charge of supervising the test also scored it on the same day. This year, three teachers had the answer sheet read by the computer, and they then input the data and scored the dictation and free description parts, while waiting for the CASEC results.

5.4 Placement Procedure

We first totaled up scores for both CASEC and New e-TAC and divided the students into 30 classes based on their test scores. Next, for the students in the four lower classes, we placed more importance on the Grammar and Structure section in New e-TAC, as it assessed their grammatical abilities at a basic, junior high school level. We also took into account that these students might not have understood the article and the lecture in an academic context in the Reading and the Listening section and that the scores for these sections might have been the result of guesswork. Based on this assumption, we placed priority on the Grammar and Structure section rather than the Reading and Listening sections and reordered the students by carefully reviewing the scores for the Grammar and Structure section. As for the higher classes, the highest class was regarded as being the elite class. To form the elite class, we eliminated those who had significantly low scores in any particular section in either CASEC or New e-TAC. Consequently, the number of students in this elite class was only 6. That was a remarkably small class, as the average number of students in other classes was about 14. We will keep track of the outcome of this class and report on it after FY 2015 ends.

6. Results of New e-TAC

6.1 Basic Statistics

The basic statistics of the results of New e-TAC are shown in Table 6. The maximum possible points achievable in this test were 63. The mean score was 43.93; thus the mean percentage of correct answers was 72.91%. As mentioned earlier, each section of New e-TAC was designed to have different levels of difficulty. The mean percentage of the Grammar and Structure section was 84.60%. Considering that this section includes questions of a junior high school level, this high percentage was as expected. The mean percentage for the Reading section, which was expected to have the same level of difficulty as the former e-TAC, was 60.20%. This also met our expectations. The mean percentage for the Listening section was 37.00%. This number was slightly lower than our expectations, but gave us an insight into how Nature News and TED talks on a scientific subject were challenging to most of the freshmen.

Table 6 The Basic Statistics of The Results of New e-TAC (n=420)

Section	Mean (63)	Mean (100%)	Max.	Min.	SD
1. Grammar & Structure	33.84	84.60%	40	16	5.07
2. Reading	9.03	60.20%	15	2	2.49
3. Listening	3.23	37.00%	7	0	1.36
Total	43.93	72.91%	60	22.5	7.45

6.1 Reliability

Reliability of New e-TAC was measured using Cronbach's alpha coefficient, a widely used reliability measurement. The value obtained from New e-TAC was 0.78. Taking into account that a test with an alpha value above 0.7 is generally considered to be reliable (e.g. Morgan, Leech, Gloeckner, & Barrett, 2013), and also that the value of New e-TAC was higher than that of the former e-TAC (0.57 to 0.65), the reliability of New e-TAC was improved as well as acceptable. This is assumed to be the result of an increased number of test items, from 13 to 40, as it is widely believed that having more items leads to an improved reliability value (e.g. Hatch and Lazaraton, 1991; Henning 1987).

6.2 Correlation between New e-TAC and CASEC

Spearman's correlation coefficient was used to evaluate the relationship between New e-TAC and CASEC. A strong correlation was found between the overall scores of New e-TAC and CASEC ($r = .75$). What should be paid attention to is that each section shows varying correlation values. As Table 7 indicates, sections except for the Listening sections in both New e-TAC and CASEC have moderate correlations. On the other hand, the Listening section (Comprehension) of CASEC has only a weak correlation with every part of New e-TAC. The Reading section of New e-TAC also has a weak correlation with the Listening sections (Comprehension and Dictation) of CASEC. This indicates that the Listening section (Comprehension) of CASEC and the Reading section of New e-TAC have different criteria of assessment compared with the other sections. CASEC aims to assess English communication ability, while New e-TAC was designed to assess students' English competence in an academic context and also to serve as a complementary placement test along with the standardized test, CASEC. Of course, communication occurs in any context, but CASEC measures English communication proficiency level in daily life, at school or in business. In other words, what CASEC aims to assess is daily life-based. On the other hand, New e-TAC, excluding the Grammar and Structure section, uses a scientific article and lecture. In the Listening section, students not only just chose answers based on what they had heard, but they also had to understand the lecture and summarize a part of it to answer the free description question. The results obtained here suggest that English communication competence required in daily life and English competence required in an academic context are different.

Table 7 Spearman's Correlation Coefficient between New e-TAC and CASEC (n = 420)

		CASEC				
Section		1. Vocabulary	2. Expression	3. Listening (Comprehension)	4. Listening (Dictation)	Total
New e-TAC	1. Grammar & Structure	0.58	0.57	0.39	0.44	0.63
	2. Reading	0.56	0.51	0.38	0.38	0.59
	3. Listening	0.47	0.49	0.40	0.48	0.58
	Total	0.68	0.66	0.48	0.53	0.75

7. Evaluation of New e-TAC in accordance with objectives

This section evaluates New e-TAC in accordance with the three objectives mentioned in Chapter 4. As for the first objective, "efficient operation", formatting necessary directions in a movie file and using computer-scored answer sheets worked effectively, making it possible to finish conducting the test and scoring it on the same day. However, since CASEC was fully implemented on a computer, it is time for us to consider developing a computer based complimentary test. Yet, we should take into consideration that new freshmen's computer proficiency is not as high as that of a university student in higher years. Secondly, for the second objective, "effective placement", New e-TAC played a significant role in placing students into appropriate classes in an academic setting, in our case, the Project-based English Program. Although CASEC is now a widely acknowledged test and it alone can be utilized to assess the students' communication abilities, students in our program will utilize and develop abilities not just for communicative purpose in general. A weak correlation between the Listening sections of CASEC and New e-TAC suggests that some students can understand what they listen to in a daily life context, but cannot necessarily understand academic lectures or summarize what they have listened to. Also, a weak correlation between the Reading section of New e-TAC and all sections of CASEC implies that some students can understand academic articles but do not have as high communication ability in daily life. More analysis is required to determine the factors that explain these different correlations. Finally, for the third objective, "assessing students' development", we plan to carry out our research in various phases, by using the results of different tests in different phases, as this is only the first year for us to introduce New e-TAC in our colleges.

8. Discussion

University entrance examinations, a major renewal of the National Center Test for University Admissions, are scheduled to begin in FY 2020. However, for the subject of English, an external examination is supposed to be utilized, such as TOEFL; this illustrates that we struggle with the assessment of English compared with other subjects whose new tests will be developed with no

external source needed. Indeed, developing an English proficiency test might be a burdensome task. However, starting with the Test of English for Academic Purposes (TEAP)⁸⁾, which has been introduced to the English entrance examinations of some universities, and the changeover of STEP (The EIKEN Test in Practical English Proficiency) level 2 from three skills to four skills with the addition of a writing dimension⁹⁾, many developments and improvements to English evaluation methods are underway. It is not too much to say that the field of English assessment is now in “the age of laissez-faire.” If entering a period of upheaval, we make at least a plain statement that this report is not intended to participate in specific controversies.

Testing is necessarily regarded as an “indirect” measurement for some reason. In this context, we must accept its limitations; in other words, there is no definitive theory of evaluation. Additionally, as there are constraints on administrative factors such as time limitations, we are not very reluctant to pursue a “mixed model” wherein existing evaluation models are combined as a complement. We expect to continue doing our best to conduct appropriate placement wherever possible, and we will also aim at enhancing the quality of English educational programs eventually. In this regard, expectations will certainly be high in New e-TAC.

The aim of the Project-based English program is to improve professional skills, which encompass the skills in conducting presentations, discussions and debates. These skills are required when conducting research, as well as being necessary in expressing the process and the results of the research. Therefore, it would be ideal to assess these skills right after the start of their first year, in the placement test. Yet, it is obviously not realistic to include actual presentations in the placement test due to time and logistic constraints. Thus, in an attempt to make the test as authentic as possible, we used a scientific article and presentation, which are similar to the types of material that students might use in their own research in the future. Here, authentic does not only refer to the test material itself but taking New e-TAC is regarded as effective practice for their own future projects, as the process of carrying out the test is similar to the process of conducting research, which involves reading an article, listening to a lecture on a related subject, and then writing about it. Taking New e-TAC would serve as if students had already started their academic research in English, and in that way, we believe that we will make greater advances in creating a test that will optimally assess the students’ development in our program. By the time students have finished their first year English project, we plan to carry out further research to measure how effectively we have placed students using New e-TAC along with CASEC in light of their performance and growth level in the Project-based English program.

Notes

- 1) This is a ranking by Quacquarelli Symonds, a rating agency of world universities in the U.K. (<http://www.topuniversities.com/university-rankings>). THE World University Rankings (<https://www.timeshighereducation.co.uk/world-university-rankings>) by Times Higher Education, etc., are also well known. The Times Higher Education Asia University Rankings 2015 revealed that Japan has only 19 universities in the

- prestigious top 100, down from 20 universities in 2014. On the other hand, China has 21 universities for the first time, leading to the conclusion that “China has overtaken Japan as Asia’s number one nation for producing top universities.” (<https://www.timeshighereducation.co.uk/news/asia-university-rankings-2015-results-announced>)
- 2) See the Ministry of Education, Culture, Sports, Science and Technology’s “Top Global University Project” (http://www.mext.go.jp/a_menu/koutou/kaikaku/sekaitenkai/1319596.htm) and “Re-Inventing Japan’s Project” (http://www.mext.go.jp/a_menu/koutou/kaikaku/sekaitenkai/), etc. Here one can glimpse the government’s intention to select already competitive or promising universities and let them achieve positive results in a short period of time by strategically budgeting.
 - 3) Despite their contentiousness and divisiveness, various approaches and trials are included, such as obligatory study abroad for one year, which is a well-known policy of Akita International University and the School of International Liberal Studies, Waseda University (<http://web.aiu.ac.jp/abroad>), etc., or The Village: English, Enjoyment, Education administrated by Kindai University (<http://www.kindai.ac.jp/e-cube/>) for the purpose of producing space for students to utilize English outside the classroom, etc.
 - 4) In Europe, the Common European Framework of Reference for Languages (CEFR) has now been in widespread use. The framework was officially developed and opened in 2001 by the Council of Europe, for the purpose of certifying a common frame for various elements such as second language use, educational guidelines, learners’ achievement, etc., intended to enhance foreign language education in Europe. Since it is effectively utilized as an index for immigration control and conditions of employment in Europe from the point of prevailing, CEFR is regarded as a standard with social impact that exceeds the context of language education (Dunlea, 2009)「英検とCEFRの関連性について Part 1」[STEP 英語情報 11・12月号]及びDunlea, J. (2010)「英検とCEFRの関連性について Part 2」[STEP 英語情報 1・2月号]. However, there must be many controversies in the consideration of criticism toward Can-do descriptors, which distinguish the levels of CEFR. Further, there is the question of whether we can use European guidelines in the same way in Japan, despite the differences of sociolinguistic background.
 - 5) According to the TOEIC Official Website (<https://www.ets.org/toeic/succeed>), it emphasizes its standardization. It refers “They can be compared regardless of where or when the test is administered”.
 - 6) <http://www.jiem.co.jp/en/business/testing/>
 - 7) The vocabulary levels were determined by the JACET list of 8000 basic words, published by JACET Basic Word Revision Committee, 2003
 - 8) <http://www.eiken.or.jp/teap/>
 - 9) http://www.eiken.or.jp/association/info/2015/pdf/20150715_pressrelease_writing2.pdf

References

- Education Testing Service. “Statistical Relationships between the TOEIC Bridge Scores and the TOEIC Listening and Reading Total Scores”, 2007
- Hatch, E, & Lazaraton, A., The Research Manual: Design and Statistics for Applied Linguistics, New York: Newbury House, 1991.
- JTEM, CASEC Official Website, 2005
- Kondo, Y. and Yamanaka, T. “Development and Implementation of “English Test in Academic Context” (e-TAC) as a Placement Test for “Project-based English Program” in the College of Life Sciences and Pharmaceutical Sciences: Achievements and Challenges” *Ritsumeikan Higher Educational Studies*, No. 14, 2014, pp.131–146.
- Morgan, G. A., Leech, N. L., Gloeckner, G. W, & Barrett, K. C., IBM SPSS for Introductory Statistics: Use and

Interpretation (5th ed.), New York: Routledge/Taylor and Francis, 2013.
van der Linden, W. J. *Elements of Adaptive Testing*. New York, NY: Springer, 2010.

表面妥当性を備え学習者にとって「実感の持てる」プレイスメント・テストの模索
—既存の英語テストのもつ脆弱性の考察と具体的解決に向けての試み—

YAMANAKA Tsukasa (Associate Professor, College of Life Sciences, Ritsumeikan University)

KONDO Yukie (Lecturer, College of Pharmaceutical Sciences, Ritsumeikan University)

要 旨

本報告は、既存の英語習熟度を測定するプレイスメント・テストの多くが、学習者の英語コミュニケーション能力及びその「伸び」を、必ずしも適切に評価し切れていないとの問題意識に立ち、生命科学部・薬学部における独自の英語プレイスメント・テストである New e-TAC の開発と実践について述べるものである。New e-TAC は、生命科学部・薬学部における統一の独自英語カリキュラムである「プロジェクト発信型英語プログラム」における 1 回生のプレイスメント・テスト用に 2009 年度より開発・実施している e-TAC (English Test in Academic Context) を大幅に刷新し、2015 年度より運用を開始した評価モデルである。本報告では、既存の評価モデルや指標の問題点を具体的に検討すると共に、New e-TAC のコンセプト、問題構成、実施結果、考察について詳細に報告し、今後の課題についてまとめるものである。

キーワード

大学英語教育、評価論、New e-TAC、プレイスメント・テスト、プロジェクト発信型英語プログラム

