

A LONGITUDINAL STUDY OF THE ENGLISH PROFICIENCY OF NON-ENGLISH MAJORS AT A TERTIARY INSTITUTION IN JAPAN

Gregory Minehane
(minehane@meijo-u.ac.jp)
Meijo University, Japan

Abstract

General English classes for non-English majors from six faculties at a university in Japan were reorganized into the Liberal Arts English Program (LAEP). The program was based around the Cambridge English exam suite and was aligned with levels of the CEFR, namely A2, B1 and B2, incorporating standardized textbooks, syllabi and final examinations. At its peak, the program was taught by 50 teachers and catered annually to 2700 students. University entrants were given a proficiency test (CASEC) for placement into class levels and students were tested again at the end of the academic year. In this paper, annual scores for these CASEC tests (2006 to 2013) will be analyzed and discussed. The data show that in any academic year, averages point to little, if any, measureable improvement in English proficiency. An independent study, employing a four-skills Cambridge PET test of 31 students in the LAEP (Nagao et al., 2012) adds support to this finding. Detailed examinations of the data reveal trends of interest for curriculum planners. However, it appears two weekly language lessons are insufficient. The common model of compulsory university English for non-English majors needs to be reconsidered if we are aiming to bring about meaningful improvement in proficiency.

1 Introduction

1.1 The Liberal Arts English program (2006 – 2013)

In 2005, it was decided that liberal arts education, including English education, would be managed university-wide and fall under the auspice of the Liberal Arts Program (LAP) run by the University Education Development Center. Eventually, a total of six of the eight faculties at our university would join the Liberal Arts English Program (LAEP) with first-year student participation averaging over 1700 a year and second-year students totaling another 1000. From 2006 to 2010, the LAEP curriculum centered on two examinations from the Cambridge ESOL examination suite, KET and PET. The program was unified and standardized with teachers following strict guidelines of what to focus their lessons on. By late 2010 however, the current of support which had created LAP was beginning to disappear - a reorganization of management and a change in university policy concerning liberal arts saw the tightly controlled program begin to unwind. In the transitional 2011 academic year, instructors were largely free to teach students what they thought was important through self-selected texts and materials. In 2012 and 2013, faculties provided guidelines for instruction which broadly recommended focusing lessons on reading skills. In 2014, only the framework of LAEP is still reflected in the names of classes and

even these are slated to change in the coming academic year. As the university returns the administrative power of deciding what is universally important in liberal arts education to the province of individual faculties and completes a large-scale organizational restructuring, it is a time to reflect upon the successes and failures of eight years of the LAEP (2006-2013).

1.1 Research questions

All first-year students in the LAEP were given a proficiency test of English in April during the university orientation period and again towards the end of the academic year in December. The aim of the current research is to examine the improvement in English proficiency of these English non-majors over the course of one academic year. The first few years of the LAEP were taught in a unified and standardized curriculum; the latter years were far less controlled. Consequently, the results may also indicate whether this change in program style and management affected student outcome. It is hoped that what we learn from this analysis can help us understand the Japanese undergraduate non-English major and be applied to the management and implementation of English programs in the future.

1.2 Research context and English language program

After completion of a trial program in 2006 (see Tadaki & Minehane, 2007) the LAEP was designed with the aim of giving all students standardized training in communicative English. Standardized assessment tools based on the Cambridge ESOL exam suite were used as end of semester final exams and classes of students with a similar level of English ability were taught with the same textbooks using the same standardized syllabus. The program was evaluated internally through the use of student and teacher surveys, Cambridge ESOL (KET and PET) examinations, teacher coordination sessions and course development meetings. Broadly based on the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), the program aimed to improve the ability of students in all four language skills. Students in the program studied English twice weekly – once with a native English instructor (NES), and once with a Japanese or non-native English instructor (NNES). NES instructors focused their teaching on the productive skills of speaking and writing while the NNES focused on the receptive skills, listening and reading. Textbooks focusing on preparation for the Cambridge ESOL KET and PET examinations (see Capel & Sharp, 2005; Naylor & Hagger, 2004) were shared by instructors by having them teach alternating units, concentrating on their allotted language skills. English lessons were 90 minutes each and it was compulsory for students to take classes in the first and second semesters. For this type of English class at a university with non-English majors in Japan, relatively small class sizes of between 20 and 30 students were standard in our program.

Students in the LAEP were given a placement test upon enrollment at the university. From 2006 onwards, this test was the Computerized Assessment System for English Communication (CASEC) which will be discussed further below. Based on CASEC scores, students were divided into four levels of ability ranging from ‘Basic’ to ‘Intermediate’. Table 1 shows the levels of the LAEP, CASEC scores and the commensurate levels in the CEFR and Cambridge ESOL exam suite.

Table 1. Corresponding levels - LAEP, CASEC, CEFR and Cambridge ESOL Tests

<u>LAEP</u>	<u>CASEC</u>	<u>CEFR</u>	<u>Cambridge ESOL</u>
Advanced	n/a	B2	First Certificate in English (FCE)
Intermediate	550+	B1	Preliminary English Test (PET)
Pre-Intermediate	500~600		
Elementary	400~500	A2	Key English Test (KET)
Basic	~400		

The four levels of LAEP class described here, and a fifth level (Advanced) for second year students remained consistent until the end of the 2013 academic year.

2 Methodology and testing instruments

2.1 Subjects

Data for this study were collected from the first year university students who entered the various faculties that were participating in the Liberal Arts English Program (LAEP). In 2006, this included three faculties, namely Agriculture, Business Management, and Economics. English proficiency tests were taken by 1575 students in this year. In subsequent years up to 2013, three additional faculties, those of Human Studies, Pharmacy and Urban Planning became part of the program. Student numbers of CASEC test takers fluctuated with yearly faculty enrolment between a maximum of 1888 in 2008 and a minimum of 1712 in 2012. Over the eight years of the program in which CASEC scores were recorded, more than 13,000 students were tested. CASEC tests were taken by students at the start of the academic year in the first week of April. A further CASEC test was also then taken near the end of the academic year in December. Efforts were made to track down as many students as possible who failed to take either the April placement test or the test in December. Supplementary exam periods were scheduled for this purpose. Students who failed to take either the April or December test were removed from the statistical analysis of 12,262 students presented in this paper, a loss of 4% to 10% annually of student numbers.

2.2 The CASEC test

While not as well known as other Japan based proficiency tests such as TOEIC¹ (Test of English for International Communication) or STEP Eiken² (Test in Practical English Proficiency), the Computerized Assessment System for English Communication (CASEC) test has been used since 2001 and has over 150,000 test takers annually in Japan. The test is based on Item Response Theory (IRT) (see Lord, 1980) and claims to run an adaptive system varying the difficulty of items with the proficiency level of the test candidates. The test takes 40 – 50 minutes on average and was sat in the various computer rooms at our university usually on a faculty-by-faculty, all-in all-out rotational basis. In the first sitting in April, testing was conducted over two days. In the second sitting of the test in December, a one week period was

¹ See <http://ets.org/toeic>

² See <http://stepeiken.org>

allotted for students to write the test after regular fourth period classes had ended. Instructors were present at all times during the tests although issues of collusion were of little concern owing to the fact that tests based on IRT will invariably have different problems showing on different computers screens and because answers are recorded electronically. The CASEC test comprises of four sections: 1). vocabulary 2). knowledge of phrasal expression and usage 3). listening ability for the main ideas of conversations and 4). listening ability for specific information via dictation exercises.³

While some concerns surround such computer-adaptive language tests, particularly with issues of security (Carr, 2006) CASEC was regarded favorably at our institution for a long period. Clearly context and purpose of the test were ideal; test scores were available almost instantly and class placement was begun by program administrators on the day of testing. Anecdotal evidence from instructors in the LAEP also suggested that on the whole, students were very well placed in level appropriate classes.

It should be noted that at no time did instructors in the LAEP teach their lessons with the goal of improving CASEC scores per se. That is, classes were not designed like those of cram schools focusing on CASEC test taking skills. CASEC can be seen as a general proficiency test lacking a speaking component. Standard for-credit university classes in the LAEP were designed to teach the four-skills using textbooks aimed at Cambridge ESOL test takers. LAEP program coordinators were conscious of the type of negative wash-back discussed by Chapelle & Douglas (2006) that could have resulted in simply teaching for improvement of the limited skills required by computer-based tests. Rather, instructors focused on the communication skills that were required by the four-skills approach seen in the Cambridge ESOL exam suite. However, because CASEC is a proficiency test, there is certainly a reasonable expectation that improvement in general English ability should lead to higher scores in CASEC. To ensure that students were motivated to take the December CASEC test, 20% of their second semester grade was assigned to their performance.

2.3 KET/PET tests

Because Cambridge ESOL was central to the LAEP in the early years, many instructors who taught in the program became qualified Cambridge KET/PET/FCE Oral Examiners. During the eight years of the LAEP official KET/PET tests were administered at our university on three occasions. On two of these occasions the tests were held at the end of the academic year. The aim of these tests was not to gauge whether students English ability had improved, so consequently, are not directly related to the aims of this paper. These tests were done as an internal check to make sure the initial placement of students was to appropriate class levels and to subsequently adjust the CASEC bands of scores for placement if necessary. The third occasion when accredited ESOL examinations took place using students in LAEP is reported in an independent study by Nagao et al. (2012). Their research focus was on the attitudes of teachers and students in the LAEP to a PET-based curriculum and the data they report regarding performance in PET is longitudinal so their results have greater bearing on the present study. This research will be further discussed in the following section.

³ See <http://global.casec.com/about>

3 Results

3.1 Overall yearly improvement

The change in CASEC scores between the April placement test and the end of the academic year December test is shown for all first-year students from 2006 to 2013 in Figure 1. Paired-samples dependent t-tests were conducted separately for each year, determining that student scores had improved significantly in each year of the program (paired t-test $P < .001$) (see Table 2). While perhaps a pleasing result at face value, the low magnitude of improvement suggests that a further examination of the data is necessary to reveal what trends may be present.

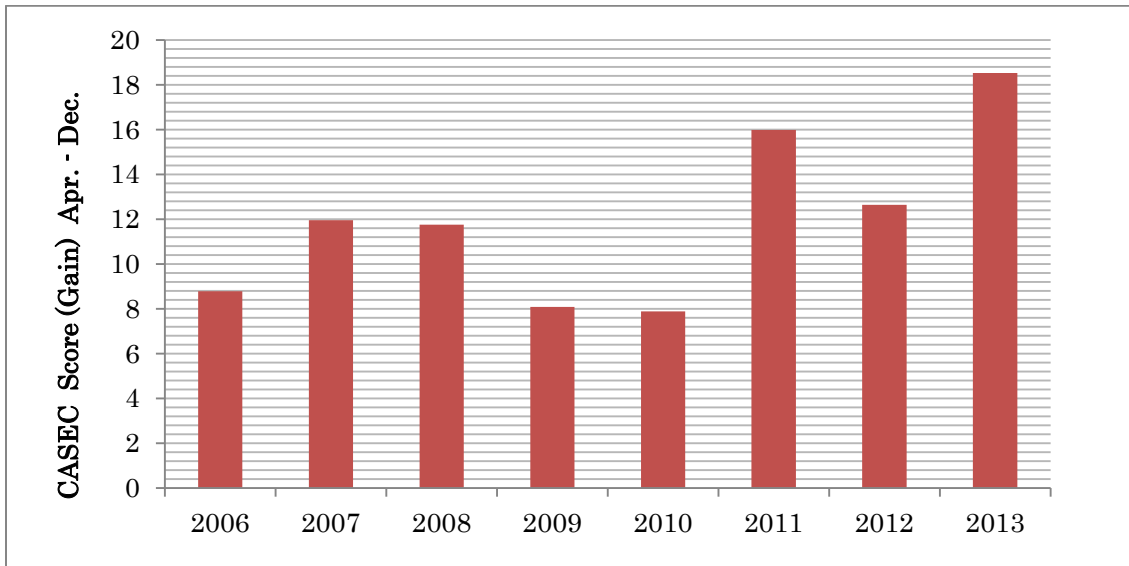


Fig. 1. Yearly average change in CASEC score (Apr. to Dec.) all students (2006-2013)

Table 2. Descriptive statistics for change in CASEC score (Apr. to Dec., 2006-2013)

	2006	2007	2008	2009	2010	2011	2012	2013
Apr. (mean)	465.65	460.69	452.79	456.23	447.56	446.42	437.53	442.56
Dec.(mean)	474.44	472.65	464.54	464.38	455.45	462.41	450.17	461.09
Mean diff.	8.79	11.96	11.75	8.09	7.89	15.99	12.64	18.53
DF	1149	1451	1685	1612	1551	1561	1628	1620
t-ratio	5.58	8.62	9.10	6.17	5.64	11.81	8.81	12.84
Probability	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

3.2 Analysis by class level

To determine whether gains in examination scores were consistent throughout the student body, CASEC scores for the four class levels of the LAEP were examined. Scores were compiled for students from Basic to Intermediate level and improvement in CASEC scores over the academic year was measured. A graphic representation of the results appear below in Figure 2. Students of lower initial English ability who joined the LAEP improved more than students of a greater

initial ability. Mean improvement for Basic class students was the largest at 34.52, while mean improvement for the Intermediate class was close to zero. The bottom line of each bar in the chart below shows the average CASEC score in April. The top line of the bar shows the average score at the end of the year.

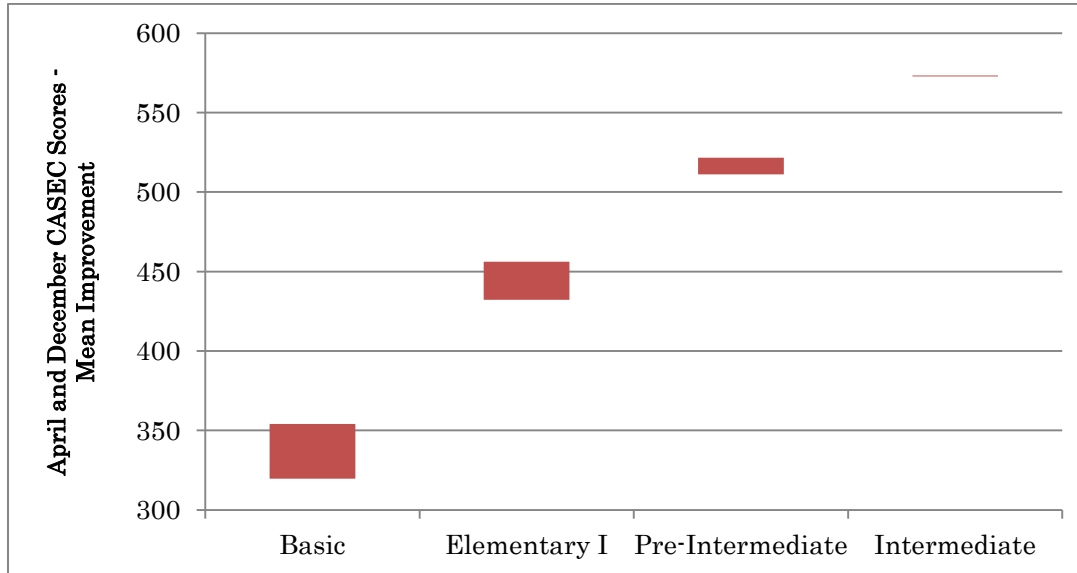


Fig. 2. Gain in mean CASEC score by class level (2006-2013 aggregate)

As has been previously noted, the CASEC test students took in April was used as a placement test. Based on a student's score, he or she would be placed in a class of commensurate ability ranging from Basic to Intermediate. While all efforts were made to make this placement as appropriate as possible, examination of CASEC threshold scores showed a large variation from year-to-year and from faculty-to-faculty. The reason for this was that when a faculty had fewer than say 20 Intermediate students, Intermediate classes were not created. In this case, students would have studied in the lower level Pre-Intermediate class. With four different class levels of ability and six faculties, situations such as this occurred somewhat frequently.

Consequently, it should be seen that the wide variation in CASEC threshold scores for entry to LAEP classes makes any statistical analysis impractical - such an analysis would be so specific to the circumstances of the LAEP that generalizations could also not be made easily. To counter this, it was decided that a statistical analysis of students' improvement in CASEC scores independent of LAEP class level was necessary. If it is possible to tie CASEC scores back to a standard reference of ability such as CEFR it would make the current research more meaningful to university programs outside of the LAEP. To do this, it is important to briefly examine the results of Cambridge ESOL exams taken by students at our institution.

3.3 Cambridge ESOL KET and PET examinations

Several years into the administration of the LAEP it was decided to gauge the performance of students on the official KET and PET tests. This plan was instigated with the aim of confirming that our students were studying level appropriate material, or more specifically, that based on the CASEC score we used for ranking and placing students, that they were studying at ability

commensurate Cambridge ESOL levels. Unfortunately due to the costs of administering KET and PET we weren't able to test many students. There was also a tendency on the part of teachers and administrators to select students who had done well over the academic year to take the test and hence may have been outliers in terms of score and motivation. Information on the CASEC homepage relying on self-reported TOEIC(LR) scores suggests that a CASEC score of 500 may approximate 400 on TOEIC(LR). Research by Tannenbaum and Wylie (2005) established threshold or cut scores for TOEIC(LR) in relation to CEFR, finding agreement among expert panelists that a cut score (the minimum level required to qualify a candidate) for B1 was 550 on TOEIC(LR). Although some extrapolation of this result was necessary to set our bands, it was decided that with 400 TOEIC level a student could most probably pass KET (A2), and be ready to study at the B1 level. With the limited data we were able to obtain through official ESOL examinations of our own students we also found this approximation useful. Of the 20 LAEP students who took KET in 2009 for example, the lowest CASEC score recorded that led to a pass in KET was 506. From this perspective, it seems reasonable to conclude that students above the level of CASEC 500 are linguistically capable of beginning to study at the PET or B1 level. The class structure employed by the LAEP reflected this division, as shown in Table 1.

3.4 Analysis of improvement based on ability at university entrance

An analysis of the CASEC scores obtained from 2006-2013 was carried out with a CASEC score of 500 being used as a cut-off – the underlying idea, as detailed above, being that this represents an appropriate level separating students above and below B1 CEFR. Paired-samples dependent t-tests were conducted on the data obtained over the entire eight years of the LAEP involving some 12,262 students. Improvement in English language proficiency over the course of one academic year was gauged by comparing April and December test scores. It was found that a significant improvement was evidenced for students of A2 ability (CASEC < 500) in end of year testing (M= 419.27, SD= 56.60) when compared with English proficiency at university entrance (M= 398.60, SD= 56.60); paired $t(8215) = 33.11$, $p = .0001$. For students above the 500 CASEC level, those judged as CEFR B1 and above, a comparison of the April (M= 555.93, SD= 45.17) and December scores (M= 550.88, SD= 45.17) however, found no improvement $t(4045) = -7.10$, $p = .0001$. On average, mean scores had decreased by 5.04 points. The score between first and second attempts at CASEC decreased in 52.2%, improved in 46.8% and was identical in 0.9% of test takers. A negative t-value recorded in the pair-wise analysis suggests that students' ability was actually more likely to decrease. A graphic representation of the analysis of performance is shown in Figure 3.

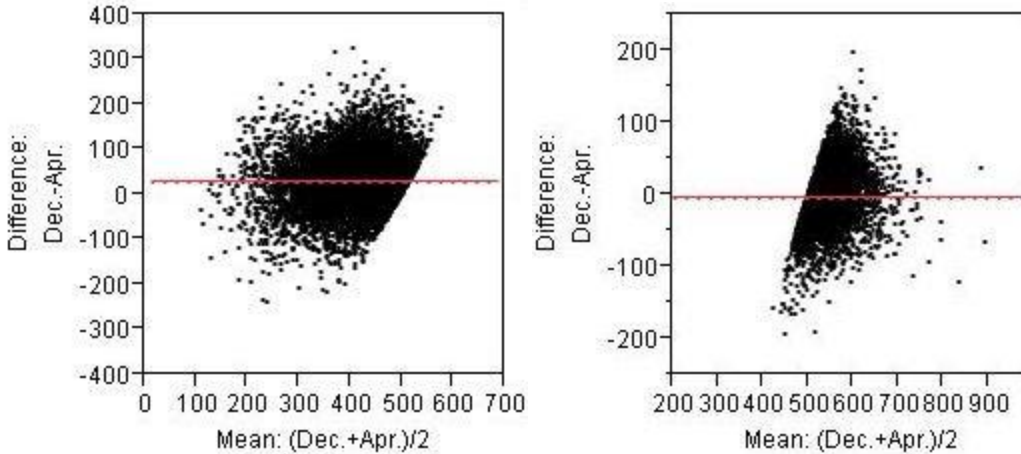


Fig. 3. Scatter plot distribution of CASEC scores <500 (left) and >500 (right)

Note the difference between the December and April scores (y) is plotted against the mean of the December and April scores (x). Students scoring below 500 (CEFR A2 or below) were more likely to score higher at the end of the academic year, thus showing an improvement in proficiency. Students who entered university at a higher level (over 500 CASEC or CEFR B1 level) had difficulty in maintaining their proficiency, tending to score lower in December. The analysis suggests the proficiency level of these students tended to decrease despite the year of English study.

3.5 An independent study of PET level students studying in the LAEP

Nagao et al. (2012) report an independent study involving 31 students studying in Pre-Intermediate and Intermediate classes in the LAEP (at the same institution as the current study). Their research focuses on the attitudes of teachers and students towards the PET-based curriculum; however, they also state they document and analyze students' performance in the test over the course of an academic year. At no point does Nagao et al. (2012) mention CASEC test scores of their students, but given that CASEC was the standard for placement to class levels in the LAEP, we can assume that student scores were over or well over 500. As has been shown above in the current research, the group of students more proficient at university entry is the hardest to affect a positive change in. Likewise, Nagao et al. (2012) reports that while a minor improvement in speaking ability was seen between the official April and December PET tests, scores for the other sections of the test (reading, writing and listening) all slightly decreased. They also report that students were surveyed on their expectation of their test performance prior to the second sitting of the PET test and that 70%, when asked, 'Do you think your score will be better than last time?' answered 'no'. Nagao et al. (2012) say little to address why a curriculum based around the PET such as LAEP, has failed to produce an improvement in PET scores. And yet, it is clear that not only did the students' English ability measured by the PET test fail to show an increase over the academic year, students themselves would not have been surprised by this result.

4 Discussion and conclusion

Various examinations of the data from eight years of the LAEP have shown us:

1). Overall, first-year non-English majors made minor, yet statistically significant gains in English proficiency over the period of one academic year. Results were consistent in any year of the program (2006-2013), regardless of the form of the LAEP curriculum and resulting syllabi.

From 2006 to 2013, average mean gains in English proficiency varied between 7.89 and 18.32 points. It is debatable whether this small numerical average gain alone would equate to a real-life perceptible improvement in English ability. If a student can advance only a handful of points each year, it may take an impractical number of years to achieve a high level of proficiency. Despite the statistical gain of the group as a whole, examination of individual performances suggests that only minority outperformers seem to have benefited greatly.

From a managerial perspective, the first few years of the program entailed far greater work and commitment from instructors towards the accomplishment of a common goal. It was undoubtedly much more expensive for the university to run this type of program as it required the time and effort of administrative staff. That the measured increase in student ability was no greater for an intensively managed program than a loosely managed one and that in either case the overall improvement was low, calls for a fundamental reassessment of the two-lesson a week approach to compulsory English for non-English majors (in Japan).

2). The more difficult the level of class content (seen from Basic to Intermediate) the smaller the improvement in English proficiency scores.

The greatest improvement in CASEC was seen in students who began at the lowest level. There are two possible reasons for this: 1). the LAEP worked better with low-level students, or 2). lower-level students had less prior contact with English than their more advanced peers. Looking at the question of whether tertiary English education had been successful in such a large scale program as the LAEP, however, it is hard to argue that the result with the lower-level learners brings much satisfaction. If for example, a student enters university at a low A2 level, and after a year of study, their level has improved to a slightly higher A2 level, even though improvement has been realized, the bar seems set too low to call it true 'success'. Students would have to continue at a high rate of progress for several years running to reach the B2 or higher levels. As a university graduate there is a societal expectation of proficiency that probably isn't met by an A2 standard in Japan. Employers focusing on 'can do's' will also have demands of their future employees in the work place that A2 level will have difficulty in coping with (see Schmidt, 2010). As such, it is perhaps fair to say that at least a B1 level is desirable. One year of English will not be enough – English programs should last for two or more years if they are to have a greater chance of fulfilling their purpose.

3). CASEC proficiency test scores were equated with the more commonly used CEFR scales. It was found that the English ability of non-English majors who entered university at the B1 level or above, actually most probably decreased by only studying English twice-a-week. Independent testing in PET carried out by Nagao et al. (2012) adds support to this finding.

While in principle instructors may want to rail against this result, for many it may also be clarification of an intuitive belief formed over years of teaching at university level in Japan: students' English simply isn't getting better with the limited amount of time they spend studying it. Most high school students in Japan study English daily and then cram for university entrance examinations. After entry, in the case of faculties where English is not the students' major, it may only be taught once or twice a week. In such cases, the number of hours students devote to English will decrease dramatically compared to their high school years and as the current research indicates, higher-level students may even have difficulty in maintaining the English level that they formerly achieved, much less improving upon it.

The English proficiency tests, CASEC, KET and PET discussed in this study, were unknown to the students at our university prior to encountering them in the LAEP. It would be understandable if many students, in fact, had scored lower comparatively in their first attempt at the tests, due to unfamiliarity. Practice effects improving performance in tests of cognitive ability are well documented in research literature. In a meta-analysis of 50 studies, Hausknecht (2007) notices that practice effects are even greater when test coaching accompanies the practice. In the early years of the LAEP there was considerable training for taking KET and PET tests as this was the focus of the test-based curriculum. Consequently, the results and implications of the present study – low or no improvement in academic-year-end tests in CASEC and PET for higher level students – appears even more worrisome.

It is very important for universities to develop programs that work. Non-English majors in compulsory English classes may just be the hardest group to work with. And yet, if this group were able to achieve a higher level of English proficiency – when combined with the knowledge gained from studying their majors – would have the potential to contribute the most. The question of why our teaching isn't more successful is both antagonist and motivator for the enthusiastic instructor. The current research may not provide an answer to this question but it does highlight it – and this is important in a field where the data often seem to point to gains, and yet the classroom environment is a constant of mediocrity. The LAEP will end entirely at the conclusion of this academic year and individual faculties will decide on their own English language education policy and curriculum. What form these take will hopefully depend on the lessons learnt here. In closing, I would also like to acknowledge the tireless efforts of the staff and instructors who worked for the LAEP over the years and also to applaud the LAEP on its transparency and accountability, without which the current research would not have been possible.

References

- Capel, A., & Sharp, W. (2005). *Objective KET*. Cambridge: Cambridge University Press.
- Carr, N.T. (2006). Computer-based testing: Prospects for innovative assessment. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching. CALICO Monograph Series, Vol. 5* (pp. 289-312). San Marcos, TX: CALICO.
- Chapelle, C., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge: Cambridge University Press.
- Council of Europe, (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T., Moriarty G., & Meghan, O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*(2), 373-385.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum Associates
- Nagao, J., Tadaki, T., Takeda, M., & Wicking, P., (2012). The attitudes of teachers and students towards a PET-based curriculum at a Japanese university. *Cambridge Research Notes, 47*, 27-36.
- Naylor, H., & Hagger, S. (2004). *Insight into PET*. Cambridge: Cambridge University Press.
- Schmidt, M.S., Naganuma, N., O'Dwyer, F., Imig, A., & Sakai, K. (Eds.) (2010). *Can do statements in language education in Japan and beyond*. Tokyo: Asahi Press.
- Tadaki, T., & Minehane, G. (2007). Issues in developing programme evaluation procedures for a tertiary institution – a case study: Evaluating the Liberal Arts English programme at Meijo University. *Journal of Research Institute of Meijo University, 6*, 135-144.
- Tannenbaum, R., & Wylie, E.C. (2005). Mapping English language proficiency scores onto the Common European Framework. *TOEFL Research Reports, 80*, 1-53.