CHAPTER III

RESEARCH METHODOLOGY

The research methodology here described was designed to compare the effects of the focus-on-form techniques integrated in the task-based instructions on the students' English learning achievement and their writing ability. This chapter presents the research design and instructional design, subjects, instructional treatments, research instruments, data collection, and data analyses.

3.1 Overview of the Experimental Design

This study was a *matching-only pretest-posttest comparison group design*. Six groups of students who took the pretest measuring English proficiency were randomly assigned to two treatment groups and one control group. The individuals in each group were matched according to their pretest scores and paired up with others who were judged to be at a similar score (± 2) in the other two groups. The design was as shown in Table 3.1 below:

Table 3.1: Research Design

Treatment group$_1$	O_1	M	X_1	O_2
Treatment group$_2$	O_1	M	X_2	O_2
Control group	O_1	M	X_3	O_2

The O_1 in this design refers to the pretest and the O_2 refers to the posttest. The M refers to the fact that the participants in each group will be matched. The X_1 is the FTBI experimental group, X_2 the TBI experimental group, and X_3 the CI control group.

The aim of this study was to examine the effects of the three types of instructions focusing on different types of grammatical treatment integrated in each type of instruction. The impact of three treatments – task-based instruction (TBI), form-and-task-based instruction (FTBI) and conventional instruction (CI) – on the English learning

achievement and the students' writing ability was compared quantitatively. The population was about 500 first-year diploma students at Rajamangala's University of Technology Suvarnabhumi Nonthaburi. The subjects were the first year technical students learning English as a foreign language (EFL) in Thailand. Initially, a total of 96 subjects, or 32 matched pairs selected for 3 classes, participated in this present study and the pretest scores of 96 subjects were analyzed to indicate the comparability among the groups. The post test scores of 92 subjects were analyzed to see the effects of the treatments because four subjects left the university. One pretest and one post test were used in order to examine the effects of the instructions. Three writing tests were given during the interval of five weeks between the pretest and posttest. The results of the writings were analyzed. The instructional treatments lasted 45 hours, including a three-hour pretest and a three-hour posttest. A questionnaire was provided right after the post test to investigate the students' attitude towards each instructional treatment. Figure 3.1 illustrates the experimental design of the present study.

Figure 3.1: Experimental and Instructional Design
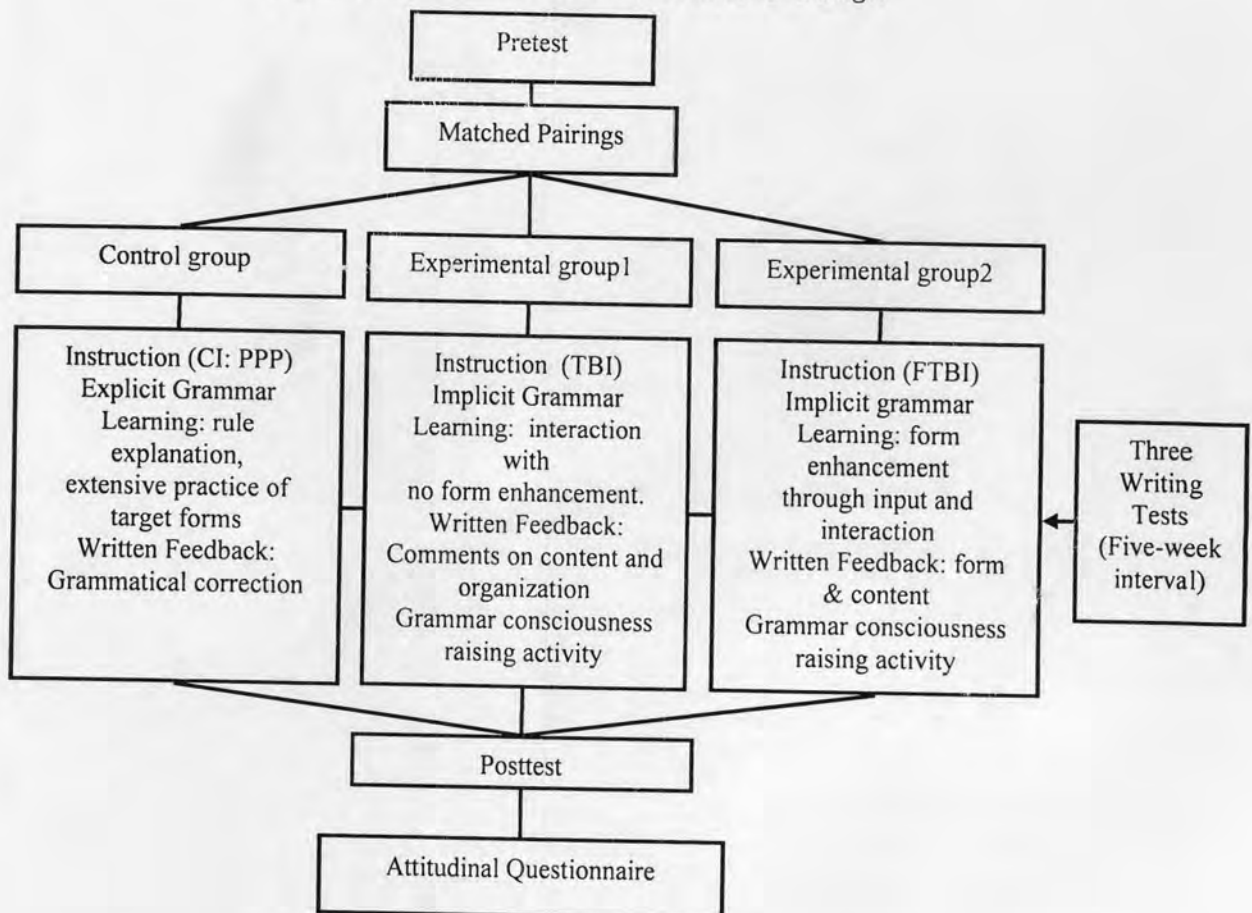
3.2 Population and Samples

3.2.1 Population

The population was 500 first year diploma students at Rajamangala's University of Technology Suvarnabhumi Nonthaburi (RMUTS Nonthaburi), attending the first fundamental English course, namely English 1. They were technical students who registered for this course, in the first academic year 2006. They were likely similar in many aspects as follows:

1) They were all technical students. To enter this university, they obtained their certificate in vocational and technical study. Due to their exclusively studies of their subject matters, they were not different. That is, they were interested in science and technology such as new modern computer technology; they liked to work with mechanical equipment or machines. Because of these similarity, either the opportunities to learn English provided by their previous educational system, if any, or the seek of the opportunities to get involved in English for themselves were very little.

2) The students' English proficiency was similar. The previous record of their achievement scores of the English 1 course was calculated using SPSS. Five groups of students (n= 149) were randomly selected. SPSS results rejected the null hypothesis that the distribution was normal, the Kolmogorov-Smirnov Z test = 1.637, Asymp. Sig. (2-tailed) = .009, $p < .05$ (See. Appendix D1).

3) All students had obtained the certificate level study. In their first year of study, they learned the same foundation subjects such as Thai, English, social science, physical education, and basic practice in the workshop. All students must take only one course of English – English 1, in the first semester. No other English course is provided for students in the first semester of the first year study at RMUTS Nonthaburi.

4) The school was a technical-based institution; support English study could not be equal to the subject matters related. Consequently, the institution supported very little concerning English. There was no English self-access learning center for the practice of English. The study of English was mostly from the teaching and learning in the classroom setting.

Therefore, it could be assumed that the population has similar characteristics though they might be slightly differed in affective factors, such as motivation, attention, interest, and learning in learning the English 1 course.

3.2.2 Sampling Method and Sample Size

The sampling method was cluster or group random selection. Random selection was performed to ensure the generalizability.

Since the students were generally assigned to the English classes according to their field of study, cluster random sampling, therefore, is the most appropriate method among all types of random sampling methods. It is a selection of sample in groups rather than individually sampling selection.

3.2.2.1 Steps of selection

The steps of selection were as follows:

1) The researcher informed the academic committees that two experimental and one control groups were to participated in this study, and that the class size would be 30 students or more. Due to the variety in numbers of the students in each class ranging from 20 to 35 students, six classes were to be sufficient for the selection so the academic committees randomly selected 6 classes using the lucky draw technique. The six classes were two classes of Automotive Technology, one Communication Technology class, one Sound system & Avionics class, one Electrical Power Technology class, and one Building & Construction Technology class.

2) The academic committees submitted the department tentative schedule to the central academic committees with the condition that six groups needed to be paired. Each pair must have the same date and time of the study. Unfortunately when the timetables for all students were completed, there were only two pairs, not three as expected. The timetable was:

Days	Times	No. of students	Classes
Tuesday	11.00 - 14.00	32	Automotive Technology
		23	Communication Technology
Thursday	8.00 - 11.00	31	Automotive Technology
		27	Sound system & Avionics
Thursday	14.00 - 17.00	33	Building & Construction Technology
Friday	8.00 - 11.00	32	Electrical Power

3) The students' English proficiency was measured by the pretest. The pretest was specifically designed and was already validated by the experts and piloted to find its reliability and validity. The pretest material contained three parts: 1) listening, 2) conversation, grammar& structure and reading, and 3) writing. Data collection was conducted in the same rooms (air-conditioned rooms for speaking and listening, well-ventilated room for reading and writing). The time for the pretest was the same as appeared in their timetable. All students took all paper-based skill tests within the time provided; for speaking test, some students made the appointments with the interviewer (the present researcher) with the time available for them. All interviewed were audio-recorded and rated by three raters (two were the English teachers who were trained how to rate the interview with the speaking test rubric, the other was the present researcher)

4) The scores of speaking test, reading test, listening test and writing test were calculated. The group means scores were compared on the basis of the total score of the speaking, listening, reading, writing abilities and linguistic ability. Although this study would like to see three dependent variables: the learning achievement, the grammatical accuracy and writing ability, the scores of the pretest were not separately compared. If the groups do not differ significantly in terms of their general English ability, they would be equivalent with regards to the grammatical accuracy and the writing ability.

5) The individuals in each group were matched and paired up with other students who were judged to be at a similar score (± 2). The level of the students' English ability was used as the factor to determine the comparability. To make the three groups comparable, the technique of *mechanical matching* (Fraenkel and Wallen, 2000: 294) was performed. One individual from one group whose pretest score was similar to other individuals from other two groups was matched to make one matched pair. It was

inevitable that some students in each class were eliminated from being the subjects in the study when 1) their scores could not be matched, and 2) their scores could be matched but no "matchee" left for them. Ninety-six students were pair-matched. By this matching technique, the three groups was equivalent in terms of the size (32) and means score. The means scores of group 1, 2 and 3 were 30.489, 30.498, and 30.514 accordingly. The group variances were 40.857, 39.915, and 41.742 accordingly (See. Appendix D2).

6) The three groups of students were randomly assigned to two separate experimental groups: Experimental group 1 (Task-based instruction or TBI), Experimental group 2 (Form-and-task-based instruction or FTBI). The contrast treatment was given to a contrast group (CI).

7) It could be, therefore, taken that the groups were equal in terms of statistical concerns – group means and variance. However, to increase confidence that they were really statistically comparable, the data of the subjects were analyzed. A one-way analysis of variance (ANOVA) was run after testing the normality of the subjects. No significant difference among the groups was revealed, $F(2, 93) = 0.00$, $p = 1.00$. It could be inferred confidently that any difference that might occur would be from the treatment only.

3.2.2.2 Sample size

Due the number of population, the sufficient samples should be about 217 (Krejcie and Morgan, 1970). This study is a classroom-based research study; a class of over 30 students is considered sufficient (Fraenkel and Wallen, 2000:118).

The samples resulting from the mechanical matching were 32 pairs. Therefore, each class contained 32 students, that is, the subjects in this study were 96. However, during the study 4 students left the university. The number of subjects used in data analyses was 92.

This study had variations in the numbers of the subjects used for data collection. For research questions 1 and 5 relating to the quantitative research method, all subjects were the target samples whereas for research questions 2, 3, and 4 relating to the careful analyses of the students' writing ability and the grammatical accuracy of the written production, the samples were purposively selected from each class. The

selection was based on their language ability: three high, four average, and three low students of one class. To select, the pretest scores of one class were calculated to find the percentile rank using SPSS. The first quartile was at 26, the second 29 and the third 34. The ones who were below the percentile of 25 (or the score of 26) were marked as low students. The ones who were above the percentile of 75 (or the score of 34) were marked as the high group. After grouping, each group was randomly selected: 3 from high, 4 from average, and 3 from low group of students. After the one-class selection, the matched pairs from the other two classes were also selected.

3.3 Research Instruments

There were two types of instruments: instructional instruments and research instrument. Each will be discussed in turn.

3.3.1 Instructional treatments

There were three instructional treatments: conventional instruction, task-based instruction, form-and-task-based instruction. Next section describes each teaching method.

3.3.1.1 Conventional instruction

Figure 3.2 illustrates the conventional instruction which is a structure-oriented approach to tasks. According to Skehan (1998), this structure-oriented approach places its centrality to specified forms and allows planning and systematic treatment to specific language structures.

This approach contained Presentation, Practice and Production. At the presentation stage, a target language was presented either explicitly or implicitly. The aim was to maximize the chances that the underlying rule would be known and internalized, and at this stage the focus was on the declarative knowledge development. At the practice stage, the exercises that provided ready made meaning were practiced straightforwardly. At the production stage, learners produced language based on meaning they wanted to express. When the students worked in groups, the teacher let them to take their time – teacher intervention was purposively ignored. When they

finished the task, they were asked to present their work orally, and submit the written version. With the written production, the teacher corrected the grammatical errors and returned to them. Then, the students analyzed the correction and asked any points of language they were not sure of. The teacher explained and gave remedial exercises as necessary. After checking their understanding and making sure that the students had no other point to make, the students were asked to correct and re-write their work and submit it (See Appendix E1).

Figure 3. 2: Conventional Instruction (CI)
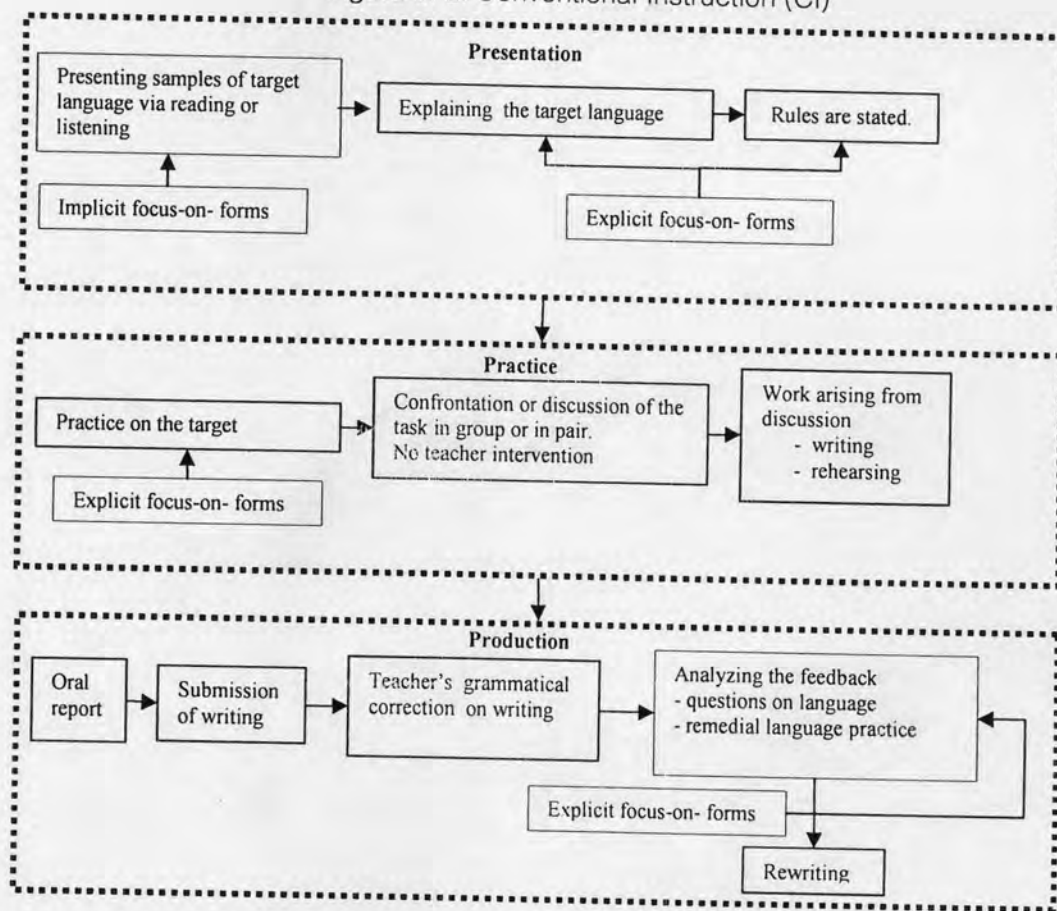
3.3.1.2 Task-based instruction

The framework of Willis's TBL was applied in this present study since it allows variety in the choice of options in each phase. To apply Willis's (1998) framework of task-based instruction, the requirement of the course objectives involves development of the four language skills – listening, speaking, reading, and writing. Comparing them with those of Willis, the writing production is an optional. In this study, writing had to be an essential component in the framework. So, at the post-task stage

the students were required to present the outcome of the task in written form as an end product of the task after they had presented the outcome orally. It was a promising teaching strategy because the framework encouraged a recursive process of writing. At the stage of during task, the student made a draft with teacher assistance, revised and edited focusing on content and organization before involving in the language focus stage which was grammar consciousness-raising activities. The students could revise and edit again before they produced a complete version by re-writing at the end stage of the task-based learning (See. Appendix E2).

1) The pre-task phase: Figure 3.3 illustrates the pre-task stage. The teacher introduced and defined the topic and task. To activate words and phrases which related to the topic, the students were told what they were required to do and what the nature of the outcome was (Lee, 2000). Also, the teacher could activate the students' content schemata of the given topic by providing activities, e.g., brain storming, mind-mapping (Willis, 1996). And to scaffold the students' performance to the task so that they were conceptually and semantically prepared, the teacher could provide the students with a set of activities, being as a path to the main task performance (Prabhu, 1987), a demonstration of the task in order to decrease the cognitive load on the students (Skehan, 1996; and Willis, 1996); conversational strategies that the students would become adaptable, creative inventive and above all independent (Nunan (1989:81). Before the students got involved in the during task stage, the students were given with planning time. This led to fluency and complexity of the language (Skehan, 1996).

Figure 3.3: Pre-task Phase and Instructional Options

Pre-task Stage	Examples of Instructional Options
Introducing topic and defining task	Providing the structure of the task Establishing the outcome of the task
Content and Task preparation	Performing a similar task A task demonstration Vocabulary-based activities (e.g. brain storming, mind mapping) Providing conversational strategies
Planning time	Providing strategic planning of the task

2) The during task phase: Figure 3.4 illustrates the during-task phase having three main components: task discussion, planning, and report. At this stage, students began performing the task which might be based on input texts – a reading or listening text. The input would shape the students knowledge and content to the production. The students worked together in order to produce their own outcome of the task. During group working, the teacher monitored and assisted them primarily with content and organization of ideas. When they finished their task, they prepared for the report. The report was to provide students with a *natural stimulus to upgrade and improve* their language production (p: 55). In planning their report, they collaboratively prepared a report with the assistance of the more able peers, the teacher, dictionaries and grammar books. They wrote and rehearsed before presenting their outcome of the task. Actually, students started writing when they knew task instruction and task goal, that is, they had purpose to write. They knew what they would write about, why they wrote and who they wrote to. During the discussion, they might make notes. After the discussion, they corporately produced the written production which was then circulated and displayed for other students to read. Then the teacher collected the written work and gave comments focusing on content and organization. The figure 3.3 below illustrates the during-task phase.

80

Figure 3.4: The During-task Phase and Instructional Options

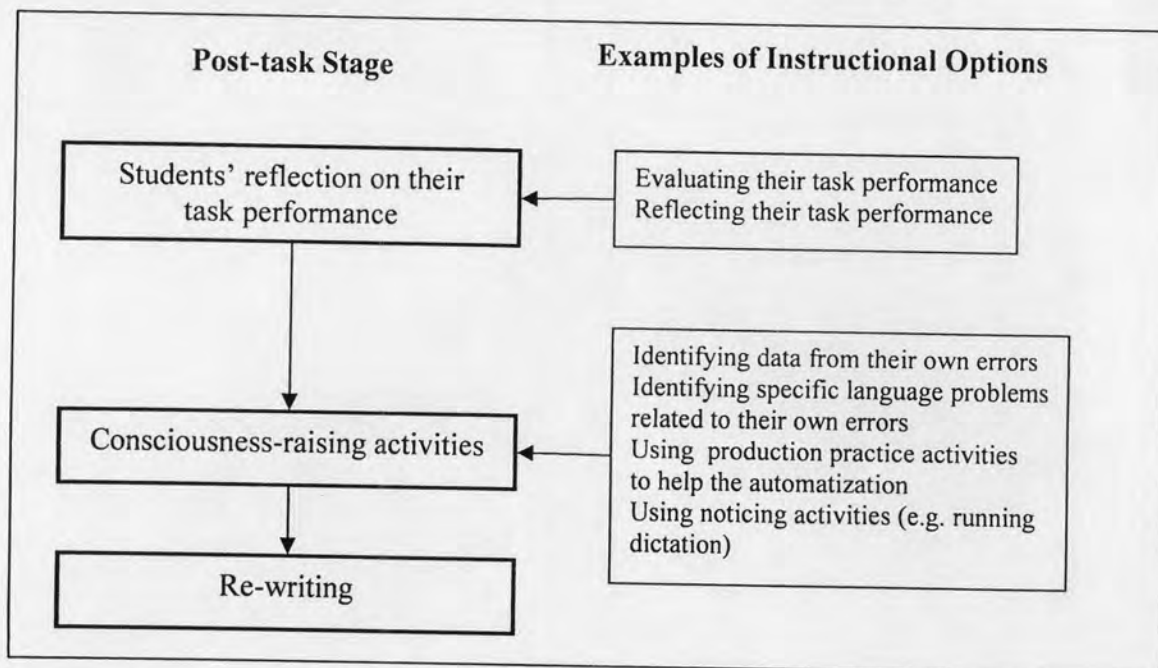| During-task Stage	Examples of Instructional Options
Discussion of the task with teacher's assistance on content	Giving unlimited time to perform the task Providing text-based input Introducing useful information
Work arising from the discussion	First draft writing
Setting up the planning (e.g. explaining the purpose, form of the report)	Encouraging self and peer-correction Consulting dictionaries
Reporting an outcome of the task	Oral presentation Written report

3) The post task phase (See. Figure 3.5). This is also called *language focus*. There are three main goals: to provide an opportunity for formal language instruction; to encourage reflection on how the task was performed; and to practice to the linguistic forms, particularly, to those that found problematic to the students when they had performed the task.

First was the task reflection. It was on any incidence either language or content as the students would like to share, and on the feedback the teacher gave on their writing which was mainly on content and organization. This was beneficial to the students in many ways: the students might recognize their particular language elements they used, leading to deeper understanding of their meaning and use; the students might remember words, phrases, structures that were useful for them; and they would use the language with more confidence.

After the students' reflection on their task performance, they were provided with activities which were consciousness-raising activities that focused on the most problematic forms and use that the teacher found in their performance. After this, they involved in the process of subsequent revision, they rewrote and submitted their successive writing.

Figure 3.5: The Post-task Phase and Instructional Options

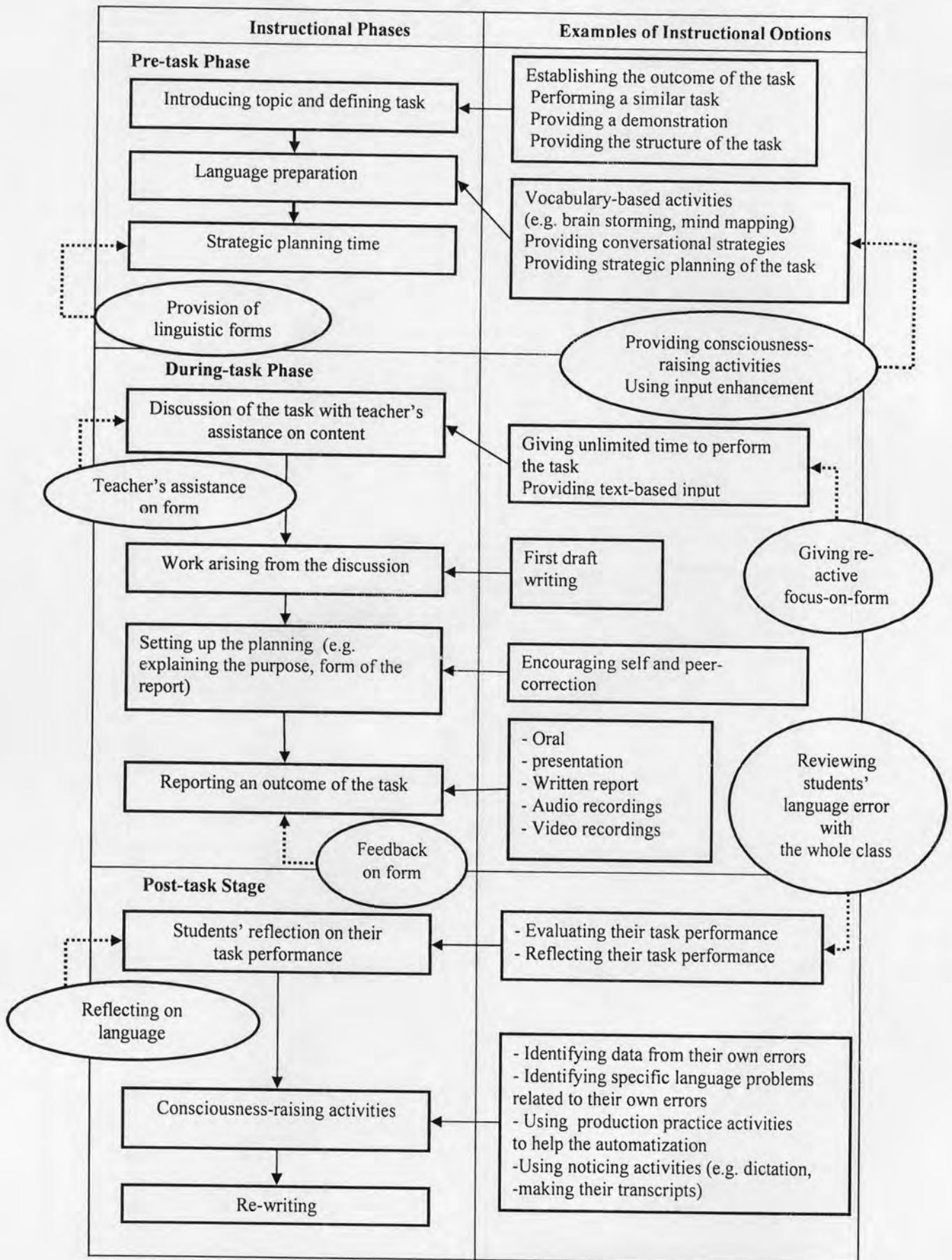| Post-task Stage	Examples of Instructional Options
Students' reflection on their task performance ←	Evaluating their task performance Reflecting their task performance
Consciousness-raising activities ←	Identifying data from their own errors Identifying specific language problems related to their own errors Using production practice activities to help the automatization Using noticing activities (e.g. running dictation)
Re-writing	

3.3.1.3 Form-and-task-based Instruction

The form-and-task-based instruction was an integration of the task-based instruction and the focus-on-form. The integration occurred in all three phases (See. Figure 3.6). The form-and-task-based instruction (FTBI) utilized the TBI framework, excepted that the FTBI involved the input enhancement in the pre-task phase, the reactive focus-on-form activities in the during-task phase, and the written form feedback into the post-task phase.

Regarding writing teaching techniques, the FTBI did two more additionally essential things. First, during the draft making, teacher gave feedback on forms. Second, for the writing paper the teacher marked the students' grammatical errors at the sentential level – graphological form, lexical form, morphosyntactic form – and the discourse level – cohesive form, information management form e.g. word order, parallelism by using correction codes, tapping with comments on the problem areas (See. Appendix E3).

Figure 3.6:   Integration of Focus-on-Form into Task-based Instruction

Instructional Phases	Examples of Instructional Options

Pre-task Phase

Introducing topic and defining task

Language preparation

Strategic planning time

Provision of linguistic forms

Establishing the outcome of the task
Performing a similar task
Providing a demonstration
Providing the structure of the task

Vocabulary-based activities
(e.g. brain storming, mind mapping)
Providing conversational strategies
Providing strategic planning of the task

*Providing consciousness-raising activities
Using input enhancement*

During-task Phase

Discussion of the task with teacher's assistance on content

Teacher's assistance on form

Work arising from the discussion

Setting up the planning (e.g. explaining the purpose, form of the report)

Reporting an outcome of the task

Feedback on form

Giving unlimited time to perform the task
Providing text-based input

Giving re-active focus-on-form

First draft writing

Encouraging self and peer-correction

- Oral
- presentation
- Written report
- Audio recordings
- Video recordings

Reviewing students' language error with the whole class

Post-task Stage

Students' reflection on their task performance

Reflecting on language

Consciousness-raising activities

Re-writing

- Evaluating their task performance
- Reflecting their task performance

- Identifying data from their own errors
- Identifying specific language problems related to their own errors
- Using production practice activities to help the automatization
- Using noticing activities (e.g. dictation, -making their transcripts)

3.3.2 The teaching materials

The steps of producing the teaching materials were:

1) Collecting the existing materials used by the Department of English during 2001-2005. They were English I (produced by the English Department, Faculty of Liberal Arts, RMUT), English I: self-study course books and video CD (This set of materials was a project of material design for all RMUT students in Thailand sponsored by World Bank), Dynamic English 1 with video CD, Breakthrough I and II with audio CD (Long and Richards, 1978). This was to see the level of language and content which was used as a basis for the design of the materials. Also, the published teaching materials that were designed for the beginner and pre-intermediate learners were investigated to find out the content or exercises that best fit the course objectives. The selected materials were Let's talk I, II and III (Jones, 2002), Focus on Grammar (Fuchs, Bonner, M., and Westheimer, 2000)

2) Producing the tasks. This study followed the course objectives and communicative-oriented task-based instruction. In consequence, the tasks were collaborative type and designed mainly having the feature of task essentialness – a task being completed with specific target language (Loschky and Bley-Vroman, 1993). Loschky and Bley-Vroman mentioned that the task essentialness would enhance the formation of hypothesis, and prepare learners to realize and learn from the feedback which occurs during the course of interaction. Eight main tasks were designed and implemented at the end of each learning unit. The tasks selected were conversation creation, describing pictures, searching and telling, and problem solving tasks.

3) Producing the consciousness-raising activities. The consciousness-raising activities were the integrated-skills. The input were either listening or reading and the students write or speak as the production.

4) Constructing the language exercises. The language exercises had two main purposes. First, the students in the CI class did the language exercises as the preparatory session before doing the tasks. And the students in the FTBI and TBI classes did the consciousness-raising activities after they performed in the task and did the language reflection.

5) Making the planning time activities

6) Writing a prototype chapter and consulting the advisor.

7) Completing the material writing (See. Appendix E4) and asking three experts in the field of teaching and learning English to evaluate the content of the material, the level of difficulty, the level of the language, the appropriateness and the correctness of language. The checklist and the framework of each type of instruction as well as the course objectives were provided. The result of the evaluation was calculated using the Item-Objective Congruence Index (IOC). The value of each item was 1.00 (See. Appendix E5).

3.3.3 Research Instruments

The research instruments implemented in this study were a pretest and a posttest, three writing tests, and a set of questionnaire. The pretest and posttest were used to measure the students' English learning achievement. The writing tests were used to measure the students' ability to write and their accuracy of the written production. The questionnaire was used to measure the students' attitudes towards the teaching and learning. Each will be discussed in turn.

3.3.3.1 Pretest and posttest

The pretest and posttest were used to measure the effect of the three teaching methods on the learning achievement of the subjects. The pretest and posttest were the same version because the length of experiment time was seventeen weeks which is sufficient to prevent the pretest-posttest effect.

3.3.3.1.1 Test Construction

To construct the test, the process started with designing the components of the test in order to be confident with the corresponding of the test content and the construct it was intended to measure, and with the test scores that could represent the intended objectives. Following Bachman and Palmer (1996), the process involved the following steps: 1) identifying the objectives of the test, the target language, and the task types; 2) defining the characteristics of the test takers; 3) defining the test construct or ability to be measured; 4) planning for the evaluation of the test usefulness: reliability,

validity, construct validity, authenticity, interactiveness, impact, and practicality; and 5) planning for the test management.

At the step of making a test specification and a blueprint (See. Appendix F1), three main tasks were: 1) developing the test tasks and a blueprint; 2) writing the test instructions; and 3) identifying the scoring methods.

The test (See. Appendix F2.1) aimed to cover the main objectives of the course. The design was based on the students' language performance, that is, how well the students could use their abilities and knowledge of language to perform each type of test. Four skills, then, were separately tested. The components of the test were:

1) The listening test. It consisted of 30 multiple choice items (See. Appendix F2.2).

2) The speaking test. It contained 2 portions: a test of components of the language and a test of speaking performance. It contained 15 items and 1 interview. The form of this part was blank-filling with multiple choices and the interview with focus questions. The interview was based on the following reasons:

- Task familiarity: familiarity with a task could reduce the transformations that a speaker would be required to.

- The material could be structured in a way which is comprehensible to the speaker.

- The questions could be designed to meet the inference about the students' abilities. Each question would give a new start with different topics during the interview.

- The interview allowed for interaction between the students and the examiner.

3) The reading test. It was arranged from less to more difficult. It contains three parts:

- *vocabulary* The reading started with sentence level reading, followed by the text level. The texts were two unrelated. One was the extract from text book and the other was the extract from magazine. The first text was followed by four-option multiple-choice lexical options. The

second text was followed by four-option multiple-choice lexical options.

- *grammar* One text was drawn from texts (both fiction and non-fiction) and instructional manual. Students were required to find out which function words of the text have been removed.

- *main idea and specific details* This consisted of three short texts followed by four-option multiple-choice options. They were drawn from articles and texts (both fiction and non-fiction), leaflets, advertisement or instructional manual. There were five four-option multiple choice questions testing content and understanding of text organization.

4) The writing test. This test appeared in the forms of sentence writing and composition writing. There were 10 sentences to write from the clues given for each sentence. The students were to write two compositions from the situational context given either by pictures. To mark the composition writings, three analytic scoring rubrics (See. Appendix F2.2) were used. The criteria of deducing the points were:

- If the content is wrong, no credit is given to the item.

- If the students do not write a complete sentence, deduct 0.5.

- Deduct 0.25 for each error (i.e., tense, voice, subject-verb agreement, part of speech, word order, sentence structure)

- Deduct 0.25 for each error in punctuation (i.e., comma, question mark, period)

3.3.3.1.2 Test Validity and Reliability

Before validating the test, the test specification and the complete set of test were approved by the research advisor. Then, the content and language of test was edited by an English native speaker teaching at KMUT North Bangkok. Then, the tests and rubrics (See. Appendix F5) were examined by three experts. To consult the experts, a checklist was provided together with the purposes of the test, definitions of the domains, and test specification. The checklist contained the objectives of the test, the item and section arrangement, and the relation of the test items to each objective (See. Appendix F3.1).

After the judgment was made by the experts, the content validity was calculated using Item-Objective Congruence Index (IOC). There were two steps of finding the content validity. First, the evaluation was calculated. The total value of IOC was 0.84. Second, the test items that have the IOC less than 0.75 were corrected based on the comments of the experts and asked the experts for the judgment again. When the experts agreed upon the correction, they were re-calculated and the result of the re-calculation was 1.00 (See. Appendix F3.2).

Due to the types of the test: objective test items and subjective test items, they were separately analyzed. The objective test was piloted with 100 RMUTS first-year technical students who had just completed English I. However, due to the incomplete marking on the answer sheets, data of the objective test from 76 students were analyzed.

For the objective test, its reliability was estimated by measuring its internal consistency. Firstly, the data were analyzed using Sukamolson's (1995) computer program of CTIA/Grading (Classical Test Item Analysis and Grading) to find the discrimination and difficulty value of each item. The acceptable discrimination value of each item was set at 0.20^{+} (Ebel, 1979, Gronlund, 1986, and Hudson, 1973). The difficulty index of each item was set between 0.20-0.80 (Sukamolson, 1995). The analysis showed that their difficulty and discrimination value of fifteen items did not reach the acceptable value. As a consequence, they were deleted and some alternatives with low value of either discrimination or difficulty were rewritten. Then, the new version of the test, containing 85 items were tried out. 174 pieces of data were analyzed using Sukamolson's CTIA. It was found out that the reliability estimated by Kuder-Richardson (KR21) was 0.806, the difficulty index was 0.376, and the discrimination index was 0.282 (See. Appendix F4).

To find the reliability of four subjective test items, the data of 44 students were analyzed after the test paper was rated following the writing rubric by three raters. Two were the English teachers and the other was the researcher. Interrater reliability as measured by Pearson run by SPSS was acceptable: correlation between raters 1 and 2 was 0.908; correlation between rater 2 and 3 was .967; and correlation between raters 3 and 1 was .956. The reliability of the test estimated by Cronbach's Alpha Coefficient was

run by SPSS/PC Version 12.0. The result of analysis showed that reliability was 0.972 which was high, indicating that the test was good in terms of the reliability. The below table shows the discrimination and difficulty values estimated by Scannell and Tracy's formulas (1975, cited in Sukamolson, 2005). The statistical values showed that the test was good with respect to its discrimination and difficulty. The discrimination indexes (Idisc) were more than 0.20^+ (Ebel, 1979, Gronlund, 1986, and Hudson, 1973). The difficulty indexes (Idiff) were between 0.20-0.80 (Sukamolson, 1995).

Subjective Tests	IDiff	IDisc
1. Sentence writing	0.346	0.498
2. Composition writing 1	0.458	0.661
3. Composition writing 2	0.409	0.657
4. Oral interview	0.436	0.536

3.3.3.2 The writing tests

The writing tests were used to see the impact of the teaching methods on the students' accuracy of written production and the students' writing ability. The tests were administered with an interval of five weeks between the pretest and posttest.

3.3.3.2.1 Test construction

There were three writing tests. Each was designed based on the course objectives and real life situations in which the students were likely to be involved and in which inferences about the students' language and writing abilities are intended to generalize. Therefore, the writing tests focused on language, content and organization of the writing production.

The process of construction started with the test blueprint and specification (See. Appendix G1). Three tests then were designed together with three writing rubric specifically and measurably designed for each test. Three tests were describing people, giving an instruction, describing past events. After the design of the tests, the situations (i.e., giving instruction and describing past events) were drawn into two series of events (See. Appendix G2.1)

The tests were then evaluated by three experts. The checklist, the test specification, the writing rubric (See. Appendix G3.1) was also provided. Interrater reliability as measured by Pearson run by SPSS was acceptable: correlation between raters 1 and 2 was 0.908; correlation between rater 2 and 3 was 0.967; and correlation between raters 3 and 1 was 0.956. The results of the evaluation were estimated by the IOC and found out that the value was 1.00. The comments from the experts were used to improve the quality of the tests and test rubrics (See. Appendix G3.2).

3.3.3.2.2 Construction of the Scoring Rubrics

The steps of constructing the scoring rubrics were:

1) Developing the framework of assessment. The questions of the writing tests and the speaking test were investigated to find the criteria which must be aligned with the task requirements and the test construct

2) Developing the scoring rubrics for writing tests and speaking tests. For writing test, there were major components to be assessed such as content and ideas, organization, voice, vocabulary, cohesive and coherence, and grammar. For speaking test, there were essential features to look into such as pronunciation, grammar, fluency, task completion. Since the measurement was very specific, the scoring rubric must be specific. Therefore, the scoring rubrics in this study needed to be analytical since there were specific components of the writing ability and speaking ability to be assessed.

3) Investigating students' level of speaking and writing ability. This was very important since the scales must be able to measure the students' level of proficiency. If not, the rubrics could not rate the students' ability because the scale points would be unable to rate the high and the low ability students. To do this, the researcher consulted the International Second Language Proficiency Ratings (Wylie and Ingram, 1999).

4) Writing the descriptor based on the students' level of proficiency and numbering the scale points for scoring student response.

5) Consulting the advisor, and the experts.

6) Piloting the scoring rubrics.

7) Changing any unclear points caused by the language and the overlapped criteria in which the raters were unable to judge the student response.

8) Training the raters before implementing the scoring rubrics.

3.3.3.2.3 Test validity and reliability

The three tests were tried out with 30 students and rated by three raters. Then, the reliability estimated by Cronbach's Alpha Coefficient was run by SPSS/PC Version 12.0. The result of analysis showed that reliability was 0.89 which was high, indicating that the reliability was acceptable. The below table shows the difficulty and discrimination of the tests 1, 2, and 3 estimated by Scannell and Tracy's formulas (1975, as cited in Sukamolson, 2005).

Writing tests	IDiff	IDisc
1. Describing people	0.592	0.497
2. Giving an instruction	0.517	0.476
3. Describing past events	0.486	0.674

The difficulty index (IDiff) and the discrimination index (IDisc) indicated that the tests were acceptable. The discrimination values were more than 0.20^{+} which showed that the tests could discriminate different ability groups (Ebel, 1979, Gronlund, 1986, and Hudson, 1973). The difficulty indexes were between 0.20 - 0.80 (Sukamolson, 1995), indicating that the tests were not too difficult nor too easy.

3.3.3.3 Questionnaire: Semantic Differential Scale

The semantic differential (SD) is a procedure for observing and assessing affective responses (Kerlinger, 1973). The SD measures people's reactions to stimulus words and concepts in terms of ratings on bipolar scales defined with contrasting adjectives at each end (Heise, 1970). Osgood, Suci, and Tannenbaum (1969: 58) believed that language was an index of meaning and *selection among successive pairs of common verbal opposites should gradually isolate the meaning of the stimulus.* Accordingly, a bipolar adjectival scale was created. An example of an SD scale is:

Fast   _   _   _   _   _   _   _   Slow
         3   2   1   0   1   2   3
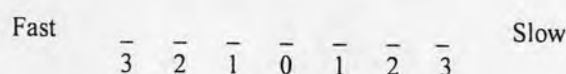
Usually, the position marked 0 is labeled 'neutral", the 1 positions are labeled "slightly", the 2 positions "quite", and the 3 positions "extremely". Osgood et al. mentioned that this

type of scale allowed both the directional measurement (e.g. fast versus slow) and also intensified measurement (e.g. slight through extreme).

In their book *The Measurement of Meaning*, Osgood, Suci, and Tannenbaum (1957) mentioned that three semantic dimensions: evaluative, potency and activity were essential variables within the SD. They reported similar results of three studies on factor analysis that the evaluative factor was the most identifiable, followed by the potency dimension and activity dimension. Carroll (1969:111-2) has affirmed that these three *principal* SD dimension are fundamentally necessary for an individual to make judgment on things in his environment, and thus claimed that: 1) the evaluative dimension informs the individual's tendency to accept the stimulus or to avoid it; 2) the activity dimension provides the answer to the necessity or non-necessity of adjusting the stimuli; and 3) the potency dimension suggested the amount of adjustment that is made to a stimulus or the amount of effort somebody put into the response of the stimuli. Based on the results of studies by Osgood et al. and the claim made by Carroll, the semantic differential scales are to contain the scales related to the evaluative, activity and potency dimensions.

The technique has three features that distinguish it as an instrument for social psychological research. First, the SD is easy to set up, administer, and code. This gives it favorable cost-effectiveness. Second, the SD contains three dimensions: evaluative; potency; activity (EAP). The EPA measurements yield rich information about affective responses to a stimulus. The information that the EAP gives is lost with alternative measures. Third, the form of an SD is basically the same whatever the stimulus, so research using the SD would be very cumulative.

In reviewing the variety of studies implemented the SD (e.g. Barclay and Thumin, 1963; Beardslee and O'Dowd, 1961; Gusfield and Schwartz, 1963; Jakobovits, 1966; Osgood, 1964; Osgood et al., 1957; Prothro and Keehn, 1957; Rodefeld, 1967; Triandis, 1959; and Williams, 1964; 1966) and a number of studies focusing their work on the correlation of three-dimensional bipolar adjective ratings (e.g. DiVesta,1966; Heise,1965; Osgood, *et al.*, 1957; Suci, G.J.,1960; and Wright, 1958), Heise (1970) concluded that the SD was a valid technique for the attitude measurement and that ratings on bipolar adjective scales were correlated. In comparison with other measurement scales, Osgood (1975) reported his findings that the SD highly correlated

to Thurstone (r = .74 - .82) and to Guttman (r = .78). Saiyod (1971, cited in Boonchuduang 1983) reported that the correlation coefficient value between Likert, and SD was .73. It can be concluded that the demonstrated validity of SD, the high correlation of ratings of SD scales as well as the high correlation of the SD with other scales of measurement give it a theoretically and empirically supported measure of attitude.

The questionnaire was used to survey the students' attitudes towards the teaching methods. It was administered right after the students had finished their posttest since the learners' sense of achievement was evaluated.

3.3.3.3.1 Questionnaire construction

The questionnaire in this study was a semantic differential scale. The steps of the construction were as follows:

1) Collecting the information as a basis for the design of adjective pairs. One class of first-year diploma students – automotive technology students – participated in this activity. The students were asked a series of questions orally and answer on the provided sheet of paper. The questions which were designed to cover the components of each teaching instruction can be listed below:

What are three words most representing your reaction towards .....

-learning English

-English teaching

-pair work

-group work

-reporting in front of the class

-talking in English with your peer

-talking in English with your teacher

-writing activities

-reading activities

-listening activities

-teacher's feedback on your writing

-preparation before working in group

-exhibiting your writing to your classmates

-reflecting on your language errors

-re-writing your work

2) Categorizing the data. The words expressed for each question were tallied to find the top ten most frequent expressed. The selected words were categorized into 3 dimensions: evaluation (e.g. good, useful), potency (e.g. confident, successful, difficult), and activity (e.g. boring, joyful, relaxed). The samples were from a group having similar characteristics of the population, and there was no attempt to add the theory-driven words. As a consequence, the portion of words belonging to each dimension for each question that was not equal.

3) Finding a contrasting adjective. *The Roget's Thesaurus* was used as a tool to find a contrasting adjective for each selected word. After this, the questionnaire was written following the format characteristics of semantic differential scale. It consisted of a title of the investigation, instructions (i.e., greetings and telling the purpose of the study, demonstration how respondents should answer the questions), questionnaire items, and final thank you.

4) Writing the questionnaire. The questionnaire was designed in Thai version in order to prevent any confusion or ambiguous information resulting from the meaning of the English language. Due to the polar ends of each pair of adjectives, to prevent the students from superficial responding, the position of negative pole and positive pole was variedly sequenced – the positive pole was alternated between being on the right and the left side. And to provide the logical data processing, each question ended with the pair of adjectives representing the absolutely evaluative dimension (e.g. good-bad). The length of this questionnaire was 7 pages long (See. Appendix H1). Three students were asked to complete this questionnaire. It was found that the time for the completion was between 10-15 minutes.

3.3.3.3.2 The validity and reliability of the questionnaire

The complete version of questionnaire was initially investigated by the research advisor. Sensitive words were the topic of discussion. After the compromise agreement had been reached, a checklist for evaluating the questionnaire was constructed. The checklist contained the objectives of the test, the item and section

arrangement, and the relation of the test items to each objective (See. Appendix H2.1). Three experts were asked to validate it. The data from the experts were calculated by Item-Objective Congruence Index (IOC). Analysis showed that all of the experts agreed upon the content of every item (IOC= 1.00).

Ninety-nine students who were attending their study of English I in week 10 were asked to answer the questionnaire. The time to the completion the questionnaire did not exceed 20 minutes. The data was run by SPSS estimated by Cronbach's Alpha Coefficient. It was found that the reliability was 0.977.

3.3.4 Data collection

This research study was conducted the Department of English, Faculty of Applied Liberal Arts, RMUT Suvarnaphumi, Nonthaburi, Thailand. The participants were 92 diploma students who were just entered and studying in their first year at this university. The participants were selected into three classrooms which are air-conditioned so that all the three groups of students learned under the same classroom atmosphere; using matched pairing technique based on their pretest score. The research instruments in this study were the pretest, the post test, three writing tests, and one set of attitudinal questionnaire. The duration of the study lasted 17 weeks which equaled one semester. The instructional period was 15 weeks. The other two weeks were for mid-term examination and final examination.

The data collection from the three groups was structured as shown below:

Week	TBI class	FTBI class	CI class
1-2	PRETEST		
	Unit 1: Greeting and introducing		
3-4	Unit 2: Request and permission		
5	Unit 3: Apologies and thanks		
	WRITING TEST 1		
6-7	Unit 4: Instructions		
8	Review		
9	MID-TERM EXAMINATION		
10-11	Unit 5: Describing objects		
	WRITING TEST 2		
12-13	Unit 6: Describing present events		
14-15	Unit 7: Describing past events		
	WRITING TEST 3		
16	Unit 8: Describing future events		
17	POSTTEST AND QUESTIONNAIRE		

All data were collected during the first semester academic year 2006. The data were collected from the subjects of the study as well as from the teacher. The subjects from the three classes were the present students attending the course "English 1". The teacher was the present researcher who was assigned full responsibility for teaching and evaluating all 96 students. After the participants feel relaxed, the questionnaire was administered. The reason for this is that the learner self-achievement is a part of the study.

Week 1

This week was an orientation and a pre-testing measure. The orientation and the pre-testing measure were presented to the six classes on the scheduled timetable. The researcher explained to the students the pedagogical goal for the research – to develop a new teaching technique for teaching English, emphasizing that this study was an actual learning and teaching and that the researcher would take full responsibility of the class as an actual teacher. The researcher also briefly described the experiment, and what they needed to do in order to participate in the study. The researcher also informed them that not all students would participate in this study, but some would be selected

based on their score that could be matched up with the other scores, the rest would learn with the other three teachers who also took full responsibility to the teaching and learning process. All students were asked if they were willing to take part in this study. They all agreed to participate so the pretest began.

Week 2 to week 16

The instructional treatments took place on the following times and dates in a week:

Instructional treatments	Date	Time
FTBI	Tuesday	11.00 - 14.00
TBI	Thursday	08.00 - 11.00
CI	Thursday	14.00 - 17.00

The instructional treatments followed the teaching framework of each type of instruction. Right after each teaching, the researcher recorded the teaching and learning events in her diaries. After completing each unit, the students answered the retrospective questionnaire. In week five, ten, and fifteen, the writing tests were administered. Three raters graded the paper and produce the measured score for each subject.

Week 17

The post testing measure took place on week 17 (September 25 [th] – September 29th, 2006). Since the students had to sit in for the Final Examination which scheduled by the university, the fixed time, date and place made it impossible to arrange the papar-pencil based tests (listening, semi-speaking, reading and writing) at the same time as the pretest arrangement. The schedule of the posttest for all participants in this study was arranged as the following times and dates:

Types of test	Date	Time
Semi-speaking, Reading, Writing	September 26th, 2006	13.00-15.00
Listening	September28[th], 2006	12.00-12.30

The oral interview occurred on different date and time. Each group of students made appointment with the researcher according to their availability. The time and date for the interview can be shown below:

Classes	Students	Date	Time
FTBI	Automotive Technology(1/1.1, 1/1.2)	September 27th, 2006	12.00
	Communication Technology (1/5A)	September 27th, 2006	16.00
TBI	Automotive Technology (1.1.3, 1/1.4)	September 29th, 2006	12.00
	Sound system & Avionics (1/5B)	September 29th, 2006	9.00
CI	Building & Construction Technology (1/6)	September 25th, 2006	9.00

The questionnaire was also implemented in this week after the paper-pencil based test on September 26th, 2006. The questionnaire was provided with the test. The students were informed before the test started that they could complete the questionnaire after they had finished the test.

3.3.5 Data Analyses

3.3.5.1 Analyses of English Learning Achievement

The data obtained from the posttest were used to compare the effects of the TBI, the FTBI and the CI on the students' English learning achievement and their effect sizes. The posttest scores of TBI, FTBI and CI were first calculated by Shapiro-Wilk test which was carried out on a computer using SPSS software (version 11.5 for window) to find the normality and variance of the variables. It was found that there was a significant difference ($df = 31$, $p = .041$) in TBI scores. Consequently, selection of a test focused on non-parametric tests (See. Appendix I1).

This study involved teaching methodology with three methods of teaching, each method was tested using independent samples, so the Kruskal-Wallis H test (as cited in Heiman, 2006) was performed. The Kruskal-Wallis H test was carried out on a computer using SPSS software (version 11.5 for window). The result revealed that there was no significant difference among the three variables, so no post-hoc comparison test was performed and the effect size was not investigated.

3.3.5.2 Analyses of Grammatical Accuracy

Quantity and quality of the writing productions was analyzed to compare the grammatical accuracy of written production of students learning by the TBI, the FTBI and the CI. The accuracy of each writing production was analyzed. The analyses focused on the amount of the accuracy of linguistic features of the writing tests 1, 2 and 3 of 10 selected subjects from each group. The analyses followed T-unit analysis suggested by Polio's (1997) guideline for T-units, clauses, and errors. The accuracy ratio was calculated by dividing error-free T-units by T-units. Consequently, the data was a ratio score so each set of data was calculated using the Kruskal-Wallis H test. The Kruskal-Wallis H test was carried out on a computer using SPSS software (version 11.5 for window). It was found out that there was no significant difference among three variables in writing tests 1, 2 and 3, so no post-hoc comparison test was performed and the effect size was not investigated (See. Appendix I2).

3.3.5.3 Analyses of Writing Ability

Quantity of students' writings was analyzed to answer research question 3, which was 'Is there any significant difference among writing ability of the students learning by the TBI, the FTBI and the CI? If so, how much are the effect sizes?', three sets of scores of writing tests of the three groups of every subjects were calculated. The data obtained from the writing test measurement were analyzed. The Shapiro-Wilk test was used to find the normality of the data. The significant differences were found, so the Kruskal-Wallis H test was performed to find the significant difference among the rank scores of the three groups. It was found out that there was a significant difference in writing test 3; therefore, a post hoc comparison – the Rank Sums test – was performed on every pair of conditions (See. Appendix I3). The pairs of scores of writing test 3 being compared were re-ranked before the calculation.

To perform the rank sums test (Heiman, 2006),

1) Compute $\sum R$ for each group using Microsoft Excel software.

2) Compute the expected sum of rank ($\sum R_{exp}$) on a computer using Microsoft Excel software. Use the formula

$$\sum R_{exp} = \frac{n(N+1)}{2}$$

3) Compute the rank sums statistics, symbolized by Z_{obt}

$$Z_{obt} = \frac{\Sigma R - \Sigma R_{exp}}{\sqrt{\dfrac{(n_1)(n_2)(N+1)}{12}}}$$

4) Find the critical value of z in the z-table.

5) Compare z_{obt} to z_{crit}

To find a significant relationship, compute the eta squared, using the formula

$$\eta^2 = \frac{(Z_{obt})^2}{N-1}$$

The eta squared value was interpreted following Hopkins's (2002) effect size scale as shown below:

Eta Squared Value	Level of relationship
> 0.2	trivial
0.2 to 0.6	small
0.6 to 1.2	moderate
1.2 to 2.0	large
2.0 to 4	very large
< 4	nearly perfect

3.3.5.4 Analyses of the Relationship between Grammatical Accuracy and Writing Ability

Quantity of the writing accuracy and writing ability of each subject from each group was analyzed to answer research question 4, which is 'Is the relationship between the students' grammar ability and their writing ability significant? If so, how much is the effect size?'. The data were obtained from ten mix-ability subjects from all three groups, so the outliers were not excluded from data analysis. Three sets of data derived from the T-unit analyses and writing ability measures of writing test 1, 2 and 3 were computed using Pearson Product-Moment Correlation run by SPSS software

(version 11.5). The relationship value was interpreted following Kurtz's (1999: 282) level of association as shown below:

r - value	Level of association
.75 to 1.00	very high positive
.50 to .74	high positive
.25 to .49	moderate positive
.00 to .24	low positive

The magnitude of the relationship (r^2) between grammatical accuracy and writing ability was interpreted using Montcalm and Royse's (2002: 215) scale:

r^2	Interpretation
> .20	Slight; inconsequential
.20 to .40	Small; low correlation
.40 to .70	Moderately correlated
.70 to .90	Strong correlation
.90 to 1.00	Very strong correlation

The magnitude of the treatment effects, or the effect size, on the correlation between grammatical accuracy and writing ability was investigated by computing the effect size from the r- value using Furr's formula:

$$d = \frac{r}{\sqrt{p1p2(1-r2)}}$$

Where p1 is the proportion of participants who are in group 1 and p2 is the proportion in group 2.

3.3.5.5 Analyses of Students' Attitude

The data obtained from the measurement of attitude by using the semantic differential scale questionnaire were calculated to find the students' attitudes toward the TBI, the FTBI and the CI. Initially, each questionnaire item was assigned a score according to the answer the respondent gave. The scores were straightforwardly framed with the meaning of the scores for each item, i.e., + 3 = 7, +2 = 6, +1 = 5, 0 = 4, - 1 = 3,

- 2 = 2, - 1 = 1. Then, the scores of each item were keyed in the spreadsheet program (*i.e.*, Microsoft Excel). The mean scores of each item belonging to each group were calculated. Then, the data were categorized according to three dimensional attitudes: evaluation, potency and activity. The mean scores of each item belonging to each category were presented in a radar diagram (See. Appendix I4).