

# THE TOEFL ACHIEVEMENT OF KU STUDENTS

*Chariya Pholprasert*

## ABSTRACT

This research emphasizes the whole picture of the TOEFL achievements of the students, in three consecutive steps—before being given TOEFL instruction and techniques to do the test, in the middle of the semester, and at the end of the course—all of which are presented in the form of TOEFL scores converted from their raw scores.

To look for the students' English language problems after having taken a one-semester course, the researcher has analyzed their achievement of each item contained in the final examination and presented her analysis in the form of correct answers and percentage. Any item on which the students' achievement as a whole was lower than 50% was considered to be a problem.

It was discovered that the means ( $\bar{X}$ ) of the 31 students' achievement of the final examination was 482, the highest score was 590, and the lowest one was 400. The most serious problems the students had were with the listening section, followed by structure and reading in that order.

One very important fact about learning a foreign language is that it takes time, study, and practice. It is therefore suggested that the test-taker arrange a regular program of language study to best improve his general ability to understand and use the English language. As such, he will achieve higher and more desired TOEFL scores in no time at all.

**Key words:** *TOEFL, achievement, paper-based test*

# I. ANALYTICAL EXPOSITION

## Statement of Purpose

The purpose of this study is to work out the language achievement of Kasetsart University students after having taken a TOEFL PBT course and to point out the items that are their problems.

## Rationale

Since foreign students are required to take TOEFL before furthering their studies at advanced levels in the United States, Australia and other countries including Thailand, the researcher is interested in her students' TOEFL achievement as well. As TOEFL tests cover contexts of various academic fields—spoken as well as written language—it is worthwhile to do research on her students' language development during a one-semester TOEFL preparation course by having them do the TOEFL test from time to time during the semester, i.e. the pre-test, the post-test, and the final examination. The results of their final achievement are also suited to being analyzed to look for the items that are language problems for them. As such, the researcher could provide valuable suggestions to test-takers who are preparing themselves to take TOEFL for further studies both at home and abroad.

## Definitions of Terms

For the purpose of this research, the following terms are provided:

1. TOEFL stands for "Test of English as a Foreign Language." In this research, the TOEFL tests include the TOEFL Sample Test, the TOEFL Final Test Exercise, and the Final Examination, all of which are prepared by test specialists at Educational Testing Service. All items in these tests are standardized and have different levels of difficulty.
2. KU stands for "Ka-setsart University," Bangkhen Campus.
3. TOEFL achievement is the result of the students' ability to do the pre-test, the post-test, and the final examination, the raw scores of which are converted into and displayed in the form of TOEFL scores.
4. KU students refer to 31 Kasetsart University students who took the course entitled "English for Standardized Tests : TOEFL" in the summer session (2005).
5. The pre-test is the TOEFL Sample Test.

6. The post-test is the TOEFL Final Test Exercise.
7. The final examination is a real TOEFL test, chosen to assess the researcher's students' achievement in this research.
8. Core parts are the basic and most important parts of the sentence, which include S (Subject), V(Verb), O (Object), and C (Complement).
9. Appositives refer to noun phrases and adjective phrases that clarify or qualify the preceding or following nouns, with a comma or commas before or after them.
10. Collocation is the combination of words formed when two or more words are frequently used together in a way that sounds correct.
11. TOEFL PBT is the TOEFL paper-based test, which largely includes the listening, structure, and reading sections.
12. TOEFL CBT is the TOEFL computer-based test, which comprises the listening and structure adaptive sections, the linear reading section, and the essay section.

## Scope of the Research

This research emphasizes the whole picture of the TOEFL achievements of the students, in three consecutive steps—before being given TOEFL instruction and techniques to do the test, in the middle of the semester, and at the end of the course—all of which are presented in the form of TOEFL scores converted from their raw scores.

To look for the students' language problems after having taken a one-semester course, the researcher has analyzed their achievement of each item contained in the final examination and presented her analysis in the form of correct answers and percentage. Any item on which the students' achievement as a whole is lower than 50% is considered to be a problem.

## Assumptions

Because of the nature of this research, the following assumptions are made in carrying out the investigation :

1. It is assumed that the students who are the subjects in this research do not have the same English language ability even though all of them have fulfilled the three pre-requisite English background courses required by the university before taking the TOEFL course.

2. It is also assumed that after taking the pre-test, the students have attended class regularly, been taught and informed of all the techniques in taking TOEFL, and also completed the self-practice exercises assigned by the instructor-researcher before taking the post-test in the middle of the semester and the final examination at the end of the course.

### **Discussion of Operational Aspects**

The researcher will demonstrate the development of the students' TOEFL achievement by tabulating the results of their pre-test, post-test, and final examination in the form of the TOEFL scores, which are converted from their raw scores. The results of their post-test as well as their final examination scores will also be discussed to give a clear picture of their achievements as well.

To find out the items with which the students have language problems, the researcher has analyzed their final examination item by item and presented it in the form of the number of students with correct answers together with the transformation of the raw score to percentage. The students' achievement of any item lower than 50% is considered to be a problem.

### **Expected Findings**

After a thorough study of the TOEFL achievement of KU students, the following findings were expected:

1. After the pre-test, the students' TOEFL achievements of the post-test and the final examination should be better because they had been instructed in the English language—both spoken and written—and informed of all the techniques to tackle TOEFL problems in class. In addition, they had to systematically do self-practices, as planned and assigned by the instructor-researcher, outside the classroom as well.

2. The results of the item analysis and the students' achievement of the final examination would reveal the language points that are problems for Thai students. The TOEFL test-taker and the instructor would then be aware of the problems and thus pay close attention to them. As a result, the test-taker could achieve a higher TOEFL score than expected.

## II. CRITICAL CONTEXT

### Related Literature

In the TOEFL PBT as well as CBT tests, the test-takers have an opportunity to demonstrate their language ability in listening, reading, and writing skills. The three skills mentioned will enable those who want to further their studies at the advanced levels to communicate and attain their educational goals.

Since the beginning of the new millennium, there have been some scholars who have done relevant researches (TOEFL, 2003) concerning TOEFL as follows:

In Spring 2003, Lawrence J. Stricker conducted a research on "The Performance of Native Speakers of English and ESL Speakers on the TOEFL CBT and GRE General Test." The purpose of this study was to replicate previous research on the construct validity of the paper-and-pencil version of the TOEFL test (TOEFL PBT) and extend it to the TOEFL CBT. Two samples of GRE test takers were used: native speakers of English specially recruited to take the TOEFL CBT, and ESL test takers who routinely took the TOEFL CBT recently. Native speakers performed well on TOEFL, relative to ESL test takers and to the maximum possible scores on the test, and varied less in their test performance than did ESL test takers; TOEFL CBT scores were highly but not perfectly correlated with GRE General Test scores for both groups of test takers; regressions of the General Test verbal scores on the TOEFL scores for ESL test takers were nonlinear, and the regressions of the other General Test scores were linear; trends in the variances of the General Test verbal scores associated with TOEFL scores were also nonlinear, and the trends were either unsystematic or negatively linear for the other General Test scores. All of the findings are consistent with previous results with the paper-and-pencil TOEFL, support the construct validity of the TOEFL CBT, and illuminate its interplay with ability tests for ESL test takers.

Also in the same year, Michael Rosenfeld, Philip K. Oltman, and Ken Sheppard carried out a study on "Investigating the Validity of TOEFL: A Feasibility Study Using Content and Criterion-Related Strategies." The purpose of this study was to investigate the feasibility of two complementary approaches to assessing the validity of the TOEFL examination. One approach used evidence based on test content. In the context described in this report, evidence based on test content refers to the degree to which the items on the TOEFL

examination are representative of the knowledge and skills required to demonstrate English proficiency in undergraduate and graduate programs throughout the United States and Canada. The content-oriented approach used in this pilot study involved item-rating procedures that were designed to evaluate and document the relationship between the language tasks or behaviors previously identified as important for academic success and the test items used to measure them. The second approach used a criterion-related validation strategy. In this aspect of the study, experimental rating scales were developed for use by faculty to evaluate students' current level of English language proficiency. These scales were designed to sample the domain of behaviors previously identified as important.

In November 2001, Claudia Leacock and Martin Chodorow studied "Automatic Assessment of Vocabulary Usage Without Negative Evidence." As part of the TOEFL program's effort to develop performance-based measures of communicative competence, they implemented and evaluated an automated statistical method for assessing an examinee's use of vocabulary words in constructed responses. Their error-detection system, ALEK (Assessing Lexical Knowledge), infers negative evidence from the low frequency or absence of constructions in 30 million words of well-formed, copy-edited text from North American newspapers. ALEK detects two types of errors: those that violate basic principles of English syntax (e.g. agreement errors as in *\*a desks*) and those that show a lack of information about a specific word (e.g. treating a mass noun as a count noun as in *\*a pollution*). The system evaluated word usage in essay-length responses to TOEFL prompts. ALEK was developed using 3 words and was evaluated on an additional 20 words that appeared frequently in TOEFL essays and in a university word list. The system performed with about 80% precision and 20% recall. False positives (correct usages that ALEK identified as errors) and misses (usage errors that were not recognized by ALEK) were analyzed, and methods for improving system performance were outlined.

In August the same year, Ginther did a survey on "Effects of the Presence and Absence of Visuals on Subjects' Performance on TOEFL CBT Listening Comprehension Stimuli." Now that TOEFL is computer-based, listening items are being created that include both audio and visual information. This study was conducted in order to begin to understand the effects of different types of visual presentations. The design examined the effects of language proficiency (high or low), still photos (present or absent), and type of stimuli (dialogues/short

conversations, academic discussions, minitalks with context visuals, minitalks with content visuals) on performance on standard multiple-choice listening items. Three two-way interactions were significant: proficiency by type of stimuli, type of stimuli by time, and type of stimuli by visual condition. The weakest of these interactions, type of stimuli by visual condition, was the most interesting and indicated that the presence of visuals results in facilitation of performance when the visuals bear information that complements the audio portion of the stimulus. The majority of the subjects indicated a strong preference for the presence of visuals.

## Context of Research

1. To demonstrate the development of the students' TOEFL achievement before, during, and after a one-semester course, the researcher has administered the pre-test, the post-test, and the final examination. The students' raw scores of each test are converted into the TOEFL scores and tabulated accordingly. The ID Codes with the last three undisclosed digits have been arranged from the largest to the smallest number of students from each faculty, enrolled in the TOEFL course, as follows:

**Table 1** The TOEFL scores of the pre-test, the post-test, and the final examination with the students' ID Codes.

No.	ID Codes	Pre-test	Post-test	Final
Faculty of Humanities				
1.	44112XXX	447	(+10) 457	(+56) 513
2.	45110XXX	427	(+13) 440	(+37) 477
3.	45112XXX	567	(-20) 547	(+10) 557
4.	45112XXX	487	(+46) 533	(-26) 507
5.	45113XXX	437	(+3) 440	(-10) 430
6.	45113XXX	497	(-14) 483	(+24) 507
7.	46110XXX	447	(+20) 467	(-7) 460
8.	46110XXX	470	( - ) 470*	(-17) 453

Table 1 (continued)

No.	ID Codes	Pre-test	Post-test		Final	
<b>Faculty of Science</b>						
9.	45041XXX	450	(+10)	460	(+30)	490
10.	45042XXX	453	(+20)	473	(+20)	493
11.	46041XXX	383	(+10)	393	(+10)	403
12.	46045XXX	403	(-6)	397	(+20)	417
<b>Faculty of Engineering</b>						
13.	44053XXX	447	(+43)	490	(-17)	473
14.	44055XXX	423	(+24)	447	(-17)	430
15.	44055XXX	473	(+4)	477	(+6)	483
16.	46054XXX	453	(+107)	560	( - )	560*
<b>Faculty of Social Sciences</b>						
17.	45080XXX	437	(+50)	487	(-14)	473
18.	45082XXX	530	(+13)	543	(+20)	563
19.	46082XXX	457	(+50)	507	(-54)	453
<b>Faculty of Architecture</b>						
20.	45190XXX	353	(+54)	407	(+13)	420
21.	45190XXX	437	(-20)	417	(+10)	427
22.	45190XXX	473	(-10)	463	(+57)	520
<b>Faculty of Veterinary Medicine</b>						
23.	46090XXX	463	(+17)	480	( - )	480*
24.	46090XXX	570	(-17)	553	(+37)	590
25.	47090XXX	507	(-47)	460	(+37)	497

Table 1 (continued)

No.	ID Codes	Pre-test	Post-test		Final	
Faculty of Economics						
26.	45071XXX	547	(+6)	553	(-3)	550
27.	46071XXX	430	(+7)	437	(+30)	467
Faculty of Forestry						
28.	44031XXX	400	(+7)	407	(-7)	400
Faculty of Fisheries						
29.	46021XXX	417	(-10)	407	(+30)	437
Faculty of Education						
30.	46062XXX	360	(+97)	457	(-7)	450
Faculty of Business Administration						
31.	46133XXX	507	(+10)	517	(+36)	553

2. To find out the items that are language problems to her students, the researcher had analyzed the questions in the final examination. Those in the listening and reading comprehension sections were primarily analyzed in terms of stated questions, implied or inferred questions whereas the grammatical points in the structure and written expression section were mainly based on the core parts and the parts of speech in each item.

In the listening section, the stated questions refer to the questions to which the students could choose the answer from their understanding of the wording they have heard in the conversation, e.g.

Woman : *The seminar schedule for today has been cancelled.*

Man : *That's too bad. But at least the one for next week's still on.*

Narrator : *What does the man mean?*

Answer : **(D) Another seminar will take place the following week.**

As for the implied or inferred questions, the students have to form an opinion or guess from the information they have heard, e.g.

Man : *I miss class this morning. Could you lend me your notes?*

Woman : *My notes? You've never seen my handwriting, have you?*

Narrator : *What does the woman imply?*

Answer : **(C) Her handwriting is difficult to read.**

The frequency analysis of the students' achievement of the items that were their language problems (lower than 50%) in the listening section is presented together with the percentage as follows :

**Table 2** The analysis of the type of question in the listening section, together with the number of students with correct answers, and the percentage.

Item	Type of question : Stated, Implied or Inferred	No. of students with correct answers	%
6.	Implied	14	45.16
7.	Stated	15	48.39
12.	Inferred	13	41.94
13.	Implied	14	45.16
14.	Implied	13	41.94
15.	Inferred	10	32.26
16.	Inferred	12	38.71
17.	Stated	11	35.48
19.	Inferred	5	16.13
20.	Inferred	12	38.71
21.	Implied	13	41.94
22.	Implied	9	29.03
24.	Implied	15	48.39
25.	Stated	13	41.94
27.	Inferred	4	12.90
29.	Inferred	14	45.16
30	Inferred	12	38.71

Table 2 (continued)

Item	Type of question : Stated, Implied or Inferred	No. of students with correct answers	%
35.	Stated	10	32.26
37.	Inferred	4	12.90
38.	Stated	14	45.16
40.	Stated	15	48.39
45.	Stated	13	41.94
46.	Stated	13	41.94
47.	Stated	5	16.13
48.	Stated	15	48.39
49.	Stated	10	32.26
50.	Stated	5	16.13

It is interesting to note that, of the 30 questions in Part A (Short Conversations), 26 are implied and inferred questions, with only 4 stated questions. The test-takers have to interpret what is said during the short conversations so that they can understand what the speakers mean. On the contrary, in Parts B and C (Long Conversations and Talks), 18 questions out of 20 are stated questions, with only 2 inferred questions. Therefore, if the test-takers concentrate on what is said in the conversations and talks, there will be no problems choosing the correct answers at all.

The structure and written expression section is designed to measure the test-takers' ability to recognize language that is appropriate for standard written English. There are two types of questions, with special directions for each type. Questions 1-15 are incomplete sentences with four choices to best complete each sentence, e.g.

During the early period of ocean navigation..., any need for sophisticated instruments and techniques.

(A) hardly was	(B) so that hardly
(C) there was hardly	(D) when these hardly was

In questions 16-40 each sentence has four underlined words or phrases. The test-takers have to identify the choice that must be changed in order for the sentence to be correct, e.g.

Serving several term in congress, Shirley Chisholm became an important United States  
 (A) (B) (C)  
politician.  
 (D)

Therefore, in the structure and written expression section, the researcher has chiefly pointed out the grammatical points tested in terms of the core parts and the parts of speech. Modifiers in the forms of adjective and adverb phrases are also given attention to. Subordinate clauses—the larger modifier strings—in the complex sentences referred to in this research include noun clauses, adjective clauses (both defining and non-defining), and adverb clauses. Appositives, collocations, and parallel forms of parts of speech are analyzed as well. The analysis of the students' achievement of the items that were their language problems (lower than 50%) in this section is presented together with the percentage as follows :

**Table 3** The analysis of the grammatical points tested in the structure and written expression section, together with the number of students with correct answers, and the percentage.

Item	Grammatical points tested	No. of students with correct answers	%
3.	Appositive noun phrase	0	0*
8.	C : noun phrase	13	41.94
9.	Gerund : object after preposition	13	41.94
10.	O : noun phrase	12	38.71
12.	Preceding adverb phrase	12	38.71
32.	Pronoun : subject repetition	10	32.26
35.	C : comparative form	11	35.48
36.	Preposition : regardless of	14	45.16
37.	Preposition : <i>From...to</i>	12	38.71
38.	Adjective : pre-modifier	2	6.45

Table 3 (continued)

Item	Grammatical points tested	No. of students with correct answers	%
39.	O : pronoun	10	32.26
40.	A verb : part of pre-modifiers	8	25.81

Of the 15 questions in the structure part, 8 questions concern core parts, 5 questions concern parts of speech (3 appositives, 1 gerund, 1 adverb phrase), and 2 questions concern defining adjective clauses. As for the 25 questions in the written expression part (items 16-40), 12 questions test parts of speech, 8 questions test core parts, 2 questions collocations, another 2 questions parallelism, and 1 question connector of defining adjective clause.

In the reading section, the stated questions refer to the questions that require the comprehension of the details in the wording of the passage read together with the “**not**” and “**except**” questions, e.g.

Based on the information in the passage, all of the following statements about sea anemones are true except that they ... .

(A) are usually tiny	(B) have flexible bodies
(C) are related to jellyfish	(D) are usually brightly colored

But the implied questions cover those that require the readers’ ability to find out the answers from the facts given in the passage, for example, topic and main idea, vocabulary, reference, inference, e.g.

It can be inferred from the passage that hydras ... .

(A) prey on sea anemones	(B) were named after a flower
(C) are related to sea anemones	(D) are usually found in Australia

The analysis of the questions together with the students’ achievement and the percentage of the items that were their language problems (lower than 50%) in the reading section are as follows :

**Table 4** The analysis of the type of question tested in the reading section, together with the number of students with correct answers, and the percentage.

Item	Type of question : Stated or Implied	No. of students with correct answers	%
4.	Implied : inference	9	29.03
12.	Stated : detail	0	0*
16.	Implied : reference	13	41.94
20.	Implied : vocabulary	8	25.81
33.	Implied : vocabulary	8	25.81
35.	Stated : detail	14	45.16
36.	Implied : reference	14	45.16
37.	Implied : vocabulary	9	29.03
39.	Implied : inference	5	16.13
40.	Stated : detail	2	6.45
42.	Implied : vocabulary	5	16.13
43.	Stated : detail	10	32.26
45.	Stated : detail	13	41.94

Of the 50 questions in the reading section, 18 are stated questions, 14 of which ask for details whereas 4 are “not” and “except” questions. As for the 32 implied questions, 10 questions concern vocabulary in context, another 10 questions are about reference, 8 questions ask for inference, and 4 questions main idea or topic.

### III. INTEGRATION

#### Conclusions

1. The three consecutive steps of the students’ TOEFL achievements in Table 1 shows that:
  - a. From the results of the post-test, 22 students out of 31 have improved their TOEFL scores, ranging from 3-107 points whereas eight students have lowered their scores, ranging from -6 to -47 points. One student (No.8) has

shown no improvement at all as he got the same score in the post-test as in the pre-test.

- b. From the results of the final examination, 18 students out of 31 have improved their TOEFL scores, ranging from 6-57 points whereas 11 students have lowered their scores, ranging from -3 to -54 points. Two students (Nos. 19 and 20) have shown no improvement.
- c. With computer assistance, the researcher had acquired the means ( $\bar{X}$ ) of the students' achievement of the final examination as 482. The highest score was 590 and the lowest one was 400.

2. For a better and clearer understanding of the item analysis of every section in the final examination, it is recommended that Appendix C—the Final Examination with Listening Script in the research entitled "THE TOEFL ACHIEVEMENT OF KU STUDENTS" (ISBN 974-537-717-1) also be given attention to. It is interesting to note that:

- a. The data in Table 2 show that of 50 questions in the listening section, the students have problems with 27 questions, which include 12 stated questions (items 7, 17, 25, 35, 38, 40, 45-50), 6 implied questions (items 6, 13-14, 21-22, 24), and 9 inferred questions (items 12, 15-16, 19-20, 27, 29-30, 37).
- b. Table 3 shows that of 40 questions in the structure and written expression, the students have problems with 12 questions, which include 4 items concerning core parts (8, 10, 35, 39) and 8 items concerning parts of speech (3, 9, 12, 32, 36-38, 40). It should be noted that no student could pinpoint the correct answer for item 3, which tests "Appositive noun phrase," but every student could find out the correct answer for item 17, which tests "Noun: with definite article."
- c. It is shown in Table 4 that of 50 questions in the five reading passages, the students have problems with 13 questions, which include eight implied questions (items 4, 16, 20, 33, 36-37, 39, 42), and five stated questions (items 12, 35, 40, 43, 45). It is also interesting to note that every student could find out the correct answer for item 10, but no student could pinpoint the correct answer for item 12, both of which were stated questions that test detail in the first passage.

## Suggestions

The results of the TOEFL achievement of KU students and the item analysis of the final examination clearly demonstrate that the most serious problems the students had were with the listening section, followed by structure and reading in that order. Therefore, the TOEFL test-taker and the instructor should be aware of and pay close attention to the aforementioned language problems while preparing for the test and giving the instruction in class.

Last but not least, before test-takers begin to prepare to take the TOEFL test, they should remember one very important fact about learning a foreign language—it takes time, study, and practice. The best way to improve their general ability to understand and use the English language is through a regular program of language study. They should also keep in mind that every section in a TOEFL test is integrated in terms of structure, vocabulary, and context. Awareness of these will lead to a higher TOEFL score than expected in no time at all.

## Selected Bibliography

**Cambridge Advanced Learner's Dictionary.** 2003. Cambridge: Cambridge University Press.

Chariya Pholprasert. 2005. **TOEFL PBT & CBT.** Bangkok: Sujitwat Limited.

Gibaldi, Joseph and Achtert, Walter S. 2003. **MLA Handbook of Writers of Research Papers.** 5<sup>th</sup> ed. New York: The Modern Language Association of America.

Test of English as a Foreign Language. 1998. **TOEFL Test Preparation Kit Workbook.** Princeton, N.J.: Educational Testing Service.

-----, 2003. **The Researcher.** Princeton, N.J.: Educational Testing Service.