Evaluating a New Course in English for Academic Purposes

Yoshinori WATANABE

英語総合演習の効果測定

渡　部　良　典

Abstract

The purpose of the present paper was to evaluate a new course entitled English Comprehensive that was implemented in 2004. The purpose of the course was to offer advanced courses to those students who were not satisfied with the two-period per week required English courses. English Comprehensive was intended to help students acquire language skills, communication skills and academic skills. The evaluation study that has been presented and discussed in the present paper is summarized as follows. The course had positive effects overall in that it helped students improve language skills, including reading, writing, listening, vocabulary and grammar. The students themselves were also aware that they had improved in these skill areas, though the improvement still remained at a basic level rather than advanced levels. There was a tendency among the course takers that the goal of the course was too high. What needs to be done to make the course more effective in the future involves an attempt to render the course more challenging so students may really improve their skills, without giving them a sense of being overwhelmed.

Introduction

All the students at Akita University are required to complete two English courses (i.e., Listening Comprehension and Writing) during their freshman year. However, it has long been noticed, at least among the English teaching staff, that the typical schedule of two sessions per week is not sufficient, particularly for those students who wish to major in English education. English Comprehensive, an elective course implemented in 2004, was established to provide such students with more opportunities to improve their proficiency in English as well as various academic skills. The present paper reports on the effectiveness of the course that ran during the spring term of the year 2004. Effectiveness was measured by referring to various sources, including an external test, self-evaluations, and an in-house TOEFL type writing test.

FAX：018—889—2639
e-mail：yjwatana@ed.akita-u.ac.jp

Contents and structure of the course ·

In the English Comprehensive course, students were offered two 90-minute sessions during the spring term, which amounted to a total of 30 sessions. Twenty-two students initially signed up for the course, and 20 students completed it. All the students were from the Faculty of Education and Human Studies; 14 students were from the department of International Language and Culture and 6 students were from the department of School Education.

For this course, students were given a formal syllabus, specifying the entire outline. In addition, a guideline was given to students at the beginning of each lesson providing them with suggestions about learning focuses, procedures, and homework assignments. The guideline was based on what had been accomplished during each lesson. This is similar to the 'process syllabus' presented by Breen (1984), in that the present syllabus was intended to construct "a plan for the gradual creation of the real syllabus of the classroom, jointly and explicitly undertaken by teacher and learners" (p.52, italics original). A sample guideline is provided in Appendix A of the present paper. The guideline having the process nature does not mean, however, that there was no pre-planning as to what to teach prior to the course. The flexible component involved only the way of arranging teaching and learning focuses that I had had in mind before embarking on the course. I had listed each focus in a grid prior to the term, and upon finishing each lesson I placed a tick mark in the cell of the skill which was covered. The skills covered in the course and the frequency of covering each of them are displayed in Table 1.

The target skills were divided into three broad components. The first component involved language skills, including writing, vocabulary, reading, and listening, and various sub-skills of each. The second and third components were Academic Skills and Communication Strategies respectively, each of which was further divided into sub-skills as shown in the table. Each box of the second section of the table indicates one lesson, amounting to a total of 28 boxes in each sub-skill area. The last row gives the total number of lessons that had eventually dealt with each sub-skill by the end of the course. The content coverage of the course was spiral, in the sense that the same teaching and learning focus was repeated, when I was aware that that particular skill was not sufficiently acquired by the students.

It can be observed from Table 1 that there were several tendencies that might have influenced the students' achievement throughout the course. First, the largest number of lessons was devoted to developing writing skills. In addition to receiving instructions regarding how to write a formal essay, students were required to write a journal every day and submit it twice a week. The journal notebook was returned with my short comments on it by the end of the day of submission. The comments were only on content rather than language. In the first lesson of the course, students had received instructions as to how to write a journal; that is that they should keep a journal every day, that quantity rather than quality would be taken into account in the final grade, and that they should write about an academic topic, such as the course they were taking at university, world news, and so forth rather than daily chores, such as shopping, weather, dating and so on.

The table also shows that an emphasis was placed on developing presentation skills. A total of 14 lessons were devoted to this particular skill. The instructions included

Table 1

Learning Focus

Target Skills	Content coverage in each lesson	Frequency
Language Skills		
Writing		
Describing material	□■■□■□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	3
Describing person/personality	□□■□■□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	2
Definition (single)	□□■□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	1
Definition (stipulated)	□□□■□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	1
Journal writing	□■■■■■■■■■■■■■■■■■■■■■■■■■■■■■■	29
Reading		
Reading for key words (skimming)	□□□□□□□□□□□□□□□□□□■□□□□□□□□□□□□	1
Understand relations between key words	□□□□□□□□□■□□□□□□□□□□□□□□□□□□□□□	1
Reading for details	□□□□□□□■□□□□□□□□□□■■■□□□□□□	4
Listening		
Listening for main ideas (note-take key words)	□■□□■□□□□■□□□□□□□□□□□□□□□□□□□□	3
Vocabulary		
Vocabulary (personality)	□□□■□□□□□□□□□□□□□□□□□□□□□□□□□□□	1
Vocabulary building (network)	□□□□□■■□□□□□□□□□□□□□□□□□□□□□□□□	2
Vocabulary building (& reading)	□□□□□□□□□■□□□□□□□□□□□□□□□□□□□□□	1
Grammar		
Rhetoric	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	0
Review of basic grammar	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	0
Styles	□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□	0
Academic Skills		
Dictionary use	□□□□□□□□□■□□■□■□□□□□□□□□□□□□□□	3
How to take an essay test	□■□□□□□□□□□□□□□□□□□□□□□□□□□□□□	1
How to take an objective test	□□■■■□□□□□□□□□□□□□□□□□□□□□□□□□	4
Note taking	■□■■□□□□□□■□■■□■■■■■□□□□□	12
Presentation (within small group)	□□■□□□■□□■□□■□■□□□□□□□□□□□	5
Presentation (in the whole class)	□□■□□□□□□□□□□□□□□□□□□□■■■■■■■	9
Idea mapping/text-graphing	□□□□□□■□□□□□□□□□□□□□□□□□□□□□□	1
Self-evaluation	□□□□□□□□■□□□□□□□□□□□□□□□□□□□□	1
Synthesizing information	□□□□□□□□□■■■□■□□■□□□□□□□□□□	5
Communication Strategies		
Asking for information	□□□□□□□□□■□□□□□□□□□□□□□□□□□□□	1
Negotiation of meaning		
Interrupting conversations to take turns	□□□□□□□□■□■□□□□□□□□□□□□□□□□	2
Asking for definition	□□□□□□□□□□■□□□□□□□□□□□□□□□□	1
Checking spelling, pronunciation, or grammar	□□□□□□□□□□■□□□□□□□□□□□□□□□□	1
Asking for repetition	□□□□□□□□□□■□□□□□□□□□□□□□□□□	1
Confirming meaning	□□□□□□□□□■□□□□□□□□□□□□□□□□□	1
Agree/disagree	□□□□□□□□□□□■□□□□□□□□□□□□□□□	1
Responding to an utterance	□□□□□□□□□□□□■■□□□□□□□□□□□□□	2
Making an appointment	□□□□□□□□□□□□□■■□□□□□□□□□□□□	2
Leading a group discussion	□□□□□□□□□□□□□□□■■■■□□□□□□□□	4
Report on a group discussion	□□□□□□□□□□□□□□□□□□□■□□□□□□□	1

how to select an appropriate topic, how to prepare note-cards, how to organize ideas with an effective introduction, body, and conclusion, and various expressions that would be useful for effective presentation (e.g., "Today, I'd like to talk about …," "As I mentioned before …," "The next … I'll talk about…" etc). Students were then asked to give a presentation to a small group to receive comments, and subsequently to give a fully prepared presentation to the whole class at the end of the course.

Yet another skill that received a major emphasis was note-taking. The purpose of the exercise involved helping students develop a variety of listening strategies as well as note-taking skills per se. The target listening sub-skills involved how to pay focused attention to key words, top-down processing of auditory input, summary of the content, and so on. The content of the audio-visual material used for practice included narrations of various cultural topics, including Turkey, Egypt, USA, etc.

Normally, the note-taking exercise went as follows: first, students watched a video segment, which lasted approximately 10 minutes. After watching the video, they were asked to remember the words that stayed in their mind. They were then asked to compare notes with other students in a group. Several students were called on to write the words they remembered on the chalkboard, while other students copied the words they failed to note. Students then watched the video a second time and instructed to check whether the words on the chalkboard were really the ones which were used in the video segment. They then watched the video a third time to take notes of as many words as possible, which they subsequently divided into key words/phrases and peripheral words/phrases. Students eventually used these words and phrases to summarize the content of the video segment. After writing a summary, they were given a script and watched the video segment again. Despite these emphases on several particular skills, there were skills which were not covered though they had originally been intended. For example, there was no explicit instruction on grammar. Though I corrected errors of grammatical inaccuracy in their formal essays, I kept such corrections to a minimum. When giving feedback to journals, I did not correct grammatical errors at all:

instead I focused only on contents. There was also very little instruction on vocabulary. Indeed, students were given suggestions with examples as to how to expand vocabulary by using the notion of semantic network (relating a group of synonyms and antonyms), and by means of intensive reading of a text. However, no explicit exercises were given, nor did I examine whether students actually used the suggested methods in their own private study.

Despite the failure to cover several skills that had originally been intended, the medium of instruction was English throughout the course, and students were also required to use English. Thus, students spent a large amount of time using English, which in turn implies that students must have had chances to improve four skills including reading, writing, listening and speaking, though it would still be insufficient to help students develop in the areas of vocabulary and grammar.

In summary, the effectiveness of the program was expected to be observed in the form of increased proficiency in the four skill areas (i.e., reading, listening, writing and speaking) in general, and writing in particular. However, increases in proficiency might not be remarkable in vocabulary and grammar.

Instruments for evaluation

In order to measure the effectiveness of the course, various instruments were prepared. The first instrument was the Assessment of Communicative English (ACE) test, developed by Association for English Language Proficiency Assessment. This test is intended to measure the English use ability of intermediate and advanced levels of senior high school students. The test was based on Item Response

Theory (IRT), "a systematic procedure for considering and quantifying the probability or improbability of individual item and person response patterns given the overall pattern of responses in a set of test data" (Henning, 1989, pp. 108-109). Among three families of IRT, the two-parameter model was employed for the present version of the test, taking into account a person's ability and item difficulty, and a continuous estimate of discriminability. The test consists of three components; Listening (30 items), Reading (20 items), and Grammar/ Vocabulary (48 items). Two parallel versions of the test were administered at the first class of the term in April (version AL006), and at the end of the term in July (AL009). (See http://www.kirihara-kyoiku.net/BACE-ACE/ for details of the ACE test).

To measure writing proficiency, a sample TOEFL essay topic was used. The topic that was used for the present evaluation was as follows: "People work because they need money to live. What are some other reasons that people work? Discuss one or more of these reasons. Use specific examples and details to support your answer." The same topic was used for the pre-course and post-course evaluations. When students took the test in the first class, they were not informed that they would take the same test again at the last class. The written scripts were marked by two independent raters who had attended a training session. The criteria was a type of holistic scaling, whereby the scripts were evaluated without dividing the competence into sub-component writing skills. The score ranged from 6 ("demonstrates clear competence in writing on both the rhetorical and syntactic levels, though it may have occasional errors.") to 1 ("demonstrates incompetence in writing") (Educational Testing Service, 1996). Eventually, inter-rater reliability reached .92 (p<.001) measured by Spearman's rank order correlation statistic.

Another evaluation instrument was self-assessment. It is commonplace to administer self-assessment of students' performance in EFL. The major use of self-assessment has been put to raise students' awareness and monitoring of competence and as a part of their final grade. The purpose of the present evaluation was to help students monitor their level of proficiency, with the expectation that it would enable them to become aware of what they had achieved and the weakness that needs further improvement. In order to do so, a set of specific descriptors were necessary, so students could understand at which level they were and what areas need further improvement. For this purpose, the self-assessment component of DIALANG was employed. The DIALANG self-assessment system is composed of three skill areas including reading, listening and writing. Each skill is divided into three broad levels (i.e., basic user (A), independent user (B), and proficient user (C)), each of which is further divided into two levels, amounting to a total of six levels of proficiency, A1 and A2, B1 and B2, and C1 and C2. Students were asked to rate their proficiency according to a five-point scale, ranging from 'very difficult' (1) to 'very easy' (5). For the present evaluation, a Japanese version was used. The translation was made from the original English version with the cooperation of Satsuki Shirasawa, a previous student of Akita University, who was writing her senior thesis using the instrument in 2003. Sample items are as follows:

Reading A1: I can understand the general idea of simple informational texts and short simple descriptions, especially if they contain pictures

which help to explain the text.

Listening A1: I can understand everyday expressions dealing with simple and concrete everyday needs, in clear, slow and repeated speech.

Writing A1: I can write simple notes to friends.

For students to evaluate their own proficiency at each of the six levels for the three skills, DIALANG included more than one items. The number of such items that were included is shown in Table 2. (See Council of Europe (2001) for further details.)

Table 2
The number of DIALANG self-assessment items

	A1	A2	B1	B2	C1	C2
Reading	5	9	8	8	4	2
Listening	4	10	9	9	9	2
Writing	6	7	7	6	5	5

In addition to these instruments, the university-administered course evaluation was also employed. This was the one which was implemented by the Foundation Course Administrators of Akita University. There were a total of eight items, about each of which students were asked to rate according to a six-point scale, ranging from '5 strongly agree' to '1 do not agree at all.' Sample items are as follows:

5　Strongly agree（そう思う）

4　Agree（どちらかといえばそう思う）

3　Uncertain（どちらともいえない）

2　Disagree（どちらかといえばそう思わない）

1　Strongly disagree（そう思わない）

0　I don't know/Not applicable（分からない、または該当しない）

1．The purpose of the course and goals were specified.（授業の目的や到達目標は明示されていた）

2．Teaching materials were well prepared（教材はよく準備されていた）

3．Explanations (spoken and written) were easy to follow.（説明や板書は明確で十分にわかりやすいものだった）

4．The course motivated me（授業内容は興味関心・学習意欲をそそるものだった）

5．The lesson proceeded according to the levels of my understanding.（理解度に応じて進められていた）

6．I acquired the content of the lesson sufficiently.（授業の内容を十分に習得できた）

7．The instructions regarding self-study were appropriate.（授業の予習・復習（課題・問題演習等）の指示が適切に行われていた）

Evaluation of the effectiveness of the course

Students' Course Evaluation

The results of students' course evaluation are provided in Table 3. The students' responses were favorable overall, with the means higher than 4 on all the items. The only exception was item 5 (Mean = 3.78, SD = 1.31), which asked if the course proceeded according to the levels of understanding. This result seems to contradict that of item 3 (if the content was easy to follow), with a mean of 4.17. This is ironical too, because the course employed a type of process syllabus whereby I could adjust the content and pace of lessons according to the achievement of students. From the instructor's point of view, however, the course was offered to a group of advanced students, and I understood that those students would take English-related majors in the future, so I set the level of the course well beyond the supposed level of ordinary students. It may be debatable how students' subjective impressions

about the level of the course would relate to their actual progress in performance. The question will be answered in part in the sections that follow where results of actual progress in proficiency will be shown.

Table 3

Students' course evaluation

	1	2	3	4	5	6	7
Mean	4.94	5.00	4.17	4.61	3.78	4.06	4.82
SD	0.24	0.00	1.04	0.70	1.31	1.00	0.39

Progress in language skills

Students' progress in language skills was measured by the ACE test. The results are shown in Table 4. (All the data were analyzed by SPSS, Version 11.5). The table shows that they improved in all areas including vocabulary, grammar, reading and listening at the level of .01 to .05 levels. It was expected that students would make progress in the areas of reading and listening. However, contrary to the expectation, they increased in vocabulary and grammar as well, which may imply that even without explicit instructions they had chances to incidentally learn vocabulary and grammar.

It has been noted that the greatest emphasis was placed on developing writing skill throughout the course, so not unexpectedly writing scores increased. To assess the aspects of progress more objectively, analyses were made on the numbers of words, sentences, and words per sentence that were used in the writing. As shown in Table 4, the number of words and sentences increased significantly. However, the results did not show any increase in the number of words per sentence. This means that students became able to write more sentences by using more words, but not longer sentences.

Self-evaluation

As has been observed in the above sections, there were significant increases in students' overall proficiency levels, which could be interpreted to prove the effectiveness of the course. The purpose of self-evaluation was then to check to see if students would realize their proficiency had improved. Notice that the use of self-

Table 4

Differences between Pre-and Post-ACE Tests and Essay Writing Tests

	Mean	Std. Deviation	Minimum	Maximum	Z	Sig.(2-tailed)
VOC1	105.94	11.28	86.00	133.00	-2.22(a)	.026
VOC2	98.39	10.87	80.00	111.00		
GR1	112.72	12.87	89.00	141.00	-3.64(a)	.000
GR2	90.11	9.62	70.00	105.00		
RD1	219.94	31.95	188.00	300.00	-2.30(b)	.022
RD2	242.17	35.65	165.00	300.00		
LIST1	203.00	16.68	177.00	248.00	-2.77(b)	.006
LIST2	224.94	33.96	191.00	300.00		
WRIT1	2.14	0.48	1.00	3.00	-3.28(a)	.001
WRIT2	3.14	0.82	1.50	4.50		

Notes: N = 18. VOC1 = pre-course vocabulary test. VOC2 = post-course vocabulary test. GR1 = pre-course grammar test. GR2 = post-course grammar test. RD1 = pre-course reading test. RD2 = post-course reading test. LIST1 = pre-course listening test. LIST2 = post-course listening test. WRIT1 = pre-course writing test. WRIT2 = post-course writing test. (a) = Based on positive ranks. (b) = Based on negative ranks. (c) = Wilcoxon Signed Ranks Test

Table 5

Differences in the Number of Words, Sentences, and Word per Sentence between Pre-and Post-Essay Writing Tests

	Mean	Std. Deviation	Minimum	Maximum	Z	Sig.(2-tailed)
WDT1	132.44	35.20	72.00	187.00	−3.245(a)	.001
WDT2	176.00	45.46	75.00	251.00		
STT1	12.72	4.36	7.00	21.00	−2.206(a)	.027
STT2	15.39	4.96	7.00	23.00		
WDST1	10.90	2.45	7.06	15.29	−1.111(a)	.267
WDST2	11.98	2.96	8.50	18.09		

Notes: N = 18. WDT1 = the number of words in pre-course test. WDT2 = the number of words in post-course test. STT1 = the number of sentences in pre-course test. STT2 = the number of sentences in post-course test. WDST1 = the number of words used in each sentence in pre-course test. WDST2 = the number of words used in each sentence in post-course test. (a) = Based on negative ranks. (b) = Wilcoxon Signed Ranks Test

Table 6

Differences in DIALANG Self-Assessment between Pre-and Post-Course Periods - Listening[b]

	Level	Mean	Std. Deviation	Minimum	Maximum	Z[a]	Sig.(2-tailed)
A1	1	3.58	0.66	2.75	4.75	−2.594	.009
	2	4.11	0.72	2.50	5.00		
A2	1	3.09	0.56	2.20	4.40	−2.826	.005
	2	3.57	0.71	2.30	4.70		
B1	1	2.85	0.43	2.11	3.56	−1.383	.167
	2	3.03	0.61	1.78	3.89		
B2	1	1.94	0.60	1.00	3.00	−3.123	.002
	2	2.24	0.65	1.11	3.22		
C1	1	1.50	0.56	1.00	2.67	−1.377	.169
	2	1.63	0.76	1.00	3.11		
C2	1	1.16	0.35	1.00	2.00	−1.890	.059
	2	1.31	0.57	1.00	3.00		

Notes: N = 16. 1 = Self-assessment at pre-course period. 2 = Self-assessment at post-course period. a = Based on negative ranks. b = Wilcoxon Signed Ranks Test

assessment for this purpose is based on the assumption that if students would assess their own levels of proficiency accurately, then it would be more likely that they would have greater chances to improve in the future. The present paper examined this assumption. In order to do so, the results of pre-and post-course self-assessments were compared. It was predicted that there would be statistically significant increases in post-course assessments as their performance increased. The results are given in Tables 6 through 8.

It should be noted that students evaluated their performance on six different levels, though each component of the ACE test was not divided into the same levels. By examining the relationship between the increase in test scores and various levels of self-assessment, it would be possible to understand in what way the increase in test scores would relate to students' awareness of their achievement.

In order to understand if students would have evaluated themselves higher at the end of the course than at the beginning, the

Table 7

Differences in DIALANG Self-Assessment between Pre-and Post-Course Periods - Reading[b]

	Level	Mean	Std. Deviation	Minimum	Maximum	Z^a	Sig.(2-tailed)
A1	1	4.11	0.68	2.80	5.00	−1.894(a)	.058
	2	3.67	0.49	3.00	5.00		
A2	1	2.95	0.41	2.78	4.44	−1.446(a)	.148
	2	2.22	0.57	3.00	4.89		
B1	1	2.03	0.66	2.38	3.75	−1.668(a)	.095
	2	1.50	0.71	2.50	4.13		
B2	1	4.45	0.61	1.00	3.13	−.543(a)	.587
	2	3.97	0.67	1.50	3.25		
C1	1	3.20	0.46	1.00	3.25	−1.338(b)	.181
	2	2.28	0.57	1.00	3.00		
C2	1	1.81	0.64	1.00	3.00	−.933(b)	.351
	2	1.41	0.64	1.00	3.00		

Notes: N = 16. 1 = Self-assessment at pre-course period. 2 = Self-assessment at post-course period. a = Based on negative ranks. b = Wilcoxon Signed Ranks Test

Table 8

Differences in DIALANG Self-Assessment between Pre-and Post-Course Periods - Writing[d]

	Level	Mean	Std. Deviation	Minimum	Maximum	Z	Sig.(2-tailed)
A1	1	3.75	0.46	2.83	4.50	−2.676[a]	.007
	2	3.42	0.49	2.83	5.00		
A2	1	2.79	0.47	2.86	5.00	−2.136[a]	.033
	2	2.27	0.55	2.43	4.00		
B1	1	2.19	0.50	1.88	3.63	−2.079[a]	.038
	2	1.59	0.52	2.38	4.00		
B2	1	4.30	0.78	1.00	3.17	−1.604[a]	.109
	2	3.88	0.65	1.67	3.67		
C1	1	3.10	0.48	1.00	3.00	−.179[b]	.858
	2	2.51	0.57	1.00	3.60		
C2	1	2.13	0.65	1.00	2.60	−.000[c]	1.000
	2	1.59	0.67	1.00	3.00		

Notes: N = 16. 1 = Self-assessment at pre-course period. 2 = Self-assessment at post-course period. a = Based on positive ranks. b = Based on negative ranks. c = The sum of negative ranks equals the sum of positive ranks. d = Wilcoxon Signed Ranks Test

results of pre-and post-course self assessments were compared. The tables show that overall students evaluated their performance more highly at post-course evaluation than at pre-course evaluation, particularly at lower levels of the scale, though the relationships between other variables were more complex. In writing, post-course self-assessment scores were higher at the levels up till B1, lower level of 'independent user.' However, in reading, the score increase in self-assessment was found only at the lower level A1, 'basic user.' These results may reflect the amount of instructions students received. There were quite extensive writing practice sessions, whereas there were only a few chances to practice reading.

In listening, the results were more complex. Self-assessment scores were significantly higher at the end of the course than at the beginning at the level of A1 and A2 (basic user). However, there was also a

Table 9

Correlations between self-assessment and proficiency levels

Levels	Listening		Reading		Writing	
	Pre	Post	Pre	Post	Pre	Post
A1	.492*	−.019	.191	−0.06	0.13	0.11
A2	.550*	−.028	.201	−0.23	0.27	0.29
B1	.456	−.130	.303	−0.39	0.41	0.25
B2	.085	.121	−.044	−0.03	0.29	0.07
C1	.056	.037	−.121	−0.44	0.36	0.15
C2	.053	−.113	−.135	−0.38	0.41	0.06

*= $p < .05$.

significant increase at B2, though there was not such an increase at B1. There was also a tendency, though not significant, that scores increased at the highest level, C2 (proficient user). In the case of listening skills, students were given chances to listen to an extended discourse, take notes, and summarize the content using the notes. This type of skill is ranked at higher levels in the DIALANG self-assessment scale. Students might have developed confidence in dealing with this type of academic listening skill. However, they did not have much chance to practice listening to basic content such as daily dialogue, which is ranked lower in the assessment system. This may mean that students are able to develop academic skills even without developing basic skills.

The above results, though encouraging, do not tell us how self-assessment scores related to their actual proficiency levels. In order to examine the relationship, correlation analyses were administered between the two sets of scores. The results are given in Table 9.

The table shows that there were significant correlations in very few areas. The only areas that showed correlations were A1 and A2 of pre-course assessments. This means that there was very little relationship between students' perception about their proficiency and actual performance levels measured by the ACE test.

The interpretation of self-assessment results in light of progress in proficiency measured by the ACE test seems to imply that students made a progress in language skill areas, and students seemed to be aware of the fact that they had improved. However, the levels of achievements the students were aware of seem to still remain at a basic user level defined by the DIALANG scale. It should be remembered that the students' course evaluation also indicated that the level of the course was felt to be too high. In order to make the course even more effective, then, it is important to develop teaching methods and contents, so students may feel sufficiently challenged but not overwhelmed.

Conclusion

The purpose of the present paper was to evaluate a new course entitled English Comprehensive that was implemented in 2004. The purpose of the course was to offer advanced courses to those students who were not satisfied with the two-period per week required English courses. English Comprehensive was intended to help students acquire language skills, communication skills and academic skills. The evaluation study that has been presented and discussed in the present paper is summarized as follows. The course had positive effects overall in that it helped students improve language skills,

including reading, writing, listening, vocabulary and grammar. The students themselves were also aware that they had improved in these skill areas, though the improvement still remained at a basic level rather than advanced levels. There was a tendency among the course takers that the goal of the course was too high. What needs to be done to make the course more effective in the future involves an attempt to render the course more challenging so students may really improve their skills, without giving them a sense of being overwhelmed.

It should also be noted that the present study did not assess how much progress students made in communication and academic skill areas, though these were two of the most important goals of the course. There are no external tests to measure these skills, which have well established reliability and validity. Future research could concentrate on constructing measurements which would evaluate students' achievements in these areas. Finally, the present research did not compare the effectiveness of the course with other mainstream courses, because of a lack of common measures. This is also a topic that should be investigated in the future.

References

Breen, M. 1984. Process Syllabuses for the Language Classroom. In C. J. Brumfit (ed.) *General English Syllabus Design*, Oxford: Oxford University Press, pp. 47-60.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment.* Cambridge: Cambridge University Press.

Educational Testing Service. (1996). *Test of Writing English Guide, Fourth Edition.* NJ: Educational Testing Service.

Henning, G. (1989). *A Guide to Language Testing: Development, Evaluation, Research.* New York: Newbury House.

Appendix A : A Sample Day-to-Day Guideline

English Comprehensives

4th session 4/20/2004

Review of previous lessons

Previous lessons have been intended to help you get used to using English. I suppose you have had very few chances to use English in an authentic (real-life) situation. However, in order to become able to use English, you need to use English, taking as many opportunities as possible. You may say "of course!" but it's easier said than done. You may have understood by now that you need self-discipline to use English throughout a class hour. Yet, remember, the more you use English, the more likely you will become able to use English. Simply put, if you want to become a good English user, the most important thing is to break down your psychological barrier.

In addition to breaking down a psychological barrier, you have learned various language and academic skills, as shown in Table 1 (separate sheet).

Lesson today

Today's lesson involves various tasks as follows:

1. Taking a vocabulary test.
2. Learning how to expand vocabulary (i.e., the words used to describe personality) by using a method called 'vocabulary network.'
3. Learning how to describe a function of an object.

Homework assignment

1. Keep on keeping a journal. Remember, you should write your journal every day in your notebook, and submit it in the box in front of my office (3-250). The more you write, the better writer you will become. (The more you write, the more likely you will receive a good final grade.)
2. Describe your personality in one paragraph. You may want to use the vocabulary that you will learn in today's lesson.
3. Describe one 'unusual' thing in one paragraph. Your paragraph should start with a sentence that defines the term.