

Introduction

Background

Student quality in terms of English learning has always been one of the hottest issues in education at all levels. The voice urging for reformation of English teaching and learning in higher education has never stopped since the late eighties. In response to the increasing public demand, the Ministry of Education (MOE) has put forward several important educational plans during the past decade, hoping to enhance student quality and to keep national competitiveness as well.

It was believed over 80% of the students in technological colleges and universities did not have the basic level of English proficiency (Huang, Huang & Feng, 2003). Starting from year 2001 and lasting for three consecutive years, the Language Training and Testing Center (LTTC) had been assigned by the MOE to conduct elementary level General English Proficiency Tests (GEPT) for students in higher technological and vocational education. Only listening and reading skills were tested. It was found that students' listening comprehension was weaker than their reading comprehension. The passing rates were 15.8%, 14.9%, and 18.1% respectively. The findings showed that over 80% of the students in technological colleges and university did not even have the same basic level of English competence as a 9th grader should do. According to an island-wide research report (GEPT Report on Technical College Test-Takers 2003), excluding the foreign language majors, some findings are summarized below: (1) test-takers from the national colleges and universities show higher English proficiency than those in private ones; (2) students specializing in business and management demonstrate higher English language proficiency than those in engineering; (3) as for the learning habits, 55% of the non-English majors have never studied English in their free time, 57% have never read any English newspapers or magazines before; and (4) 84% of the non-English majors have never written letters or email in English.

According to the final report on the 2008 Education Yearly Plan (Secretariat MOE, 2009), there were about 13,305 out of the total 55,900 test takers in

technological universities and colleges who had passed the elementary level of English proficiency tests, which was far below the number expected in the original 2008 Yearly Plan. As to the nation-wide Higher Education Evaluation in 2009, two of its performance indicators are first introduced to assess student learning achievement; one is the development of English graduation thresholds and the other, the percentage of students who attain the threshold levels by passing English proficiency tests (Taiwan Assessment and Evaluation Association, 2009).

Ability grouping is usually believed to enhance students' learning motivation and enable teachers to ease the instructional problems owing to diverse student levels. In spite of the controversial issues on its effects in the USA and the Great Britain, ability grouping has still been getting more and more popular among higher institutions in Taiwan for the past decade.

Generally speaking, ability grouping of the freshmen English classes is usually based on two major student scores; one is the English scores on college entrance examinations and the other, scores on the freshmen placement tests. The Testing Center for Technological and Vocational Education (TVE) is the authorized organization which is responsible for the joint entrance examinations for technological and vocational institutions in recruiting students. The TVE Joint College Entrance Examinations are designed for high school graduates of the technological and vocational senior secondary education. Out of the total amount of students admitted in the technological colleges and universities in 2008, more than 85% of the freshmen had taken the TVE Joint College Entrance Examination (TVE, 2008). And if the large-scale national examinations are the exams with high reliability, it would be much convenient for college English teachers to group their students.

Purpose of the Study

Apparently extreme bimodal distributions of the English scores on the Entrance Exams every year have forewarned the highly sticky and complicated

problems. Is ability grouping an effective practice in freshmen English classrooms to narrow the gap between the two extremes and promote learning efficiency at the same time?

Since reliable English proficiency tests on the market are expensive and not affordable for most students, how can the TVE college entrance exam scores interpreted into English language proficiency benchmarks for higher institution teachers to find out their students' entry levels and to develop reasonable and achievable graduation thresholds?

The purposes of this study are to examine freshmen's English proficiency and four aspects of ability grouping in English after it has been practiced for years in the target university: (1) an analysis of student learning outcomes in different ability groups; (2) a comparison of differences in performance between listening and reading comprehension skills; (3) a comparison of differences in performance between speaking and writing skills; (4) a comparison of differences in performance between reading and writing skills; and (5) an analysis between the freshmen's TVE English scores and the corresponding proficiency rating.

Research Questions

Literature in finding college freshmen's entry levels in English language proficiency in higher education in Taiwan is very limited and almost no research can be found on relationship between the TVE exam scores and the corresponding proficiency ratings.

This study is to examine and analyze the scores of the freshmen's TVE English exam and the authorized English proficiency tests, the Global English Tests (GET), which were taken during the study period in the target university in the 2008-2009 academic year to find out answers for the following questions:

1. Are the TVE exam scores in English reliable to be translated into proficiency benchmarks useful for ability grouping?
2. How different are freshmen in English listening and reading comprehension

- skills in terms of learning proficiency?
3. How different are freshmen in English speaking and writing competence in terms of learning proficiency?
 4. How different are freshmen in English reading and writing competence in terms of learning proficiency?
 5. What are students' entry proficiency levels and the most proper English graduation benchmarks in a national technological university?

Literature Review

Ability Grouping

The practice of ability grouping in Taiwan can be traced back to the late 60's when the first practice was held in junior high schools (Liang, 2003). In the late 70's, the wave spread to the higher education and getting popular in the 90's. In 2005, the MOE decreed that there should be at least half of the students in technological colleges and universities achieved certificates of English proficiency in elementary level by year 2007 (MOE, 2005). In 2008 the Department of Higher Education published a plan called "Promoting Student Quality in the Senior Secondary Education and Higher Education," which emphasized education quality assurance and student quality control. It was mentioned in the plan that for all the college and university freshmen, they should be tested and grouped into the right levels for learning according to their entry-levels of English proficiency. Their learning process and learning outcomes also needed to be monitored and assessed to control student quality in full-scale (MOE Department of Higher Education, 2008). As for the national-wide Higher Education Evaluation in 2009, two of the performance indicators are first introduced to assess student learning achievement; one is the development of English graduation threshold, and the other, the percentage of students who attain the threshold levels by passing certain levels of English proficiency tests (Taiwan Assessment and Evaluation Association, 2009).

Ability grouping is also called tracking in the States or streaming in Great

Britain. The effects on ability grouping have already been debated for nearly a century and still not yet reached a general consensus. To proponents of ability grouping, it showed significant advantages to solve problems in academic diversity. Studies showed (Ansalone, 2003; Kulik & Kulik, 1992; Slavin, 1990), in homogeneously grouped classes, teachers were believed to have the advantage to better direct lessons; and students could build higher self-esteem because low achievers only competed with those of the same level while high-ability students got even higher achievement. Gamoran's (1990) study, on the contrary, revealed different facts on ability grouping. He pointed out the practice of ability grouping did not affect students' overall achievement; but alarmingly, it caused low achievers to learn less. It was also believed that ability grouping caused high achievers higher learning pressures because of teachers' higher expectations on them (Boaler, William, & Brown, 2000; Chen, Lin, & Feng, 2004). It was quite often the more experienced teachers were assigned to teach higher levels of classes, and the less experienced were sent to teach lower levels (Oakes, 1987) because people tended to believe low achievers did not learn much (Oakes, 1985; Clarke & Clarke, 2008).

Domestic studies showed relatively positive results toward ability grouping practices in colleges and universities in Taiwan. Han and Chang (2007) claimed, by adopting different levels of instructional materials in different levels of groups, the practice of ability grouping was proved effective. It not only promoted student achievement, but also enhanced students' learning interest in English. Tsai (2008) indicated ability grouping promoted student learning outcomes and enhanced interaction between teachers and students. Tsai recommended, in addition to the entrance exam scores, a placement test should be taken to better place freshmen into right levels of classes. Liu (2008) indicated in her research that most teachers and students involved support the practice of ability grouping; and students who benefited the most were those in the lowest level, who showed stronger support than students in the highest level. Liu's study showed students in

the high level got highest pressure due to greater competition among themselves. Lee and Su (2009) contended that students in the intermediate level benefited the most from the practice of ability grouping and gained significant progress.

English Language Proficiency Standards and Benchmarks

Due to autonomy of higher education institutions, there have been no nation-wide unified English curriculum guidelines for colleges and universities to follow or to improve their own curriculum. According to an island-wide survey conducted by a group of educators (Chang, Sue, Chou, & Chen, 2004), over 80% teachers believed the English proficiency of their freshmen students was at the level 3, all four skills were about the same level. However, over 70% of the respondents expected their college or university graduates being able to reach the level 8 which is about the high-intermediate level. On the report of this survey project, it was mentioned that there should be entry level benchmarks in addition to the graduation benchmarks in English language proficiency; and the proper graduation benchmark should be two levels higher than the real student entry levels. Moving from one level up to another, on practical experience, one needs to take at least two years of continuous English learning, spending at least three hours of study every week to achieve that goal. These educators who had conducted the survey also indicated that there were 70 percent of the higher institutions in Taiwan still had not adopted relevant proficiency benchmarks as the basis to design their own English curriculum.

As to the graduation threshold development, since freshmen quality each year is different in terms of English proficiency, every institution should keep track of students' English proficiency and their learning outcomes for at least several years before a graduation threshold is objectively developed, which must be attainable and reasonable for all students to achieve (Lee, 2003). Chang (2004) claimed that a reasonable graduation threshold is under two conditions: (1) the requirements of the threshold must be significantly related to students' qualifications; (2) these

requirements of the threshold can be reasonably acquired within four-years of college learning.

Method

The Target University of the Ability Grouping Practice

The target university is a national technological university which is situated in the central part of Taiwan. Ability grouping of non-English major freshmen for their English classes has been practiced since year 2003. There were 1511 non-English major freshmen enrolled in the 2008-2009 academic year in the target university. Among them, 1427 were the vocational high school graduates who took the TVE exams; and the other 86 students were graduates from other admission sources, who took other types of exams, and their English original scores were converted into the ones compatible to the TVE scores. Table 1 demonstrates great differences between the highest and the lowest scores among the population. These 1511 freshmen were roughly assigned into either Management or Engineering group; each of these two groups of students were separately ranked in descending order and students were arbitrarily divided into 4 levels with about 50 students or so in one class and as a total of 30 classes, which are shown below in Table 2. Tables 3 and 4 reveal the standard deviations for D-level students, which are much greater than the others.

Table 1

TVE scores of the Non-English Major Freshmen before Ability Grouping (N = 1511) 9/2008

N	Minimum	Maximum	Mean	STDEV
1511	.00	100.00	72.22	14.84

Table 2

English Placement Chart in Academic Year 2008-2009 (N = 1511)

Level	Classes in Management	TVE Exam Scores	Classes in Engineering	TVE Exam Scores
A	MA classes 1~3	84~100	EA classes 1~3	82~100
B	MB classes 1~3	78~84	EB classes 1~3	78~82
C	MC classes 1~3	68~78	EC classes 1~6	66~78
D	MD classes 1~3	0~68	ED classes 1~6	10~66

Table 3

Engineering Groups after Ability Grouping Based on TVE English Scores (N = 860) 09/2008

Level	N	Minimum	Maximum	Mean	STDEV
EA	159	82	100	87.87	4.73
EB	159	78	82	79.85	1.56
EC	270	66	78	71.45	3.26
ED	272	10	66	53.44	10.91

Table 4

Management Groups after Ability Grouping Based on TVE English Scores (N = 651) 09/2008

Level	N	Minimum	Maximum	Mean	STDEV
MA	165	84	100	89.40	4.37
MB	165	78	84	81.06	1.94
MC	166	68	78	73.60	2.66
MD	155	0	68	53.51	15.14

In addition to the TVE scores which were ranked from the highest to the lowest, the percentage or the number of students being placed into a specific level or a specific class was mainly based upon the class size. For example, there were 52 students in the management group who all scored 84 in the TVE exam; but among them, 21 students were placed in the A-level and 31 went to the B-level class.

Participants

The samples included 608 students out of the total 1511 freshmen population in the academic year 2008-2009. Among these 608 participants, 300 of them were A-level students in both Management and Engineering groups, and 308 were selected from B- and C- levels from both groups and the selection was based on students' English performance in the first semester.

Global English Tests (GET), the English Proficiency Tests

The global English Test (GET) is a two-stage examination designed to assess communicative language proficiency, which has been developed by the National Development Initiatives Institute (NDI). The first-stage examination tests reading and listening comprehension skills; and the second-stage, speaking and writing skills. The GET conforms to the Common European Framework for language assessment, and it is recognized as an authorized test of English proficiency for students and the general public in Taiwan. Domains of the GET cover general competence, communicative language competence, language activities, and social contexts of language use in English. The participants of the study took the GET in A2 level which is designed for the basic users; and the test outcomes were to be analyzed to evaluate student proficiency in English.

The GET defines six levels of proficiency, for better reference, a comparison in proficiency levels among GET, CEF and other test systems is listed in Table 5.

Table 5
A Comparison in Proficiency Level among GET, CEF, GEPT, TOEFL & IELTS

GET/CEF	GEPT	TOEFL CBT	IELTS
C1	Advanced	236 - 275	6.5 – 7
B2	High Intermediate	176 - 235	5.5 – 6
B1	Intermediate	126 - 175	4.5 – 5
A2	Elementary	96 - 125	3.5 – 4
A1			

(Source: www.ndi.org.tw)

Procedure

Before the school started in academic year 2008-2009, all non-English major freshmen had already been ability grouped into 4 different levels for their Freshmen English classes according to the TVE English scores. Freshmen English was a two-hour course every week, which is designed to train skills in reading and writing, but it focused more on reading skills. Textbooks had already been decided before the first day of the school. There were unified textbooks, English syllabus, and course schedule. The mid-term and final exams were unified mock proficiency tests of English, which counted for 30 percent of the total grades. There was also a separated required speaking and listening course for about 50 minutes every week designed for non-English major freshmen to take, and students were exactly ability grouped the same way as their Freshmen English classes with the same teachers. Oral practice was not practiced often because of limited time and large-size of classes.

There were three authorized GETs held in the campus in 2008-2009. Due to limited budget and high expenses in the authorized English proficiency tests, only level A students were required to take the first-stage tests. Among these 324 A level students, 12 were absent for the tests and another 12 did not take the 2008 TVE exams so these 24 students were not included to be examined in this study. As a total of 300 A-level participants took the first-stage GET on Jan 08, 2009, 242 of them passed. These A-level participants then were carefully evaluated by their English teachers according to their speaking and writing performance in class; 173 were picked out and allowed to take the second-stage tests in speaking and writing on May 14, 2009. Among these 173 students, 96 passed their second-stage GET test. As to levels B and C students, based on the grades and their English performance in the first semester of 2008-2009, 319 out-performed B- and C- level students were picked out and assigned to take another authorized first-stage GET on April 30, 2009; 11 were absent and 308 of them took the tests (see Table 6).

Table 6

Total Participants: 608 Freshmen

Student Level	No. of Participants	GET Proficiency Tests in English	Test Schedule	No. of Students Passed
A	300	(1) Reading (2) Listening	01/08/2009	242
A	173	(3) Speaking (4) Writing	05/14/2009	96
B & C	308	(1) Reading (2) Listening	04/30/2009	202

These three test results and the scores on TVE English exam were compared and analyzed to (1) assess the relationship between the TVE scores and the GET reading scores; (2) examine the relationship between participants' listening and reading competence, (3) examine the relationship between participants' speaking and writing competence, and (4) examine the relationship between participants' reading and writing competence; (5) extrapolate backward to estimate students' entry levels and forward to predict the most proper graduation thresholds for students in the target university. Computer software Excel and SPSS were used to do all the relevant statistical analysis.

Results

Relationship between TVE English scores and GET scores

To find out the relationship between the TVE English scores and the GET reading scores, both relevant scores of the participants' should be examined first. Two first-stage GETs were held and taken by students in level A in January and levels B and C in April (see Table 7). Since the TVE entrance examination in English was a pencil-and-paper exam, only reading skills were tested. Table 8 shows Engineering and Management groups of students' English performance on the TVE exam and the first-stage GET in both listening and reading skills.

Table 7

English Performance of 1st GET Participants on Reading & Listening Tests (N = 608)

Date	Test Level	Ability Grouping Level	No. of Participants	No. who passed	Percentage
2009/1/08	A2	A	300	242	80.67%
2009/4/30	A2	B & C	308	202	65.56%

Table 8

Mean Proportion of Scores on TVE and GET 1st Reading Tests (N = 608)

Grouping Class	TVE	GET Reading	GET Listening
EA (<i>n</i> = 145)	87.8690	78.8759	67.5241
EB (<i>n</i> = 106)	79.9623	72.7264	67.7353
EC (<i>n</i> = 29)	76.0000	69.4138	63.6897
MA (<i>n</i> = 155)	89.4645	77.9871	67.2000
MB (<i>n</i> = 113)	79.8058	73.9709	69.6241
MC (<i>n</i> = 60)	75.2667	67.6500	64.0167

1. Prediction of GET scores based on TVE scores

Based on freshmen's TVE exam scores, is it possible to predict student performance in terms of English competence in a proficiency test? A simple linear regression was calculated to predict 608 participants' reading scores on the GET. A significant regression equation was found ($F(1,607) = 57.888, p < .001$) with an R^2 of .087. The predicted reading scores on the proficiency tests are equal to $30.064 + .537$ points when the TVE scores are compared with. That means, for each TVE point earned, the average reading scores on GET increased 0.537. Another significant regression equation was also found on the GET listening scores as ($F(1,606) = 8.190, p < .005$), with an R^2 of .013. The predicted listening comprehension scores on GET are equal to $51.572 + .184$ points when the TVE scores are measured; that is, for each TVE point earned, the average GET listening scores increased 0.184.

2. Relationship among TVE, GET listening and reading scores

608 participants' scores on both TVE exam and the first-stage GETs were examined to find out how well that ability grouping was implemented in the target university. SPSS 15.0 was used to compute the Pearson correlation coefficient, or simply the Pearson r , to determine the strength of the linear relationship between these scores. Preliminary analyses indicated that all dependent variables, the English scores on the TVE, the GET listening and reading tests, had normal distributions (skewness < 1). The one-way ANOVA indicated that there was no significant difference in the implementation of ability grouping between Management and Engineering groups in the target university; so this was excluded from subsequent analysis. However, TVE scores, GET Listening, and reading scores were significantly correlated with each other at all ability group levels ($p < .01$ and $.05$).

Ability grouping effects and relationship between receptive skills

The Pearson correlation coefficient was adopted to examine the relationship of English performance between GET listening and reading scores. A positive correlation was found ($r(606) = .489, p < .001$), indicating a significant linear relationship between listening and reading competence. That means, if a student scored higher on his GET listening test, he also scored higher on the reading test; as shown above in Table 8, participants' listening scores are comparatively lower than their reading scores at all levels and these differences are significant in classes in all 3 levels: EA (paired $t(144) = 14.310, p < .001$), EB (paired $t(105) = 5.036, p < .001$), EC (paired $t(28) = 3.788, p < .001$), MA (paired $t(154) = 14.256, p < .001$), and MB (paired $t(112) = 5.689, p < .001$) and MC (paired $t(59) = 2.19, p < .05$).

The effect of ability grouping among 3 different levels was assessed in 6 classes at levels A, B, and C, which are: EA, EB, EC, MA, MB, and MC. A paired-sample t test was calculated to compare the learning outcomes: (1) the mean

score between the TVE and GET listening test and (2) the mean score between the TVE and the GET reading test. Since the TVE (pre-test) and GET (post-test) exams were not measured with the same scale, the scores were converted to the z-scores before conducting the *t* test (Cronk, 2006; Salkind, 2006). No significant difference was found on student learning outcomes from the TVE exam to the GET receptive tests.

A Multivariate ANOVA was conducted to compare these 6 classes of 3 levels in GET reading and listening scores respectively. A significant difference was found among these 6 classes ($F(5, 602) = 14.846, p < .001$). Tukey's *HSD* was used to determine the nature of the differences among these 6 classes of 3 levels. In terms of the GET reading scores, this analysis revealed that students who were placed in the level A classes scored higher ($m = 78.88$ and $77.99, sd = 9.69$ and 11.64) than students who were placed either in the level B classes ($m = 72.72$ and $74.03, sd = 9.51$ and 10.36) or in the level C classes ($m = 69.41$ and $67.65, sd = 10.54$ and 12.16). However, there was no significant difference found on GET listening scores ($F(5, 602) = 2.204, p > .05$) among these classes.

Ability grouping effects and relationship between productive skills

A Pearson correlation coefficient was computed to examine the relationship between the scores of the participants' GET speaking and writing tests. A weak significant correlation was found ($r(171) = .101, p > .05$). The scores of the GET speaking test are not related to the scores of the GET writing test. However, it can be seen from Table 9 that, GET speaking scores are lower than GET writing scores at all classes. These differences are significant in some classes, paired $t(27) = 4.036, p < .001$ in EA1 class, paired $t(30) = 5.183, p < .001$ in EA2 class, and paired $t(31) = 3.364, p < .05$ in MA2 class.

Table 9

Mean Proportion on GET 2nd Stage Speaking and Writing tests (N = 173)

A-Level Classes	GET Speaking/SD	GET Writing/SD
MA1 (<i>n</i> = 28)	71.36 (14.36)	73.14 (5.10)
MA2 (<i>n</i> = 31)	67.52 (8.06)	72.29 (4.14)
MA3 (<i>n</i> = 32)	66.31 (15.39)	69.63 (3.23)
EA1 (<i>n</i> = 24)	62.38 (15.88)	75.75 (4.90)
EA2 (<i>n</i> = 32)	67.88 (9.14)	75.84 (5.63)
EA3 (<i>n</i> = 26)	70.88 (8.12)	73.19 (4.89)
Total (<i>n</i> = 173)	67.77 (12.36)	73.20 (5.00)

173 A-level students' learning outcomes in terms of the GET speaking and writing scores were also assessed. From the TVE exam (the pre-test) to the GET productive tests (the post-tests), no significant difference was found on learning outcomes as the result of ability grouping practice.

A Multivariate ANOVA was conducted to compare the following 6 participants' classes at A-level in GET speaking and writing scores. These classes are: MA1, MA2, MA3, EA1, EA2, and EA3. A significant difference was found on GET writing scores among the 6 A-level classes ($F(5, 167) = 7.842, p < .001$). Tukey's *HSD* was used to determine the nature of the differences among these 6 classes in level A. This analysis revealed that students who were placed in the MA3 class scored significantly lower ($m = 69.63, sd = 3.23$) than students who were placed in the classes of EA1, EA2, EA3 and MA1 ($m = 75.75, 75.84, 73.19$, and $73.14, sd = 4.90, 5.63, 4.09$, and 5.10). Students who were placed in the MA2 class scored significantly lower ($m = 72.29, sd = 4.14$) than students who were placed in the EA2 class ($m = 75.84, sd = 5.63$). However, no significant difference was found on GET speaking scores ($F(5, 167) = 1.847, p > .05$).

Ability grouping effects and relationship between literacy skills

A Pearson correlation coefficient was calculated to examine the relationship between participants' scores on the GET reading and writing tests. A positive correlation was found ($r(171) = .211, p < .005$), indicating a significant linear

relationship between these two variables. Students who scored higher on the GET reading tests tend to score higher on their GET writing tests. It can be seen from Table 10 which shows that there are significant differences in all 6 classes, paired t (27) = 10.246, $p < .001$ in MA1, paired t (30) = 9.483, $p < .001$ in MA2, paired t (31) = 8.572, $p < .001$ in MA3, paired t (23) = 2.544, $p < .05$ in EA1, paired t (31) = 2.640, $p < .05$ in EA2, and paired t (25) = 7.294, $p < .001$ in EA3.

Only a one-way ANOVA was conducted to compare 6 A-level participants' classes on their GET reading scores because the GET writing scores were already illustrated in the previous section. A significant difference in GET reading performance was found among these classes (F (5, 167) = 3.596, $p < .005$). Tukey's *HSD* was used to determine the nature of the differences among the classes. This analysis revealed that students who were placed in the MA1 class scored significantly higher ($m = 83.89$, $sd = 6.81$) than students who were placed in the MA3 class ($m = 78.84$, $sd = 5.63$); and there was no significant difference among the rest of the classes on the GET reading scores.

Table 10

Mean Proportion on GET Reading and Writing tests (N = 173)

	GET Reading	GET Writing
MA1 ($n = 28$)	83.89 (6.82)	73.14 (5.10)
MA2 ($n = 31$)	82.06 (5.46)	72.29 (4.14)
MA3 ($n = 32$)	78.84 (5.62)	69.63 (3.23)
EA1 ($n = 24$)	80.08 (6.92)	75.75 (2.90)
EA2 ($n = 32$)	79.31 (6.27)	75.84 (5.63)
EA3 ($n = 26$)	83.54 (6.19)	73.19 (4.08)
Total ($n = 173$)	81.20 (6.18)	73.21 (5.00)

Participants' English Language Entry Levels

Table 11 shows the distribution of the total population's TVE English scores. The average is 72.225. Table 12 illustrates about the English language proficiency of the A-level students in both receptive and productive skills. After 2 semesters of the ability grouping practice, 96 students, or 32% of the A-level participants achieved the elementary level of English language proficiency in all 4

skills. Table 13 demonstrates the English language receptive proficiency of the participants in levels B and C. It is worth noticing, although the lowest TVE score for C-level students to take the GET is 72, the students who proved to pass the first-stage tests all gained 74 or higher scores on the TVE. That means students who gained the TVE scores 72 or lower have very slim chance to pass the first-stage GET.

To assess whether or not there is an impact of ability grouping on the 608 participants' English performance after a semester or two of the instruction, a paired-sample *t* test was adopted to calculate and compare the mean TVE scores and both of the listening and reading scores on GET. Unfortunately, no significant difference was found from the TVE exam to the GET tests. This means that, despite the practice of ability grouping, students remain about the same entry proficiency levels after spending several months of English learning in the campus.

Table 11

Frequency & Percentage of TVE Scores (N = 1511) 09/2008

TVE Scores	Frequency	Valid Percent	Cumulative Percent
.00	1	.1	.1
10.00	1	.1	.3
20.00	4	.3	.5
30.00	8	.5	2.1
40.00	12	.8	4.7
50.00	14	.9	8.4
60.00	45	3.0	18.6
70.00	72	4.8	38.6
72.00	67	4.4	43.1
74.00	109	7.2	50.3
80.00	116	7.7	70.2
84.00	94	6.2	83.3
90.00	41	2.7	95.0
98.00	1	.1	98.4
100.00	24	1.6	100.0
Total	1511	100.0	-

(Note: The average is 72.225.)

Table 12

A-Level Student Performance on GET 1st Stage Reading & Listening Tests
(N = 300) 01/2009

Class	N	Min. Score on TVE	1 st Stage Passed	Proportion	2 nd Stage passed	Proportion
MA1	51	90	40	78%	22	43%
MA2	54	86	43	80%	17	31%
MA3	50	84	42	84%	12	24%
EA1	48	88	42	88%	8	17%
EA2	47	86	40	85%	17	36%
EA3	50	82	35	70%	20	40%
Total	300	-	242	81%	96	32%

Table 13

B- & C- Levels Student Performance on GET 1st Stage Reading & Listening Tests
(N = 308) 04/2009

Level	N	Min. Scores on TVE	Min TVE Score Passed	N passed	Proportion
MB	113	78	78	83	73.45
MC	60	72	74	26	43.33
EB	106	78	78	76	71.70
EC	29	74	74	17	58.62
Total	308	-	-	202	65.58

Discussions

TVE English Exam versus GET Tests (608 participants in Levels A, B, and C)

According to regression analysis in the comprehensive empirical analysis, the scores on the GET reading and listening tests are exactly consistent with the regression equations found, indicating that freshmen scores on the TVE exam in English can be interpreted into corresponding levels of English proficiency. By using the Pearson correlation coefficient acquired through statistic analysis, the authors find a strong correlation among scores in TVE English exam, the GET reading and listening tests. This further implies that TVE English scores are

proven to be reliable and useful to be adopted as a basis for proper ability grouping students into different levels of classes. These findings are against the claims of the “2008 Investigation Report on the TVE Exams” made by the National Teacher’s Association (2008) saying that the English exam is not discriminating enough to tell student levels in terms of English proficiency.

The findings demonstrate that expensive campus-wide placement tests for colleges and universities freshmen prior to the Freshmen English classes are not necessary if students are carefully grouped according to their TVE scores in English. Whether or not a student should be placed into either the Management or the Engineering groups is not one of the concerns.

Reading Skills versus Listening Skills (608 participants in Levels A, B, and C)

Based on the Pearson correlation coefficient, a significant relationship is found between reading and listening comprehension. The findings indicate students’ overall comprehension in reading is lower than listening at all levels; and when a student gains higher score in reading, he/she gets comparatively higher score in listening, too. It can also be seen in Table 8 that comprehension difference between reading and listening of A-level students are much greater than that of the students in both B and C levels. That means the difference between reading and listening comprehension increases with increased ability levels, and this is not in agreement with the previous research findings which indicate exactly the opposite (Diakidoy, Stylianou, Karefillidou, & Papageorgiou, 2005). A few possible culprits for this result of the current study are the traditional grammar-translation type of teaching method, the pencil-and-paper entrance exams always focusing more on reading skills and lack of practice in listening comprehension, etc.

As to the effect of ability grouping in terms of students’ reading comprehension, a significant difference is found among the classes of these 608 participants who were placed in 3 different levels, A, B and C. The findings show

students in the highest level achieve the most in reading comprehension; and students in higher levels achieve more than those in lower levels, the lowest students achieve the least. These findings are consistent with previous findings (Boaler, William, & Brown, 2000; Chen, Lin, & Feng, 2004). With regard to the learning outcomes of students' listening comprehension, it shows no significant difference among different levels with the practice of ability grouping. Once again, this can be attributed to reasons like pencil-and-paper exams, instructions focusing more on reading skills, lack of proper facilities and class hours for training in listening, etc.

Writing Skills versus Speaking skills (173 A-level participants)

A previous research finding draws the conclusion that there is a positive relationship between students' speaking and writing proficiency, indicating that a student who writes better speaks better because of better mastery of words and sentences (Zhu, 2007). In Zhu's research, the subjects were a group of ESL students studying in the USA with plenty of exposure in English environment. In the current study, however, based on the analysis using Pearson correlation coefficient, the finding is different; it indicates no significant relationship between student writing and speaking comprehension. This is because students in Taiwan are usually lack of exposure to English writing and speaking experiences. Besides, big-sized classes with more than 50 students in each class, test-oriented instructions, short of or no class hours in speaking and writing are all reasons to be blamed for.

As it can be seen in Table 9, the overall student scores in speaking test are lower than scores on the writing test. The difference is especially significant in classes of EA1 and EA2. As to the effect of ability grouping in terms of writing proficiency, a significant difference is found among all 6 A-level classes. A-level students in the Engineering group perform better than those in the Management group; and yet no significant difference can be found among these classes in speaking proficiency. As to the reasons why students in the Engineering group

write better than those in the Management group, a further follow-up study needs to be done.

Reading Skills versus Writing Skills (173 A-level participants)

It is a truth commonly acknowledged that reading does not only help enhance language learners' writing ability, but also provide them with opportunities to get familiar with knowledge in linguistic and cultural background which are necessary for good writing. An analysis using Pearson correlation coefficient indicates that there is a positive correlation between students' reading and writing comprehension; students with higher reading proficiency tend to acquire higher writing competence.

When it comes to the effect of ability grouping, among these 6 classes of 173 A-level participants who took listening and reading tests on the first-stage GET in mid January, and speaking and writing tests on the second-stage GET four months later, the authors find a significant difference in the reading performance, especially on the class MA1 if compared with MA3. There are no significant differences among classes in the Engineering group.

Freshmen's English entry level versus graduation thresholds

The Canadian Language Benchmarks (CLB) describes 12 benchmarks as 12 levels of English language proficiency: (1) the basic level is described as Stage 1, levels 1 to 4; (2) the intermediate level as Stage 2, levels 5 to 8; and (3) the advanced level as Stage 3, levels 9 to 12. If compared with the CEF proficiency scales, CLB 5 and CLB 6 are respectively equivalent to CEF B1 (Intermediate) and B2 (High Intermediate), CLB 3 and CLB 4 to CEF A2 (Elementary) first-stage and second-stage respectively; CLB 1 and 2 are described as no proficiency at all (CLB, 2000).

In the research on *Developing English Language Proficiency Benchmarks for Taiwanese Students in Higher Education*, the CLB proficiency scales were

advocated and the following six items were recommended in the research report (Chang et al., 2004):

1. There should be both entry level benchmarks and graduation benchmarks developed for students in colleges and universities.
2. There should be 3 kinds of graduation benchmarks developed as levels 4, 6, and 8.
3. Graduation benchmarks should be 2 levels higher than the entry levels; if the entry level benchmark is 4, the graduation benchmark should be developed at level 6.
4. Entry level benchmarks and graduation benchmarks should be developed by each institution according to its specific professions or features of each department, school, or educational group.
5. Graduation benchmarks should be adjustable to meet students' special needs; for instance, the graduation benchmark developed at level 4 in listening, speaking and writing proficiency, but at level 6 in reading comprehension in order to enable certain groups of the college graduates to have enough reading skills to well perform the office work in future job environments.
6. Freshmen's entry levels should be carefully examined by each institution. If there are first-year freshmen below the minimum proficiency requirements, remedial instruction should be given.

The total freshmen population in the target university in the 2008-2009 academic year was 1511, while the average scores on the TVE exam in English is 72.225 with the standard deviation 14.84. According to the Testing Center for Technological & Vocational Education, the total TVE test-takers in 2008 were 169,974 with the average 45.6978 in English and the standard deviation 21.377. In comparison with both the means and standard deviations between the total TVE test-takers and the freshmen population in the target university, it can be found that freshmen in a national technological university were in better quality in terms of

the English language proficiency; and yet there were 49.7% of the first-year students in the target university not equipped with the kind of English competence equivalent to level 3 (see Table 11). According to the findings of this study, students who passed the first-stage GET at least scored 74 or higher in their 2008 TVE exam in English. Therefore, the score, 74, can be considered as a reference indicator or an important entry-level passing grade to estimate students' graduation levels on the relevant proficiency scales. If students who scored 74 or higher in the TVE, they are equipped with the ability of level 3 or the higher levels; their graduation benchmarks should be developed at level 5 in CLB or B1 in CEF. For students who scored lower than 74 but higher than 60, they were very close to level 3 and should be offered intensive training courses in the first year focusing on both reading and listening to help them keep up with those who scored higher than 74. As to the students who scored from 0 to 59 in their TVE in English, they were about at the levels 1 or 2 and should be given extensive remedial instruction according to their weaknesses in specific skills; and the reasonable and reachable graduation benchmarks for these students should be developed at about level 3 because for each further level achieved, one needs to spend at least two years of continuous English learning, at least three hours of studying every week in order to achieve the proficiency demanded at that level (Chang et al., 2004). If there are unified campus-wide graduation thresholds for all non-English majors in each institution as most colleges and universities do at present, for those who have already achieved the threshold before their coming to the college campus, they may not have the strong motivation to encourage themselves to move forward to higher English proficiency levels; and for those low-achievers, on the other hand, their anxiety and pressure may influence their learning interest because of the unreasonable and unreachable thresholds long developed before their coming to the campuses. Of course, the score, 74, means a lot to 2008-2009 students and may not be applied to students in other academic years without further investigation because different TVE exam questions may yield different results.

Instead of the recommended graduation thresholds for colleges and universities suggested by a group of well-known educators and professionals mentioned above as levels 4, 6, and 8 in CLB, according to the findings of this current study, more reasonable and attainable graduation thresholds are proposed as levels 3, 5 and 6 for non-English majors in technological colleges and universities, which are respectively equivalent to A2 the elementary first-stage, B1 the intermediate second-stage and B2 the high intermediate first-stage in CEF proficiency scale.

Conclusion

This study is based on theoretical arguments about the effects of ability grouping practice and the relevant English language proficiency benchmarks. Empirical analysis suggests that ability grouping practice in the target university does not illustrate significant difference. This corresponds to the statement mentioned above that significant improvement in English takes at least two years of continuous studying. Ability grouping should not be considered as a one-size-fit-all solution, other factors like teaching methods, instructional materials, class-size, learning motivation, testing, etc. are all crucial to a success formula. Other findings of this study provide English teachers with evidence that students' listening comprehension is behind their reading comprehension, and better reading comprehension skills enhance writing skills. These results allow teachers to pay more attention to students' listening and writing skills. The findings also indicate the TVE English scores are reliable and convenient enough to be adopted as the basis of ability grouping, providing technological institutions with an indicator to examine student entry levels and develop attainable and practical graduation thresholds which can be applied to all non-English majors.

Before starting ability grouping, first-year freshmen should always be examined first to find out their entry levels because students are different every year in terms of English language proficiency. As to the decision made about

ranking students into certain levels, it should be carefully considered and evaluated and not to misplace students into inappropriate classes of levels. Simple assignment of an arbitrary percentage of students on a ranking list into certain ability levels without any supporting empirical data should be avoided. Besides, freshmen student entry levels should be adopted as a basis; dividing freshmen into at least three groups of levels: ability 0 to 2 level (in CLB) students, level 3 students, and level 4 students.

In current practice of the college evaluation, the number of certificates in the English language proficiency tests that an institution can get has become a very important indicator of education performance. High-achieving students are usually assigned with better teachers, get more resources, and use better facilities in order to get the best learning outcomes in terms of the English proficiency certificates. Lee (2003) claimed the English education in Taiwan is not designed for the low-achievers because there have not been any effective measures taken to solve the problem of bi-modal distribution in English and thus make the problem worse. Actually, English education in higher education should be functional and has its own specific purposes (Chang, 2003). If English teachers have always spent a great deal of time and energy designing and offering remedial instruction to those low-achievers who were from the high schools, what would be the specific purposes in the English education for colleges and universities then?

Ability grouping can never be an educational panacea to solve problems in English education in Taiwan. Despite the national plans and measures of all kinds in English education after all these years in Taiwan, student proficiency doesn't seem to improve much. It is frustrating to notice that there were 72.68% or 123,538 test-takers in the 2008 TVE English Exam who scored lower than 59, suggesting that about two thirds of the vocational high school graduates needed to get certain amount of remedial instruction if they all get into college campuses. If this problem does not improve from the very beginning stage of the education system, the English classrooms in the elementary and junior high schools, the

current English education dilemma will probably persist.

References

- 中華綜合發展研究院 (2008) , Global English Test. http://www.get.org.tw/get/e_GET/index.htm
- 全國教師會新聞稿 (2008)。九十七學年度四技二專統一入學測驗試題檢視報告。中華民國全國教師會電子報 970523(星期五)。
- 李文瑞(2003)。全國各大專校院普設英語畢業門檻標準之省思。英語教育電子月刊, EJEE。第三期。 <http://ejee.ncu.edu.tw/>
- 李家同(2003)。為什麼英語會有雙峰現象? 英語教育電子月刊, EJEE。第三期。
- 李金玲, 蘇綉惠 (2009)。英語分級教學成效分析—以中區某科技大學為例。新竹教育大學人文社會學報, 第二卷第一期, 頁234-253。
- 技專校院入學測驗中心(2008)。 http://www.tcte.edu.tw/page_new.php
- 社團法人台灣評鑑協會(2009)。「98 年度科技大學評鑑表冊」。 <http://utce.twaea.org.tw/about.htm>
- 教育部技專校院學生英語能力檢測計畫, 2004。「九十二學年度英語能力檢測成績統計報告」。 <http://www.lttc.ntu.edu.tw/2003ANNUALREPORT.pdf>
- 教育部祕書室(2009)。「97年度施政績效報告」。 http://www.edu.tw/secretary/content.aspx?site_content_sn=19757
- 教育部祕書室(2008)。第二期獎勵大學教學卓越計畫, 附件 2: 95-97 年度績效報告, 頁 45。 <http://www.edu.tw/secretary/>
- 教育部高教司(2008)。「提升高級中等以上學校與生素質計畫」。 <http://www.edu.tw/high/>
- 教育部(2005)。94 至 97 年度中程施政計畫。國立教育資料館, 頁 6-65。 http://www.edu.tw/content.aspx?site_content_sn=19439
- 張武昌(2003)。中華民國英語文教育政策目標及策略。英語教育電子月刊, EJEE, 第二期。 <http://ejee.ncu.edu.tw/prevlist.asp>
- 張顯達(2004)。台灣大學的英語能力檢定。英語教育電子月刊, EJEE, 第八期。 <http://ejee.ncu.edu.tw/>
- 張顯達, 蘇以文, 周碩貴, 陳美華(2004)。建立我國大學生英語能力指標之研究, 2004 教育部研究訂計畫。 <http://ejee.ncu.edu.tw/>
- 黃雪雲、黃靜雲、馮敏毅 (2003)。勤益學生在「全民英文能力檢定」中的定位與評估。勤益學報, 第 21 卷第 2 期, 頁 25-34。
- 蔡金治 (2008)。國立高雄海洋科大四技大一英文分級教學實施成效及檢討。海洋學報, 第22期, 頁169-184。
- 韓瓊芳、張燦明(2007)。英文教材分級對學習成效之研究：以修平技術學院為例。 Paper presented on 2007 Conference on Innovation of Industry Management. <http://www.hit.edu.tw>
- Ansalone, G. (2003). Poverty, tracking, and the social construction of failure: International perspectives on tracking. *Journal of Children & Poverty*, 9.

- Boaler, Jo, William, Dylan, & Brown, Margaret. (2000). Students' experiences of ability grouping – disaffection, polarization and the construction of failure. *British Education Journal*. 26(5), 631-648.
- Chen, Mei-Hua, Lin, Yi-Chiu & Feng, Hsiu-Chen (2004). Students' and Instructors' Perceptions of Ability Grouping in English Listening Learning. *Journal of Chang Jung Christian University* , 8(1), 107-123.
- Clarke, Doug & Clarke, Barbara (2008). Is Time Up for Ability Grouping? *Curriculum Leadership*. 6(5).
- Council of Europe (2001). *Common European Framework of Reference for Languages CEFR*, from: http://www.coe.int/t/dg4/linguistic/CADRE_EN.asp
- Cronk, B. C. (2006). How to use SPSS: A step-by-step guide to analysis and interpretation. CA: Pyrczak Publishing.
- Diakidoy, I. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology*, 26, 55-80.
- Gamoran, A. (1990). *Classroom instruction and the effects of ability grouping: A structural model*. Paper presented at the annual meeting of the American Educational Research Association, Boston: MA.
- Kulik, James A. & Chen-Lin C. Kulik (1992). "Meta-analytic findings on grouping programs". *Gifted Children Quarterly*. 36(2), 73-77.
- Liang, T. L., (2003). Cooperative learning: An alternative to ability grouping. *Journal of Paisa Humanities and Social Sciences*. 2, 141-175.
- Liu, Hui-Ju (2008). An Analysis of the Effects of Ability Grouping on Student Learning in University-Wide English Classes. *Feng Chia Journal of Humanities and Social Sciences*. 16. 217-249.
- Oakes, Jeannie (1985). *Distribution of Knowledge. Keeping Track: How Schools Structure Inequality*. Yale University Press.
- Oakes, Jeannie (1987). *Tracking in Secondary Schools: A Contextual Perspective*. Santa Monica, CA: Rand Corp.
- Pare, Valerie. (2004). Exploring the Conflicts Involved with Ability Grouping. *The National Research Center on the Gifted and Talented*. Summer 2004 Newsletter. July, 2008, retrieved from: <http://www.gifted.uconn.edu/nrcgt/newsletter/summer04/summer04.pdf>
- Salkind, N. J. (2006). *Exploring research*. NJ: Pearson Education Ltd.
- Slavin, Robert E. (1990). "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis". *Review of Educational Research*. 60(3), 471-499.
- Testing Center for Technological & Vocational Education (2008). *Year Report*. <http://www.tcte.edu.tw/en/index.html>

The Canadian Language Benchmarks (2000). *The Centre for Canadian Language Benchmarks*. Retrieved

from: http://www.language.ca/display_page.asp?page_id=1

Zhu, Xin-hua (2007). What do we know about the relationship between speaking and writing in college-level ESL students? *US-China Foreign Language*. 5(3), (Serial No. 42), 31-40.