Introduction

The required use of English textbooks is a well established practice in Taiwan's colleges and universities. In a survey of 65 college content area instructors at a medical university in southern Taiwan, Cheng and Hung (2002) found that 20% of the teachers used English texts entirely in the courses they taught, and 75.4% used English texts in at least 50% of the courses they taught; while 4.6% of these teachers reported that they used Mandarin Chinese texts only. By the end of 2008, the top-ranked Taiwanese universities, such as National Taiwan University and National Tsing Hua University, had announced that the courses instructed entirely in English had reached 10% (Hu, Chen & Liu, 2008). Yet, the achievement is still unknown and controversial. The students have responded controversially as well; some favor the courses whereas some expressed frustration because they could not understand the English used in these courses (Chen, Hu & Liu, 2008).

Research has also shown that 87 to 88% of the doctoral and graduate students at public universities reported experiencing great difficulties in comprehending the vocabulary words and the English syntax in their scholarly works (Yiau, 1993). Taiwanese students shun English learning materials whenever Mandarin Chinese versions are available (Cheng, 1993). When students have no choice but to use an English textbook, much of their reading time is taken up with looking up words in the dictionary and recording their Chinese equivalents. That the majority of students feel compelled to handle textbooks in English in such an inefficient manner may be due to the reader's own limited general and specialized vocabulary. Much research remains to be done on the roots of this particular reading problem among Taiwanese college students.

To attempt a first step at bringing out the causative factors with regard to their general vocabulary ability, we administered the vocabulary subtest of the Gates-MacGinitie Reading Tests (GMRT), Level 7/9 Form S, the Fourth Edition, to a group of Taiwanese high-achieving college freshmen. Explored as well were

gender differences among EFL students in term of vocabulary ability as previous research focused mainly on native speakers of English and was conducted within the field of cognitive psychology. In this paper we firstly present the review of the relevant research, and then the study, which was carried out on a sample group of 209 (134 = male; 75 = female) college freshmen. It is hoped that the results from this groundwork study will lead the way to more large-group studies, from which we may determine if and whether compensatory practices are indicated.

The Origin of Scientific Methods in Tackling Reading Problems

In the United States, existing national surveys have documented that the vocabulary of about 50 million adults only reaches the 4th and 5th grade levels (Sweet, 1997). The U.S. Department of Education's National Center for Education Statistics (2003) reports that 70% of 4th graders, 30% of 8th graders, and 64% of 12th graders read below their expected grade levels. The evidence that many students were either failing to learn to read or were performing far below expected levels has drawn the attention of educators since the turn of the 20th century (Lipson & Wixson, 1991). For example, The Reading Tent Association set up by Alfred Fitzpatrick between 1899 and 1919, Kentucky Moonlight Schools founded and led by Cora Stewart from 1911 to 1922, and Opportunity Schools established in 1920 were some of the earliest efforts to eradicate illiteracy (Kozol, 1986; Newman & Beverstock, 1990). In the early 20th century, psychologists and reading specialists initiated a two-track movement using scientific methods to explore the problems: Thorndike developed the first norm-referenced group test of reading ability in 1914, and Williams S. Gray published an oral reading test in 1915, which led to the diagnostic movement in reading and to an emphasis on remediation (Lipson & Wixson, 1991).

Nearly a century later, both tracks have led to the prolific development and use of norm- or criterion-referenced group tests and informal reading tests by schools, educators, and researchers. An analysis of the skills and areas measured in 74

formal and informal tests published before 2002 reveals that 67.57% of the tests measure one or more than one of the skills and areas in vocabulary comprehension, word analysis or word recognition, word articulation, spelling, syllable principles, and phonics; 62.16% measure passage comprehension; 5.41% measure reading speed; and 27.03% measure listening comprehension level. These results show that reading educators are vitally concerned with the roles both vocabulary and word recognition play in reading comprehension, and that they generally perceive limited knowledge of vocabulary as one of the major sources of reading difficulties.

Research into Vocabulary Ability

The correlations between vocabulary and comprehension are well documented. For example, Chall (1987) found that word meaning scores are highly correlated with reading comprehension scores. In addition, in a study using a population sample of 12,152 students from grades 7 to 9, MacGinitie, MacGinitie, Maria and Dreyer (2002) found that the correlations between vocabulary and reading comprehension range from 0.74 to 0.77. In the EFL setting, vocabulary size is also found highly correlated with reading comprehension (Qian, 1999) and writing ability (Astika, 1993; Laufer & Nation, 1995). Chall (1987) suggested that a reading vocabulary test may be substituted for a paragraph-meaning test.

In vocabulary studies, West (1953) developed a General Service List (GSL) which contains 2,000 headwords chosen mainly on the basis of frequency. Coxhead (2000) devised an Academic Word List (AWL) of 570 headwords from the Academic Corpus of 3.5 million words which consisted of textbooks from Art, Commerce, Law and Science. Nation (2004) reduced West's 2,000 headwords to 1,986 after conducting a validation study to ascertain the validity of the GSL and the AWL. When tested on the whole Academic Corpus, the two lists comprise 86.1% (Nation, 2004).

Other researchers have focused on the reading threshold of the vocabulary size or the size of vocabularies in the four language skills: reading, listening, speaking,

and writing (Hendricks, 1988; Lorge & Chall, 1963; Moe, 1974). For instance, in a native English-speaking context, it is estimated that most children entering the first grade already possess around 5,000 to 6,000 speaking and listening vocabularies (Lorge & Chall, 1963; Moe, 1974). Nagy and Herman (1987) summarized the results of their review of studies on vocabulary size and reported that at Grade 3 the vocabulary estimates vary from 2,000 to 25,500 words and at Grade 12 from 7,800 to 47,000 words. In an EFL context, Nurweni and Read (1999) conducted a study on how many words freshmen of an Indonesian university knew. The results showed that the subjects knew only 240 out of the 800 words in the University Word List's (UWL) developed by Xue and Nation (1984).

The diagnosticians are mainly interested in the miscues associated with word recognition and comprehension, and the approaches for remediation (Burns & Roe, 1989; Ekwall & Shanker, 1988; Karlson & Gardner, 1986). Classroom teachers, nevertheless, are more interested in using word tests to provide a quick estimation of a student's approximate reading level so they can assign texts at the student's approximate level when there are no other reading tests available (Burns & Roe, 1989; Lipson & Wixson, 1991). To enable teachers to locate students' reading levels, Burns and Roe (1989) developed graded word lists from preprimer to twelfth grade. Slossen (1988) developed a criterion-referenced word list. Cheng (2007) modified and administered the Slossen Oral Reading Test (SORT) (Slossen, 1988) to 145 average and below-average EFL college freshmen to identify their vocabulary grade equivalents (GE) in order to build up empirical data for understanding college students' reading difficulties in Taiwan. Descriptive statistics identify the range of vocabulary GE as being from Grade 1 to Grade 6. Both Grade 1 and Grade 6 are at the extremities of the bell curve as only 0.7 % of the students were reading at a Grade 1 level and 4.8% at a Grade 6 level. The other grade levels were distributed as follows: Grade 2, 5.5%; Grade 3, 24.15%, Grade 4, 38.6%, and Grade 5, 26.2%. The study documented that this test is very time-consuming if a great number of students are to be assessed. The researcher also suggested that before a

decisive conclusion about the vocabulary GE of college students in Taiwan can be reached, more studies with readers of different abilities need to be conducted and the widely-used norm-referenced group tests be administered.

As suggested in the study by Cheng (2007) and for building up more comprehensive data, the current study employed a widely-used norm-referenced group survey test to locate college readers' English vocabulary GE in Taiwan. Therefore, the first purpose of this study is to answer the following question: What are the distributions of English vocabulary grade equivalents of a group of high-achieving college freshmen in Taiwan?

Research into Gender Differences in Vocabulary Size

Much of the research into gender differences in cognitive abilities has been devoted to the examination of gender differences in verbal ability. The results, however, are mixed (Halpern, 2000; Hyde & McKinley, 1997). Huttenlocher, Haight, Bryk, Seltzer and Lyons (1991) found that on average, there is a 13-word difference in vocabulary size between girls and boys at 16 months of age, which grows to a 51-word difference at 20 months and a 115-word difference at 24 months. Hyde and Linn (1988), in a meta-analysis using 165 studies that had reported data on gender differences in verbal ability, found that the results indicated small differences in separate meta-analyses of vocabulary. The other meta-analysis conducted by Hedges and Nowell (1995) also found that in the case of vocabulary, the values of *d* ranged between -0.06 and +0.25 with a mean of +0.06. These estimates cluster around zero and are quite small. The result again led them to conclude that the differences are essentially nonexistent.

Nevertheless, Halpern (2000) argued that on average, females have better verbal abilities than males. He continued to argue that verbal abilities apply to all components of language usage: word fluency, grammar, spelling, reading, writing, verbal analogies, vocabulary and oral comprehension. The size and the reliability of the gender difference depend on which of these aspects of language usage are being

assessed. The mixed results of research come from the failure to distinguish between language tasks. Based on this concept, he reviewed the studies conducted before the turn of the new millennium and concluded that males comprise an unbalanced share of the extremely low-ability end of the distributions of verbal ability. Solving verbal analogies is the only type of verbal ability that shows a male advantage. By contrast, females excel at anagrams, general and mixed verbal ability tests, speech production, writing, and memory for words.

The review of related literature reveals that the majority of the studies on gender differences in verbal abilities are conducted within the field of cognitive psychology. In addition, previous research into gender differences focused mainly on native speakers of English as subjects. Little research exists on whether or not EFL males and females have similar or different sizes of reading vocabulary. Accordingly, if conclusive findings are to be reached, more comprehensive research data should be accumulated concerning vocabulary use among different populations. The second purpose of this study, therefore, is to explore whether a significant difference exists between college females and males in Taiwan with regard to their English vocabulary ability. The study seeks to determine as well which gender predominates at the extremities of the normal curve distribution as expressed in grade equivalents.

Method

Definition

General English Vocabulary: Research into vocabulary acquisition and use has divided vocabulary into four levels: general or high frequency, academic vocabulary, technical vocabulary or words for specific purposes, and low frequency words (Chung & Nation, 2003; Coxhead, 2006; Nation, 2001). In this study, “general English vocabulary” is defined to mean “words of general usefulness, not technical vocabulary or words for specific purposes or specialized words that stand for abstract concepts, such as *democracy*, *propaganda*, and *referendum*.”

Grade Equivalents (GE): In this study, grade equivalents are used as the major index as the test performance is expressed in common units with which we are all familiar (Lipson & Wixson, 1991). A grade equivalent of 7.0, for example, indicates the seventh grade, the beginning month of the first semester or September; and a grade equivalent of 7.1 means the seventh grade, the second month of the first semester or October (MacGinitie, MacGinitie, Maria & Dreyer, 2002). Interpretations of grade norms, however, can be misleading and inaccurate, as they are not equivalent to performance standards and only indicate the average score obtained by the students in the standardization sample at a particular grade level (Gronlund, 1985). Besides, the grade equivalents at the extremes are "extrapolated by projecting the average performance of students at grade levels that were tested to performance of students at grade levels that were not tested," as no individuals from the extreme grade group were included in the norming sample (Baumann, 1988, p.37).

In the study, the reason for using grade equivalents but not vocabulary size as the major index lies in the fact that published estimates of students' vocabulary size vary widely (Lorge & Chall, 1963). For example, Lorge and Chall (1963) estimated that six-year olds know about 5,000 words. The other estimates, concerned with word understanding and usage, is that six-year olds know about 6,000 different words (Moe, 1974). Summarizing the findings of studies on vocabulary size at Grade 3 and at Grade 12, Nagy and Herman (1987) reported that the variations of estimates range from 2,000 to 25,500 words at Grade 3 and from 7,800 to 47,000 words at Grade 12. As such results prove inconclusive; vocabulary size will not be used as the sole index in this study, for otherwise the results might be unreliable.

Participants

The participants were from the freshman cohorts enrolled in the departments of Medicine, Chinese Medicine, and Dentistry at a medical university in central Taiwan. Excluding 21 students who included students refusing to participate and the

overseas Chinese students enrolling in the departments, there are a total of 209 students (female 75; male 134) in this study. This includes 94 students from the Department of Medicine, 80 students from the Department of Chinese Medicine, and 35 students from the Department of Dentistry.

In this study, the high-achieving participants were selected to represent a specific level of school achievement among Taiwanese youths. The criteria used to define the ability of the participants are based on the results of the July Joint College Entrance Examination held in 2007. As the tests are written annually by a group of commissioned college professors around ten days before the examination is held, no reliability data have been established for all the tests. For example, the national mean scores of the English examination of the top one-third of the examinees in the years from 2003 to 2007 are 69 in 2003, 55 in 2004, 65 in 2005, 63 in 2006 and 57 in 2007. The data reveal that in consecutive years there exist great variations among the means except for in the two successive years: 2005 and 2006. Therefore, to better represent the subjects' ability in overall school achievement, the participants' ability was determined by the minimum college department admittance score (MCDAS) which shows the sum of the raw scores of the test subjects.

In 2007, 100,117 high school graduates registered for the examination (CEEC, 2008). Among them, 49,788 students took the examination for the science track. An MCDAS of 478.27 ranks the fifteenth among the science and engineering departments in Taiwan (See Appendix A: Mean MCDAS among the Top 15 Departments). The researcher then tallied the number of students accepted into the top 15 departments. The participants from the Department of Medicine were among the 879 students, representing the top 1.76%. The same steps were applied to the Department of Dentistry and the Department of Chinese Medicine. Table 1 displays the MCDAS, the mean of the MCDAS, and the standing by approximate percent within the 49,788 examinees of the respective department.

Table 1: MCDAS, Mean MCDAS, and Standing by Approximate Percent

	MCDAS	Mean MCDAS	Standing
Medicine	478.27	79.71	Within 1.76%
Dentistry	464.62	77.44	Within 3.02%
Chinese Medicine	460.01	76.67	Within 3.14%
National	n/a	41.89	n/a

Research Instrument

The vocabulary test in the Gates-MacGinitie Reading Tests (GMRT), Form S Level 7/9, the Fourth Edition, was used in this study (MacGinitie, MacGinitie, Maria & Dreyer, 2002). GMRT is one of the most commonly used group survey tests for measuring reading ability (Carpenter & Paris, 2005; Cook, Gerber & Semmel, 1997; Fisher, 2001; Lipson & Wixson, 1991; MacGinitie, MacGinitie, Maria & Dreyer, 2002; Nelson & Stage, 2007; Tatum, 2004). Level 7/9 contains 45 vocabulary questions with five choices for each question and two parallel forms. Each vocabulary test word is presented in a brief context frame (See Appendix B: Sample Questions). The words were chosen to represent an appropriate range of difficulty for the test level. The words are of general usefulness, not specialized words, and they represent an appropriate distribution of parts of speech. Wrong answers such as visual similarity, miscue, and association are presented in one to three of the five choices. The test is timed, with 20 minutes allowed to complete the test. Raw scores are converted into normal curve equivalents, national percentile ranks, national stanines, grade equivalents, and extended scale scores.

According to the technical report, Level 7/9 was standardized by employing a stratified random sampling design, which used a sample size of 12,153 students drawn from the 7^{th} to 9^{th} grades in 47 states in the United States. The reliability of this test was very high and was established through the following steps (MacGinitie, MacGinitie, Maria & Dreyer, 2002):

1. Kuder-Richardson Formula 20 (K-R 20) reliability coefficients were at 0.91 and 0.90 at Grade 7; 0.91 and 0.91 at Grade 8; and 0.92 and 0.91 at Grade 9.
2. Another 3,225 students participated in the equating studies. The correlations were established at 0.88 for the 7th graders who took both Level 6 and Level 7/9; at .89 for the 8th graders who took both Level 6 and Level 7/9; at .87 for the 9th graders who took both Level 6 and Level 7/9; and at .86 for the 10th graders who took both Level 7/9 and Level 10/12.
3. The third edition was also equated with the fourth edition. The correlations were found to be 0.89 at Grade 7; 0.90 at Grade 8; and 0.89 at Grade 9.
4. The stability of scores was established by computing the correlations between fall and spring raw scores. The correlations were found to be 0.88 at Grade 7; 0.90 at Grade 8; and 0.90 at Grade 9.

As presented in the technical report, criteria of validity were established through the following procedures (MacGinitie, MacGinitie, Maria & Dreyer, 2002):

1. Cultural diversity or bias review: The test items were examined by a panel of 15 reviewers: African American, Asian, Hispanic, and Native American consultants for bias and possible offensiveness and through analysis of Differential Item Functioning (DIF).
2. Other evidence of validity: The design of the Third Edition and the Fourth Edition of the GMRT was very similar (MacGinitie, MacGinitie, Maria & Dreyer, 2002), and the Third Edition was a valid and reliable group survey test of reading (Lipson & Wixson, 1991). The correlations between the two editions were very high, ranging from 0.89 to 0.90. Therefore, the studies of the validity and reliability on the Third Edition were often used to support the validity of the Fourth Edition (MacGinitie, MacGinitie, Maria & Dreyer, 2002). Significant correlations were found between the Third Edition and the verbal or English sections in Preliminary Scholastic

Assessment Test (PSAT), Scholastic Assessment Tests (SAT), American College Testing Program (ACT), and grade point average (GPAs).

The level was deemed right for this study as the maturity and difficulty of the content make it an appropriate vocabulary test for students from Grades 5 to 12.9. In addition, the subjects in this study were estimated to have been learning English for approximately seven to nine years; that is, from elementary school to the college freshman first year.

Procedure

The test was administered in the middle of the second semester. The participants were told about the purposes of the test before they began. They were encouraged to try their best, but were also told that they could stop participating at any point. Then, the vocabulary question sheets were distributed. Sample questions were practiced and questions were explained. The test was to last 20 minutes. No dictionary was allowed.

Grading

Two college English teachers hand-graded the test. The first teacher graded all the questions and tallied the total scores. Then the second teacher repeated the same grading procedure for accuracy. The raw score was then converted to grade equivalents by checking the 2006 norm provided in the Manual for Scoring and Interpretation.

Data Analysis

The SPSS version 11.5 was used to organize and analyze the data collected in the study. First reported were the descriptive statistics of the converted grade equivalents. Tables and a figure were then added to illustrate the results. Next reported was a *t*-test using the raw scores to examine whether or not a significant difference existed between males and females. The converted grade equivalents were also presented in tables and in a figure to illustrate the results.

Results

Distribution of Vocabulary Grade Level Scores

Table 2 lists the mean, median, mode, range, minimum, maximum and percentiles of the results by grade and month. The mean, median, and mode all fall within the Grade 7 level. The range is 8.80 grade levels or from the minimum, Grade 4.20, to the maximum, Grade PHS. Grade 8.2 level is in the upper 75 percentile while Grade 6.1 level shows a performance in the lower 25 percentile.

Table 2: Mean, Median, Mode, Range, Minimum, Maximum and Percentiles

N	Valid	209
Mean		7.1
Median		7.2
Mode		7.2
Std. Deviation		1.39
Range		8.80
Minimum		4.20
Maximum		PHS
Percentiles	25	6.1
	50	7.2
	75	8.2

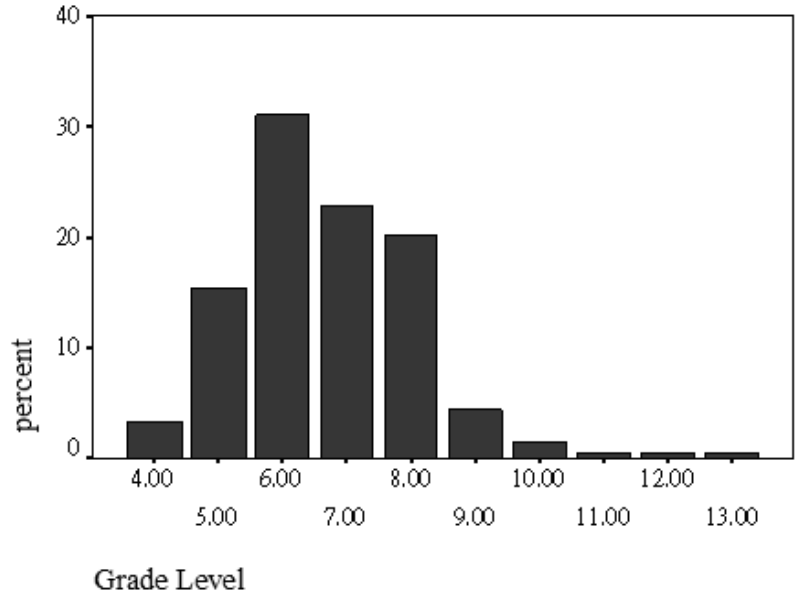
Table 3 presents the frequency and the percentage by grade. The table shows that both grade level 4.2 and grade level Post High School (PHS) are at the extreme ends of the curve, with only 3.35% at Grade 4 and 0.48% at PHS. Low percentages are also found at Grade 9 with only 4.30%, at Grade 10 with only 1.44%, at Grade 11 with only 0.48%, and at Grade 12 with 0.48%. The overall distributions by percent for the remaining grade levels in ascending order are 15.31% at Grade 5; 31.11% at Grade 6; 22.96% at Grade 7; and 20.09% at Grade 8. The table also shows that the overall performances of the majority of the participants cluster around Grade 5, Grade 6, Grade 7 and Grade 8.

Table 3: Frequency and Percentage by Grade

Grade	Frequency	Percent by 1.0 Grade	Cumulative Percent
Grade 4	07	03.35	03.35
Grade 5	32	15.31	18.66
Grade 6	65	31.11	49.77
Grade 7	48	22.96	72.73
Grade 8	42	20.09	92.82
Grade 9	09	04.30	97.12
Grade 10	03	01.44	98.56
Grade 11	01	00.48	99.04
Grade 12	01	00.48	99.52
PHS	01	00.48	100.0
Total	209	100.0	

Figure 1 visually displays the grade level distribution of the overall performances of the 209 participants by percent. The distribution spreads out in a bell shaped curve showing that the majority of the student subjects fell within grade 6, with slightly smaller groups in Grade 7 and Grade 8. A larger number also scored at grade 5. A small number scored at grades 9, 10, 11, 12 and PHS. No participants scored lower than Grade 4 level.

Figure 1: Grade Distribution by Percent



Comparing Gender Differences by Raw Scores

To test and compare the differences in vocabulary ability between female and male participants, the raw scores of the 75 females and the 134 males were used for the computation. Table 4 shows that the mean performance of females is 24.24 and that of males is 22.96. To put forward possible differences in the performances, a simple *t*-test was executed. A significant difference was found, t (207) = 2.042, $p < 0.05$, favoring the female participants. Table 5 shows the result of the *t*-test.

Table 4: Mean, Median, Mode, Range, Minimum, Maximum and Percentiles

	Female	Male
N Valid	75	134
Total Score: 45		
Mean	24.24	22.96
Median	24.00	23.00
Mode	27.00	21.00
Std. Deviation	3.834	4.594
Range	18.00	23.00
Minimum	16.00	12.00
Maximum	34.00	35.00
Percentiles 25	22.00	20.00

50	24.00	23.00
75	27.00	26.25

Table 5: Result of the *t*-test by Gender

Gender	N	Mean	Std. Deviation	*t*	df	Sig. (2-tailed)
Female	75	24.24	3.834	2.042	207	< .05
Male	134	22.96	4.594			
Total	209					

Comparing Gender Differences by Grade

Table 6 lists the mean, median, mode, range, minimum, maximum and percentiles of the results by grade and month. The mean, median, and mode of the males all fall within the Grade 6 level. The mean and mode of the females also fall within the Grade 6 level; however, the medium falls within the Grade 7 level. The grade range of the males is 9.0, spinning from Grade 4.0 to PHS, while that of the females is 7.0 ranging from Grade 5 to Grade 12. The percentiles show that Grade 8 is within the upper 75 percentile for the females, while it is only Grade 7.2 for the males. The difference is also observed within the 50 percentile, with the females at Grade 7 level and the males at Grade 6 level.

Table 6: Mean, Median, Mode, Range, Minimum, Maximum and Percentiles

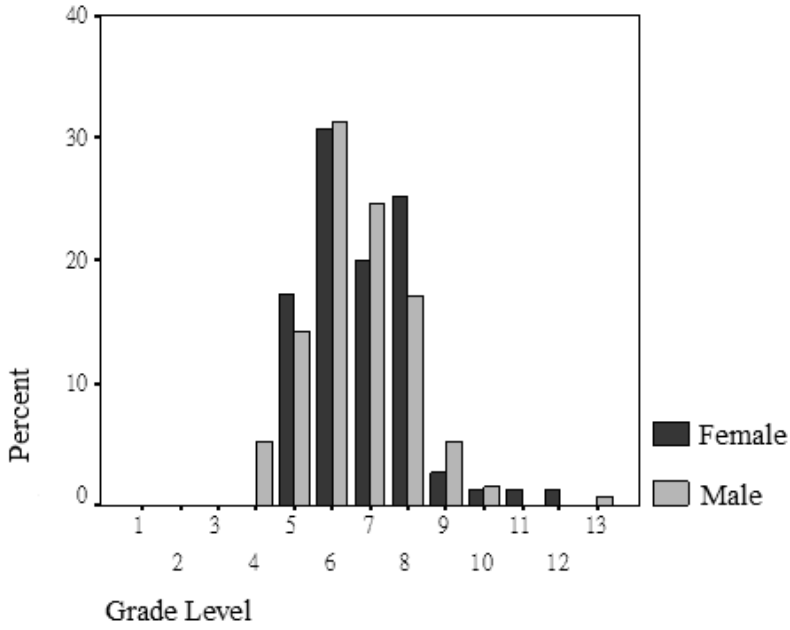
		Female	Male
N	Valid	75	134
Mean		6.8	6.6
Median		7.0	6.0
Mode		6.0	6.0
Std. Deviation		1.42	1.41
Range		7.0	9.0
Minimum		5.0	4.00
Maximum		12.0	PHS
Percentiles	25	6.0	6.0
	50	7.0	6.0
	75	8.0	7.2

Table 7 shows the frequency and percentage by grade × gender. The table reveals that no female participants scored below Grade 5; however, 5.22% of the male participants scored at Grade 4. By Grade 6 the cumulative percent difference between female and male is 48.00% vs. 50.74%, that is, a difference of 2.74%. More male participants scored below Grade 8 (75.37%) than female participants (68.00%). A greater difference is also observed at Grade 8 as more females scored at that grade (25.33%) than males (17.16%), with a difference of 8.17%. Figure 2 displays the interactions of the distributions of the performance between females and males. The figure also indicates that lower grade scores were found within the male participants.

Table 7: Frequency and Percentage by Grade × Gender

Grade	Frequency Female vs. Male	Percent Female vs. Male	Cumulative Percent Female vs. Male
Grade 4	00/07	00.00/05.22	00.00/05.22
Grade 5	13/19	17.33/14.18	17.33/19.40
Grade 6	23/42	30.67/31.34	48.00/50.74
Grade 7	15/33	20.00/24.63	68.00/75.37
Grade 8	19/23	25.33/17.16	93.33/92.53
Grade 9	02/07	02.67/05.22	96.00/97.75
Grade 10	01/02	01.33/01.49	97.33/99.24
Grade 11	01/00	01.33/00.00	98.66/99.24
Grade 12	01/00	01.33/00.00	100.0/99.24
PHS	00/01	00.00/00.75	----/100.0
Total	75/134	100.0/100.0	

Figure 2: Interaction of Distributions between Female and Male by Grade



Discussion

Vocabulary is customarily regarded by reading skill researchers as a major factor affecting the learning of reading. In the United States, word tests have been frequently included in formal and informal reading tests as a tool for reading diagnosis. In Taiwan, however, very few studies exist identifying students' vocabulary grade equivalents. This study, therefore, employed a widely-used norm-referenced group test to identify college students' English vocabulary grade equivalents with the intent of examining why most college students shun reading college textbooks in English. The goals of this study are twofold: 1) to establish the grade distributions corresponding to English vocabulary levels and 2) to determine whether if differences in English vocabulary levels exist between males and females.

The results of this study reveal that the vocabulary grade equivalents of the participants span from 4th grade level to PHS level, with only 3.35% at Grade 4 and

0.48% at PHS. Low percentages are also found at Grades 9 to 12. The mean vocabulary grade falls at Grade 7.1. Grades 6, 7 and 8 are shown to be the most common grades of those participating. The results show that nearly half or 49.77% of the subjects scoring lower than Grade 7. In addition, if Grade 10 is used as the cut-off grade, the results reveal that only 2.88% of the subjects scored at and above Grade 10, with the majority or 97.12% of the subjects scoring below Grade 10.

The results of this study reveal that 49.77% of the participants performed below 7th grade level. Grades 6, 5 and 4 are equivalent to the elementary grade levels of native speakers of English. Students within these vocabulary grade levels have greater difficulties in reading college discipline-specific textbooks in English. The results confirm our daily observations on the way college freshmen read content area textbooks. As highlighted earlier, they follow the time-consuming process of constantly looking up and writing down Mandarin equivalents of the unfamiliar English words. This modus operandi ultimately leads students to abandon English textbooks altogether and to use Mandarin Chinese versions whenever available.

What are the root causes of this common problem in learning? Some are obvious—such as gross differences in origin and structure between the two languages involved. More immediately, as Mandarin Chinese is the dominant school language, children beyond elementary school experience severe time-constraints when learning English. English at secondary school becomes only one of the compulsory school subjects among mathematics, biology, physics, chemistry, Mandarin Chinese, history, geography and civics. The time allocated to reading English at school and at home is limited. Other than learning opportunities, as Schmitt (1999) states, factors such as motivation and aptitude also contribute to a student's ability to succeed in learning a second language. Hence, to seek out the definite answer, more research is needed. The factors that have been discouraging students in English learning must be explored.

In response to the second research question, the *t*-test using raw scores indicates that a significant difference exists between males and females regarding English vocabulary ability. The comparison using grade norm further reveals that 50.74% of the male participants scored below Grade 7, whereas the percentage of the female subjects is 48.00%. The findings further reveal that no females scored below Grade 5; while more than 5% of the male participants scored below Grade 5. The interaction of the distributions between males and females shows that extremities, such as lower grades, tend to be found among male students. This finding confirms the conclusion by Halpern (2000) that the extremely low-ability end of the verbal abilities is often found among males.

What are the major causative factors leading to this discrepancy in learning? The causes are multi-grossing. Research has documented that women outperform men on verbal memory (Halpern, 2000; Kimura & Clarke, 2002). Green and Oxford (1995) found significant correlations between gender and the use of language learning strategies. Females use memory, cognitive, compensation, metacognitive, affective and social strategies more often than males. Other studies have also documented that girls possess more positive attitudes toward reading than boys (Morgan & Douglas, 2007) and that boys preferred sports and girls preferred music and art (Evans, Schweingruber & Stevenson, 2002). Even so, other causes such as preferences and time spent on learning await further examination.

Conclusions

As Singer and Donlan (1989) indicated, in the native English setting, to read materials needed on the job with a 70% level of comprehension, mechanics would have to read at a 9th-grade level, and supply clerks at a10th-grade level; while many beginning college students read at least at the 13th-grade level. The results of the current study focusing on good readers' general English vocabulary show that 49.77% of the subjects' word grade equivalent fell below Grade 7. Grades 6 and 5 are equivalent to the elementary grade levels of native speakers of English. This

finding suggests that in Taiwan, obviously not all of the college students are able to manage reading from discipline-specific texts in English successfully.

This study has shown that gender differences in English reading vocabulary ability among high-achieving EFL college students do exist. More male students scored at the lower extremities. However, this study reveals that nearly the majority of the subjects scored at and below Grade 6. These results suggest that both sexes will experience difficulties in reading college-level textbooks in English. Male college students will encounter more reading difficulties than females.

Implications for Further Research

Why do high achievers possess such low reading vocabulary ability? To explore why low English reading vocabulary also exists among high achievers such as students majoring in medicine and dentistry, qualitative studies are needed to depict their aptitudes, English learning experience, and time spent on English learning. On the other hand, qualitative research should also be conducted on the students who scored at the high extremities to demarcate the factors that have contributed to their success in learning English vocabulary.

To build up a comprehensive and reliable profile, more studies should be conducted with subjects selected among male and female students enrolled at the most prestigious universities, such as the medical students at National Taiwan University. English majors should also be included in such studies to allow researchers to collect conclusive information regarding the grade distributions of the college students in Taiwan. Vocabulary studies should also target high school students at different grade levels and genders. Together with the administering of norm-referenced reading tests to explore reading grade equivalents, the students can be informed about their grade levels precisely and the teachers can provide earlier interventions to ensure a more effective vocabulary acquisition.

As the test in this study involved only the use of general vocabulary, more

studies on students' knowledge of specific vocabulary in sciences should be designed and conducted. In Taiwan, most majors in science and engineering are assigned to read discipline-specific textbooks in English. It will be more valuable to understand how they manage to learn the science vocabulary in their textbooks. These studies, together with the findings of the current study, may also provide a way to understand why much of their reading time is taken up with looking up words in the dictionary and recording their Chinese equivalents.

Implications for Teaching

The grade norm, as this study demonstrated, provides a more precise index and covers an extensive range of ability levels, which will enable the teachers to determine students' needs and offer appropriate assistance. Classroom teachers can use them as a quick way to estimate their student's reading grade levels in order to assign reading materials at their grade levels when there are no formal reading tests available.

As regards the ability grouping for effective teaching, this study however reveals that even among good English readers in a classroom, the discrepancy of abilities covers a range from Grade 4 to Grade PHS. This broad range will pose a problem for the teacher who attempts to group students by ability. In addition, the small number of students at both extremes of the range, (Grades 4, 5, 11, 12 and PHS), certainly need special attention and different reading materials in keeping with their ability in English reading and comprehension. For cost effectiveness in educational resources, it is rather difficult for most schools to accept the idea of small group teaching due to the cost of educational resources and scheduling times.

This study, however, offers a different approach for locating the students' English reading ability in order to assign appropriate reading materials, which is especially useful if independent reading is adopted as a salient feature of instruction. Better education is possible as the student can experience more successful reading with enhanced comprehension when reading materials that match their experiences

and reading level. A slow but steady growth in English reading ability is achieved a step at a time with less chance of regression or failure. When a student is fettered with material that he or she is unable to decipher or understand, the learning curve reports little gain while the forgetting curve climbs rapidly. Little to no learning occurs and frustration abounds for such over-challenged students.

References

Astika, G. G. (1993). Analytical assessment of foreign students' writing. *RELC Journal*, 24, 61~72.

Baumann, J. F. (1988). *Reading assessment: An instructional decision-making perspective.* Columbus, OH: Charles E. Merrill.

Burns, P. C. & Roe, B. D. (1989). *Burns/Roe informal reading inventory ~ Preprimer to twelfth grade.* Boston: Houghton Mifflin.

Carpenter, R. D. & Paris, S. G. (2005). Issues of validity and reliability in early reading assessments. In Paris, S. G. & Stahl, S. A. (Eds.), *Children's Reading Comprehension and Assessment,* (pp.279~304). Mahwah, NJ.: Lawrence Erlbaum Associates.

CEEC - College Entrance Examination Center. (2008). Appoint exam profile. (Retrieved September 20, 2008 from www.ceec.edu.tw)

Chall, J. S. (1987). Two vocabularies for reading: Recognition and meaning. In Curtis, M. E., & McKnown, M. G. (Eds.) *The Nature of Vocabulary Acquisition* (pp.7~15). Hillsdale, NJ.: Lawrence Erlbaum Associates.

Chen, S.Y., Hu, C.F., & Liu, L.R. (2008). Students withdraw from courses when they no longer listen with comprehension. *The Liberty Times*. (Retrieved November 20, 2008 from http://www.libertytimes.com.tw/2008/new/sep/15)

Cheng, T. Y. (1993). The syntactical problems Chinese college students meet in reading English technical textbooks. *Resource in Education.* (ERIC Document Reproduction Service No. ED 364 094).

Cheng, T. Y. (2007). Locating college students' reading grade level through criterion-referenced vocabulary tests. *Journal of General Education*, (12), 201~213。

Cheng, T. Y. & Hung, T. I. (2002). College content teachers' attitudes toward English reading instruction in the content areas in an EFL environment. *Journal of Hsing Kuo College of Management*, (1), 25~38.

Chung, T. M. & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in Foreign Language*, 15 (2), 103~116.

Cook, B. G., Gerber, M. M., & Semmel, M. I. (1997). Are effective schools reforms effective for all students? - The implications of joint outcome production for school reform. *Exceptionality*, 7 (2), 77~96.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213~239.

Coxhead, A. (2006). *Essentials of teaching academic vocabulary*. Boston: Houghton Mifflin.

Ekwall, E. E. & Shanker, J. L. (1988). *Diagnosis and remediation of the disabled reader* (3rd. Ed.). Needham Heights, Massachusetts: Ally and Bacon.

Evans, M. E., Schweingruber, H., & Stevenson, H. W. (2002). Gender differences in interest and knowledge acquisition: The United States, Taiwan, and Japan. *Sex Roles*, 47, 153~167.

Fisher, G. (2001). Cross age tutoring: Alternatives to the reading resource room for struggling adolescent readers. *Journal of Instructional Psychology,* 28 (4), 234~240.

Green, J. & Oxford, R. (1995). A close look at learning strategies, L2 proficiency and gender. *TESOL Quarterly*, 29 (2), 216~297.

Gronlund, N. E. (1985). *Measurement and evaluation in teaching* (5th Ed.). New York: Macmillan.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd Ed.). Mahwah, NJ.: Lawrence Erlbaum Associates.

Hedges, L. V. & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41~45.

Hendricks, J. (1988). *The whole child* (4th Ed.).Columbus, OH: Merrill.

Hu, C.H., Chen, S.., & Liu, L.R. (2008). In Taiwan, 10% of college subject classes are taught entirely in English. *The Liberty Times*. (Retrieved November 20, 2008 from www.libertytimes.com.tw/2008/new/sep/15)

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236~248.

Hyde, J. S. & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104, 53~69.

Hyde, J. S. & McKinley, N. M. (1997). *Gender differences in cognition: Results from meta-analyses*. In M. Crawford, J. S. Hyde, & J. T. E. Richardson (Eds.), *Gender Differences in Human Cognition,* (pp.30~51). New York: Oxford University Press.

Karlsen, B. & Gardner, E. (1986). *Stanford diagnostic reading test* (3rd Ed.). San Antonio, TX: The Psychological Corporation.

Kimura, D. & Clarke, P.G. (2002). Women's advantage on verbal memory is not restricted to concrete words. *Psychological Reports*, 91, 1137~1142.

Kozol, J. (1986). *Where stands the republic? Illiteracy: a warning and a challenge to the nation's press*. Atlanta: Cox Enterprises.

Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics,* 16, 307~322.

Lipson, M. Y. & Wixson, K. K. (1991). *Assessment & instruction of reading disability: An interactive approach*. New York: HarperCollins Publishers.

Lorge I. & Chall J. S. (1963). Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *Journal of Experimental Education,* 32, 147~157.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates-MacGinitie reading tests – Technical report (Forms S and T)* (4th Ed.). Rolling Meadows, IL: Riverside Publishing.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2007). *Gates-MacGinitie reading tests – Manual for scoring and interpretation (levels 7/9 & 10/12)* (4th Ed.). Rolling Meadows, IL: Riverside Publishing.

Moe A. J. (1974). *A Comparative Study of vocabulary diversity: The speaking vocabularies of selected first-grade primers, and the vocabularies of selected first-grade trade books*. A paper presented at the Annual Meeting of the American Educational Research Association, Chicago. (ED 090 520).

Morgan, P. L. & Douglas, F. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children,* 73 (2), 165~183.

Nagy, W. E. & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. G. McKeown & M. E. Curtis (Eds.), *The Nature of Vocabulary Acquisition*, (pp.19~36). Hillsdale, NJ.: Lawrence Erlbaum Associates.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nation, I. S. P. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, Acquisition and Testing* (pp. 3~14). Amsterdam: Benjamins.

Nelson, R. N. & Stage, S. A. (2007). Fostering the development of vocabulary knowledge and reading comprehension though contextually-based multiple meaning vocabulary instruction. *Education & Treatment of Children*, 30 (1), 1~22.

Newman, A. P. & Beverstock, C. (1990). Adult literacy: Contexts & challenges. Newark, DE: International Reading Association.

Qian, D. D. (1999). Assessing the roles of depth and breath of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, 56, 283~307.

Schmitt, N. (1999). The relationship between TOEFL vocabulary items and meaning, association, collocation and word-class knowledge. *Language Testing*, *16*, 189~216.

Singer, H. & Donlan, D. (1989). *Reading and learning from text*. Hillsdale, NJ.: Lawrence Erlbaum Associates.

Slossen, R. L. (1988). *Slossen oral reading test (SORT)*. New York: Slossen Educational Publications.

Sweet, R.W. Jr. (1997). Don't read, don't tell: Clinton's phony war on illiteracy. *Policy Review*, 83, 38~42.

Tatum, A. W. (2004). A road map for reading specialists entering schools without exemplary reading programs: Seven quick lessons a reading specialist who worked in a high-poverty; urban elementary school provides seven lessons that can serve as guideposts for others. *The Reading Teacher*, 58 (1), 28~38.

U.S. Department of Education's National Center for Education Statistics. (2003). *The national adult literacy survey*. (Retrieved January 10, 2005, from http://nces.ed.gov/naal/resources/resources.asp)

West, M. (1953). *A general service list of English words*. London: Longman, Green, & Company.

Williams, J. E. (1994). Gender differences in high school students' efficacy-expectation/performance discrepancies across four subject matter domains. *Psychology in the Schools*, 31, 232~237.

Worrell, F. C., Roth, D. A., & Gabelko, N. H. (2007). Elementary reading attitude survey (ERAS) scores in academically talented students. *Roeper Review,* 29 (2), 119~124.

Xue, G. & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*, 215-229.

Yiau, C. K. (1993). *The approaches Chinese graduate and doctoral students use in reading both Chinese and English academic publications*. Paper presented at the 10th conference on English Teaching and Learning in the Republic of China.

Appendix A: Mean MCDAS among the Top 15 Departments

Name of University	Department	MCDAS	Mean
National Taiwan University	Medicine	525.03	87.51
National Cheng Kung University	Medicine	512.80	85.47
National Yang-Ming University	Medicine	510.44	85.07
Chang Gung University	Medicine	506.9	84.48
National Taiwan University	Electronic E[1]	419.52[4]	83.90
National Taiwan University	Dentistry	501.46	83.58
Taipei Medical University	Medicine	495.4	82.57
National Taiwan University	Physics	410.32[4]	82.06
National Yang-Ming University	Dentistry	489.54	81.59
National Taiwan University	Materials S & E[2]	407.94[4]	81.48
National Chiao Tung University	Elec. E & C S[3]	407.3[4]	81.46
Kaoshiung Medical University	Medicine	486.9	81.15

Tzu Chi University	Medicine	481.51	80.25
Fu Jen Catholic University	Medicine	480.67	80.11
China Medical University	Medicine	478.27	79.71

[1]Electronic Engineering

[2]Materials Science and Engineering

[3]Electrical Engineering and Computer Science, Honors Program

[4]Test subjects: English, mathematics for science, Mandarin, chemistry, physics

Appendix B: Sample Questions[5]

V-1. a big <u>garage</u>

Ⓚ place for cars

Ⓛ machine

Ⓜ sidewalk

Ⓝ covered porch

Ⓞ cloth sack

V-2. They will <u>close</u> it.

ⓟ stay near

Ⓠ begin

Ⓡ make

Ⓢ shut

Ⓣ go past

[5]Due to copyright restrictions, only sample questions from the Vocabulary sample page in the test booklet are taken.