

A Test Taker-Centered Computer-Based Writing Test: Meeting Half Way in the Assessment of Writing

Tanyaporn Arya
Kanchana Prapphal
Chulalongkorn University Language Institute

Abstract

Lack of resources and of the opportunity to produce drafts often leaves EFL learners at a disadvantage during writing tests. This paper reports on a study exploring the effects of provided facilitative features and prescribed drafting in a computer-based writing test on test takers writing performance scores. Participants were one hundred forty-four Thai EFL undergraduate students being required to write argumentative essays. The study employs a two-by-two factorial pretest posttest control group design. Data analyses included the use of the two-way analysis of variance, Kruskal-Wallis H and Mann-Whitney U tests. Results suggest that these two factors could potentially be included in writing tests in favor of the test taker without sacrificing the efficiency of test administration.

1. Background

Efforts to emphasize writing as a process in many ESL contexts and even in the more non-conventional EFL classrooms have become more prevalent than in the past, however, the same does not hold true for the way writing is assessed. This is supported by Hinkel (2002: 46) who has noted that although

methodology in the teaching of ESL writing has veered towards process-centered approaches over the last two decades, the assessment of ESL writing skills found in standardized and institutional writing tests has remained focused on the written product, and not on the writing process. This may be solely due to the administrative convenience of product-oriented tests that seem to be in favor of test users, rather than test takers.

Criticisms over product-oriented writing assessment have been that such tests do not allow test takers to rely on resources of any kind (e.g. East, 2006; Lewkowicz, 1997) or provide opportunities for drafting and receiving feedback during the test (e.g. Cho, 2003; Lee, 2006). These criticisms have prompted research in process-oriented writing assessment especially in ESL settings (e.g. Cho, 2003; Kim, 2002; Lee, 2006). These process-oriented writing assessments have been proven to aid the actual process of writing. During such tests, test takers are able to take advantage of the structured process, stimulus material and feedback from peers to increase the quality of their writing through the engagement of revision.

Nevertheless, process-oriented assessment does not go without disadvantages. One drawback concerns peer feedback. Leki (1990) has suggested that comments from peers may focus on surface forms rather than on ideas and organization or that comments may be vague and unproductive. Many researchers in the EFL context (e.g. Chinnawongs, 2001; Ge, 2005; Moon, 2000) have also found many EFL learners to feel ashamed, threatened and even doubtful of feedback from peers as compared to feedback given to them by teachers. Especially in EFL contexts where, culturally and affectively, learners are not comfortable sharing their work with peers, the process-oriented approach to assessing writing may not fully benefit test takers. Another challenge of process-based writing assessment goes back to the issue of practicality and resources. When testing involves large numbers of examinees, a full-scale process-based writing test may not be viable, as these tests - portfolio or workshop-based - require time, human resources and financial resources, which may not be so easy to acquire.

In an attempt to find middle ground, this study proposes incorporating facilitative features and prescribed drafts as

components which provide test takers with opportunities to improve their writing. Moreover, given that the computer is being used increasingly for language assessment and that EFL learners are becoming more familiar with the use of computers, a computer-based writing test seems to be the transcription mode of option for EFL learners who may in time encounter real-life computer-based standardized tests. Incorporating computer technology, its facilitative devices, a self-evaluation component, together with prescribed drafting might impartially render the test situation in favor of the testtaker without sacrificing the efficiency of test administration.

The test taker-centered computer-based writing test

The test taker-centered computer-based writing test (from now on T-CBWT) is a writing performance test that serves to ensure that every test taker has the opportunity to perform at their best (Hamp-Lyons, 1990:73) or to 'bias for best' (See Fulcher, 2000: 97).

The T-CBWT prompt requires test takers to produce an evaluative (argumentative) essay. The test prompt is similar to that of the Chulalongkorn University Language Institute (CULI) Writing Test prompt, which also mirrors TOEFL's independent writing section requiring test takers to express their opinion in writing and support it based on their own knowledge and experience (TOEFL iBT Tips, 2005). The controlled test topic covers general topical knowledge about social/environmental issues, since such issues are current and can easily be related to. The topics have also been adjusted to suit the socio-cultural background of the test takers.

The T-CBWT is distinct in that test takers use the Microsoft Word processor and its functions - translation (or English - Thai dictionary), thesaurus, spell-check and grammar-check functions - to facilitate their writing. An additional facilitative feature is the Self-reflective Reminder Questions (SRQ) given to test takers in checklist form to guide them through their writing. The SRQ is divided into 4 sections: what the writer should do during the pre-writing stage, during writing stage, revising stage, and editing and proofreading stage (See Appendix A for the Thai version distributed to test takers).

In addition, structured draft writing is required¹. The drafting stages encompass (1) producing an outline, (2) writing the first draft concentrating on content, (3) revising the first draft focusing on content and organization, and (4) editing the second draft, working on language, grammar, and mechanics. The Microsoft Word Track Changes device is required for those producing drafts ².

Further, the T-CBWT allows test takers 90 minutes to complete the task. This time allotment is one hour longer than that of the standard 30-minute single draft writing tests of similar formats. Translated key words are provided to give all test takers equal head start.

2. Objectives & Research Questions

The primary purpose of this study is to examine the effects of facilitative features and prescribed multiple drafts in a T-CBWT on test takers writing performance. The study also aims to explore how test takers make use of the facilitative features provided and test takers' perspectives toward the T-CBWT. The study specifically addresses the following questions:

1. Do facilitative functions (MS translation, thesaurus, spell & grammar-check devices & self-reflective reminder questions) used in the T-CBWT have an effect on test takers' English writing scores?
2. Do multiple drafts required in the T-CBWT have an effect on test takers' English writing scores?
3. Which combination of factors (facilitative functions and/or the required multiple drafts) used in the T-CBWT has a significant effect on test takers' English writing scores?
4. How do test takers make use of each of the facilitative features?

¹ In the study, the prescribed drafts were only required for two experimental groups, Groups 1 and 2.

² This is because the author also studied test takers' textual changes made from draft to draft. This qualitative section of the study can be found in the author's unpublished doctoral dissertation.

5. What are the test takers' perspectives towards the T-CBWT?

3. Methodology

3.1. Research design

To explore the first three research questions posed, quantitative analyses were performed, employing a randomized block design and a two by two-factorial pretest posttest control group design (Isaac and Michael, 1995). The two by two-factorial design was selected, being a design where the effect of different treatment combinations can be studied simultaneously. The randomized block design was selected to ensure that variance in terms of writing proficiency level in the data is reduced. Specifically, the two factors are the facilitative features and required multiple draft writing. Figure 1 illustrates the research design in this study.

Figure 1 Factors and levels

		Factor A: Facilitative Functions	
		With functions	Without functions
Factor B: Multiple Drafts	With drafts	Test Group 1 N=36 with Functions with Drafts	Test Group 2 N=36 without Functions with Drafts
	Without drafts	Test Group 3 N=36 with Functions without Drafts	Test Group 4 N=36 without Functions without Drafts

3.2. Participants

Participants are 144 Thai first-year undergraduates studying in the Faculty of Commerce and Accountancy, Chulalongkorn University, academic year 2006. They represented three writing

proficiency levels as measured by scores obtained from a computer-based writing pretest (CBWT), a test parallel to the T-CBWT. The pretest scores were used in statistically equating the participants, who were later mechanically matched and assigned to four equal groups of 36 participants, with 12 participants in each of the three writing proficiency levels (advanced, intermediate and low intermediate).

Because many researchers are uncomfortable with relying on random assignment if there are fewer than 40 cases in each group (Fraenkel and Wallen, 2000:286), the four groups were tested for equal variances. To ensure normality of each group, the Test of Normality was performed. As the sample size per group is less than 50, the Shapiro-Wilks test was studied. The significance level of .621, .642, .860, .589 were obtained for Groups 1, 2, 3 and 4 respectively, all being greater than .05 led to the conclusion that the four groups were normal.

3.3. Instruments

The following instruments used were validated by 5 experts, revised, tried out in a pilot study and revised again before utilization in the main study.

The CBWT & T-CBWT

The CBWT, the parallel of the T-CBWT, is a computer-based writing proficiency test used as a pretest to categorize subjects into writing ability levels before administering the T-CBWT (See Appendix B for a comparison between the CBWT and T-CBWT and Appendix C for prompts of both tests). The T-CBWT, described earlier, was used to assess test takers' writing performance. As recommended by Bachman (1990: 183), to assume that different forms of a test are parallel, they should be checked for equivalent difficulty level. This was carried out using Scannell and Tracey's classical formula (1975: 223) as recommended by สุขุมลันต์ (๒๕๕๗:๕๐).

Difficulty levels ³ of 0.233 and 0.277 representing that of the CBWT and T-CBWT respectively indicated that the two tests were relatively at the same difficulty level. To determine parallel forms reliability, a reliability coefficient was calculated on the scores of the two measures in a counterbalancing technique applied during the pilot study. The result of the Pearson's Correlation Coefficient, yielded a high and positive correlation ($r = .750$, $p < 0.01$), assuring parallelism of the two forms.

The analytical rating scale

The analytical rating scale ⁴ used in the study was developed through the integration and adaptation of several holistic⁵ and analytical rating scales ⁶ to ensure that as many relevant constructs as possible were incorporated. An analytic scale was chosen to enable investigation into the strengths and weaknesses in different areas of the test takers' writing performance. The rating scale focuses on three main aspects: content, organization, and language use. It consists of nine subscales measuring clarity & explicitness, and topic development & supportive examples; rhetorical organization and coherence; knowledge of grammar, vocabulary, cohesion, sentence structure and mechanics. This rating scale was used for rating both the CBWT and T-CBWT.

Inter-rater reliability

The raters consisted of two experienced English language instructors who had formerly taught a number of writing courses

³ A test with a Difficulty Index (IDiff) between 0.20-0.80 is a test that is not too easy or too difficult. A test with a IDiff of below 0.20 is a test that is too difficult, while a test with a IDiff of over 0.80 is a test that is too easy (สุกมลรัตน์ ๒๕๔๗: ๕๑).

⁴ Refer to author's unpublished doctoral dissertation for the analytical rating scale.

⁵ Writing Competency Exam Holistic Rating Scale (Retrieved August 2006), Idaho State University Writing Center Holistic Rating Scale (Retrieved August 2006), PSAT - A Writing Performance Definition (Retrieved August 2006), the Chulalongkorn University Language Institute's Scoring Scale for the Foundation English Course, TOEFL iBT Test - Independent Writing Rubrics (2005) the Michigan Writing Assessment Scoring Guide (Weigle, 2002)

⁶ Kim's (2002) Analytical Rubrics

and the researcher herself. In addition to rater training, which took place during the pilot study phase when the raters utilized and adjusted the rating scale, statistical methods were used to check for inter-rater reliability. Initially, rater consistency was checked using the Pearson product-moment correlation coefficient to compute the correlation between the three raters. Using pretest scores, a significant positive relationship existed between all the raters on all three main aspects of the text (content, organization, and language) with an r ranging from 0.787 to 0.937 ($p < 0.05$).

In addition, the coefficient alpha, as recommended by Ebel (1979) was used to compute inter-rater reliability, as more than two raters were involved. The Cronbach alpha coefficient for the pretest was 0.96, while the alpha coefficient for the posttest was 0.98, demonstrating acceptable inter-rater reliability⁷. The scores of three raters were averaged and counted as the pretest scores in order to arrive at a more appropriate proximity of the subjects' performance.

The retrospective questionnaire

A retrospective questionnaire (See Appendix D for a Thai version) was developed to elicit information on how test takers made use of the facilitative functions and on their perspectives towards the T-CBWT. The questionnaire was in the form of a 4-point Likert scale in order to reduce the 'central tendency bias' which occurs when respondents try to avoid choosing extreme responses (Vongpadungkiat 2006).

As the Cronbach alpha can account for weighted responses, it was used to estimate the questionnaires' reliability during the pilot study, showing alpha values of above 0.75. For the main study, another test using the Cronbach alpha was carried out to estimate internal consistency, again receiving an alpha value of above 0.75, indicating acceptable reliability (Hair et al., 1995).

⁷ An alpha higher than 0.75 is interpreted as an acceptable level of inter-rater reliability (Hair, et al., 1995).

3.4. Procedures

Approximately 4 weeks after the CBWT (pretest) was delivered, the T-CBWT (posttest) was administered to the three experimental groups (Groups 1-3) and the control group (Group 4) which was given the same T-CBWT test prompt without any special treatment. Participants completed the test in the same computer labs for both the pretest and posttest. The researcher and an assistant proctored all the examination sessions.

Prior to every test session, computers were preset for each test taker according to their assigned test format. Upon entering the computer lab, test takers were given a test packet that corresponded to the type of test they were to take. A test packet for Group 1, for instance, would consist of the test prompt; an instruction card explaining drafting procedures⁸; a function card explaining how to use facilitative features⁹; a self-reflective questions checklist and the retrospective questionnaire. Test takers were then requested to sit at their assigned computer, view the orientation slideshow¹⁰ and observe demonstrations delivered individually by the researcher to make sure they understood how to use the functions. The researcher and research assistant were available throughout the test in case questions arose. After having saved their work on to the computer, test takers completed and submitted the retrospective questionnaire within the test session. Later, test takers' responses were transferred to a USB drive and back up floppy discs, printed and rated by the three raters.

3.5. Data Analysis & Results

In order to examine if there were statistically significant differences in the writing performance scores among the four test

⁸ The instructions card can be viewed in the researcher's unpublished doctoral dissertation.

⁹ The functions card can be viewed in the researcher's unpublished doctoral dissertation.

¹⁰ This was the same PowerPoint slideshow previously sent to test takers via email approximately one week prior to the T-CBWT to prepare them and inform them of the specific procedures expected of them during the test. Each test group had a different slideshow corresponding to their test condition.

conditions, the two-way analysis of variance was conducted¹¹. This analysis permits the researcher to study (1) the effect of the facilitative functions, (2) the effect of the required multiple drafts, and (3) the effect of a combination of facilitative functions and required drafts on the test takers' written performance simultaneously, addressing the first three research questions.

The results of the two-way analysis of variance revealed that neither the availability of facilitative functions ($F(1, 140) = .854, p > 0.05$) nor the prescribed multiple drafts ($F(1, 140) = .847, p > 0.05$) influenced the overall writing performance of test takers in any of the test conditions. The F values being higher than .05 indicated that neither the availability of facilitative functions nor required draft writing had an effect on writing performance. Neither was there evidence to conclude that both factors combined led to better overall writing performance within any test condition. The F value of $(1, 140) = 1.826, p > 0.05$ shows no interaction effect between facilitative function and drafts. This means that writing performance was not affected by the provided facilitative functions, whether or not required drafts were written, vice versa. Because neither effect was significant, post-hoc analyses were not calculated.

Non-significant results from the two-way analysis of variance tests prompted the researcher to further examine which test condition had an effect on the performance of each analytical sub-component (i.e. clarity & explicitness, topic development & supportive examples, rhetorical organization, coherence, sentence structure, cohesion, vocabulary, grammar, or mechanics) at each writing proficiency level (advanced, intermediate, low intermediate). Thus, test takers' sub-scores from each analytical category, rather than the total posttest scores, were used for analyses. To calculate this, the Kruskal Wallis H test for independent samples was performed to accommodate the small sample size (now 12 cases per group) of each writing proficiency level. The Kruskal-Wallis test was run for every writing sub-component until results revealed a

¹¹ Prior to conducting the two-way ANOVA, initial data analysis consisted of descriptive statistics, test of population normality and a test of homogeneity of variance to check that assumptions were met.

significant difference between test performances of the four test conditions on some analytical sub-component within the writing proficiency level.

Since the Kruskal-Wallis procedure does not identify where the significant difference lies and visual inspection of mean ranks does not suffice to make conclusions, Mann-Whitney U tests were performed to indicate statistically where differences were, as suggested by พงษ์วิชัย (๒๕๔๘: ๒๒๔). The Mann-Whitney test allows us to decide when a difference between samples can be claimed (Hinton et al., 2004). This was carried out by comparing the differences of sub-scores (of the nine sub-categories mentioned in parentheses above) between two test groups or test conditions at a time across writing proficiency levels.

When studying mechanic scores of test takers of the low intermediate, intermediate and advanced writing proficiency levels, unequal mean ranks across the four test groups indicated some differences between the groups' scores. The Kruskal-Wallis H test on mechanic scores of test takers of the three writing proficiency levels, as illustrated in Table 1.1, showed Chi-Square values and Asymptotic sig., estimates of the true p value of less than 0.05, revealing that the difference between the scores on mechanics of the four test groups was significant for test takers across writing proficiency levels.

Table 1.1 Kruskal-Wallis Test on Mechanics Scores

Writing Proficiency Level	Test Statistics(a,b)		
	Chi-Square χ^2	df	Asymp Sig.
Low intermediate level test takers	11.836	3	.008
Intermediate level test takers	12.156	3	.007
Advanced level test takers	10.248	3	.017

a Kruskal Wallis Test, ($p < 0.05$)

b Grouping Variable: GROUP

To indicate where the significant difference lies for mechanics scores of the low intermediate writers, Mann-Whitney U tests were conducted. Table 1.2 presents the results of Mann-Whitney U tests of the low intermediate writers.

Table 1.2 Low Intermediate Writers' Mechanics Scores and Mann-Whitney U Test Results

Test Group (Condition)	Means (SD)	Mann Whitney U	Exact Sig. [2*(1-tailed Sig.)]	Sig.
FFD	6.25 (.621)			
D	5.16 (1.26)	34.000	.028	.014*
FFD	6.25 (.621)	43.000	.101	.050*
F	5.66 (.778)			
FFD	6.25 (.621)	21.500	.002	.001*
C	4.83 (1.19)			
D	5.16 (1.26)	57.000	.410	.205
F	5.66 (.778)			
D	5.16 (1.26)	58.000	.433	.221
C	4.83 (1.19)			
F	5.66 (.778)	40.000	.068	.034*
C	4.83 (1.19)			

Note: FFD = Facilitative Functions and Required Drafts Test Group,
D = Required Drafts Test
Group, F = Facilitative Functions Test Group, C = Control Group

The table shows that the scores of the 'with facilitative functions and drafts' test group ($\bar{x} = 6.25$) is significantly ($p < 0.05$) higher than those of the 'with drafts' test group ($\bar{x} = 5.16$), the 'with facilitative functions' test group ($\bar{x} = 5.66$) and the control group ($\bar{x} = 4.83$) ($p < 0.05$). The test also reveals that the scores of the 'with facilitative functions' test group is significantly higher than that of the control group ($p < 0.05$).

The Exact Significance value rather than the Asymp. Sig. for all of the Mann-Whitney U calculations is reported because the

dataset in each group is small ($N = 12$), thus reporting this significance level will reflect a more accurate judgment of significance (Hinton et al., 2004). Also, since the hypothesis is one-tailed (Null hypothesis: Scores of the experimental groups or groups with Facilitative Functions are higher than scores of the control group or groups without Facilitative Functions.), the p value is halved to ensure that the difference is in the correct direction (ibid.).

At the intermediate writing proficiency level, the Mann-Whitney tests for intermediate writers in Table 1.3 illustrate that Mechanics scores of the 'with facilitative functions and drafts' test group are significantly higher than those of the control group ($p < 0.05$). The Mechanics scores of the 'with facilitative functions' test group are also shown to be significantly higher than those of the control group ($p < 0.05$).

Table 1.3 Intermediate Writers' Mechanics Scores and Mann-Whitney U Test Results

Test Group (Condition)	Means (SD)	Mann Whitney U	Exact Sig. [2*(1-tailed Sig.)]	Sig.
FFD	6.83 (.389)			
D	6.16 (.834)	35.000	.033	.165
FFD	6.83 (.389)	62.000	.590	.295
F	6.50 (1.00)			
FFD	6.83 (.389)	19.000	.001	.005*
C	5.75 (.753)			
D	6.16 (.834)	54.500	.319	.159
F	6.50 (1.00)			
D	6.16 (.834)	47.500	.160	.080
C	5.75 (.753)			
F	6.50 (1.00)	40.000	.068	.034*
C	5.75 (.753)			

Note: FFD = Facilitative Functions and Required Drafts Test Group, D = Required Drafts Test Group, F = Facilitative Functions Test Group, C = Control Group

In line with the trend, results of the Mann-Whitney tests in Table 1.4 also revealed that with regard to advanced writers, Mechanics scores of the ‘with facilitative functions’ test group were significantly higher than those of the control group, the ‘with required drafts’ test groups, and the control group ($p < 0.05$).

Table 1.4 Advanced Writers’ Mechanics Scores and Mann-Whitney *U* Test Results

Test Group (Condition)	Means (SD)	Mann Whitney <i>U</i>	Exact Sig. [2*(1-tailed Sig.)]	Sig.
FFD	6.83 (.577)			
D	6.66 (.887)	64.000	.671	.335
FFD	6.83 (.577)	42.000	.089	.044*
F	7.33 (.492)			
FFD	6.83 (.577)	45.500	.128	.064
C	6.25 (.965)			
D	6.66 (.887)	40.000	.068	.034*
F	7.33 (.492)			
D	6.66 (.887)	54.500	.319	.159
C	6.25 (.965)			
F	7.33 (.492)	26.000	.007	.003*
C	6.25 (.965)			

Note: FFD = Facilitative Functions and Required Drafts Test Group, D = Required Drafts Test Group, F = Facilitative Functions Test Group, C = Control Group

Upon examining clarity & explicitness as well as topic development & supporting details scores, mean ranks observed pointed to differences across test conditions particularly among test takers of the advanced level. The Kruskal-Wallis *H* test as shown in Table 2.1 revealed results of $\chi^2 = 9.732$, $df = 3$, $p < 0.05$ with the Asmp. Sig. value of .021 interpreted as a significant difference between the scores on Clarity and Explicitness among the four test groups and a $\chi^2 = 9.005$, $df = 3$, $p < 0.05$ between scores on Topic Development and Supporting Details across test conditions among the advanced level writers.

Table 2.1 Kruskal-Wallis Test on Advanced Writers' Clarity & Explicitness and Topic Development & Supporting Details Scores

Test Statistics			
Sub-category	Chi-Square χ^2	Df	Asymp Sig.
Clarity & Explicitness	9.732	3	.021
Topic Development & Supporting Details	9.005	3	.029

Further, the Mann-Whitney U tests output seen in Table 2.2 indicate that Clarity and Explicitness scores of the 'with required drafts' test group were significantly higher than those of the 'with facilitative functions and drafts' test group, the 'with functions' test group and the control group ($p < 0.05$).

Table 2.2 Advanced Writers' Clarity & Explicitness Scores and Mann-Whitney U

Test Results

Test Group (Condition)	Means (SD)	Mann Whitney U	Exact Sig. [2*(1-tailed Sig.)]	Sig.
FFD	5.75 (1.05)			
D	6.83 (.717)	28.000	.010	.005*
FFD	5.75 (1.05)	45.000	.128	.064
F	6.25 (.452)			
FFD	5.75 (1.05)	63.500	.630	.315
C	6.00 (1.04)			
D	6.83 (.717)	39.000	.060	.030*
F	6.25 (.452)			
D	6.83 (.717)	38.000	.052	.026*
C	6.00 (1.04)			
F	6.25 (.452)	57.000	.410	.205
C	6.00 (1.04)			

Note: FFD = Facilitative Functions and Required Drafts Test Group, D = Required Drafts Test Group, F = Facilitative Functions Test Group, C = Control Group

Finally, results of the Mann-Whitney tests shown in Table 2.3 reveal that the scores for Topic development and Supporting details of the ‘with required drafts’ test group are significantly higher than those of the ‘with facilitative functions and drafts’ test group, the ‘with functions’ test group and the control group ($p < 0.05$).

Table 2.3 Advanced Writers’ Topic Development & Supporting Details Scores and Mann-Whitney *U* Test Results

Test Group (Condition)	Means (SD)	Mann Whitney <i>U</i>	Exact Sig. [2*(1-tailed Sig.)]	Sig.
FFD	5.91 (1.16)			
D	6.66 (1.07)	36.500	.039	.019*
FFD	5.91 (1.16)	63.000	.630	.315
F	5.91 (.668)			
FFD	5.91 (1.16)	56.500	.378	.189
C	5.41 (1.16)			
D	6.66 (1.07)	34.500	.028	.014*
F	5.91 (.668)			
D	6.66 (1.07)	31.500	.017	.008*
C	5.41 (1.16)			
F	5.91 (.668)	52.000	.266	.133
C	5.41 (1.16)			

Note: FFD = Facilitative Functions and Required Drafts Test Group, D = Required Drafts Test Group, F = Facilitative Functions Test Group, C = Control Group

In sum, the Kruskal-Wallis *H* and Mann-Whitney *U* tests jointly revealed that test condition had some effect on Mechanics scores for test takers of all writing proficiency levels. Low intermediate, intermediate and advanced level test takers in the ‘with facilitative functions’ test condition regardless of whether or not they also had required drafts, obtained significantly higher scores on Mechanics than writers in test conditions that did not receive facilitative features.

These two tests also revealed that particularly for the advanced level writers the “with drafts” test condition had an effect

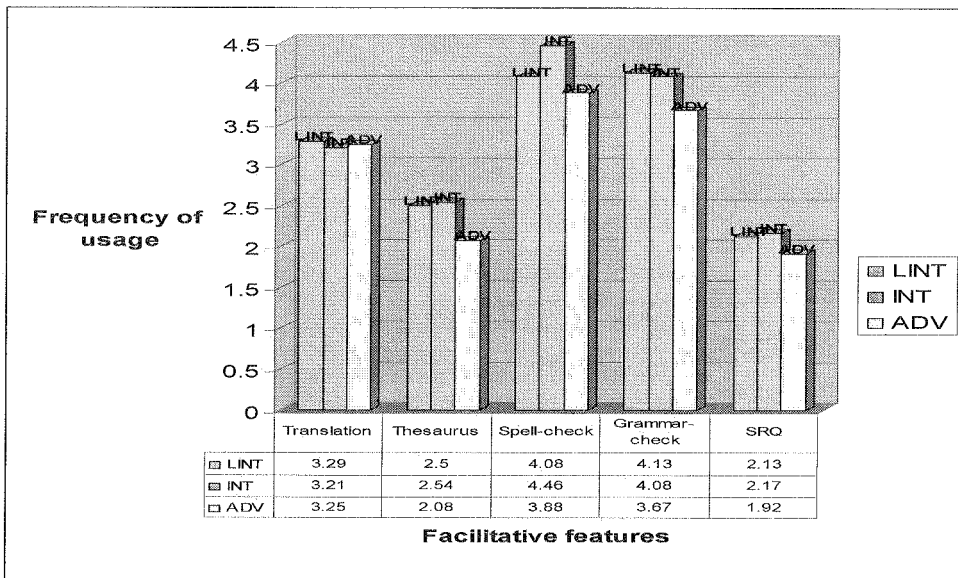
on clarity & explicitness as well as topic development and supporting details scores. Writers of the advanced level who were in the ‘with required drafts’ condition obtained significantly higher scores on Clarity and Explicitness as well as Topic development and Supporting Details (or their overall Content scores) than advanced writers in all other test conditions.

Research question four was answered by studying test takers’ frequency of usage of the facilitative functions, drawing from data collected from the retrospective questionnaire. Only test takers from Groups 1 and 3 who had access to these facilitative functions answered these questions. Prior to the analysis, the following criterion was set with mean scores interpreted as follows:

- 1.00-1.49 A very low degree (< 5 times)
- 1.50-2.49 A low degree (6-10 times)
- 2.50-3.49 A somewhat high degree (11-15 times)
- 3.50-4.00 A very high degree (>15 times)

Figure 2 presents facilitative feature usage of test takers according to writing proficiency level.

Figure 2 Facilitative feature usage



Note: LINT = test takers of the low intermediate writing proficiency level, INT = test takers of the intermediate writing proficiency level, ADV = test takers of the advanced writing proficiency level

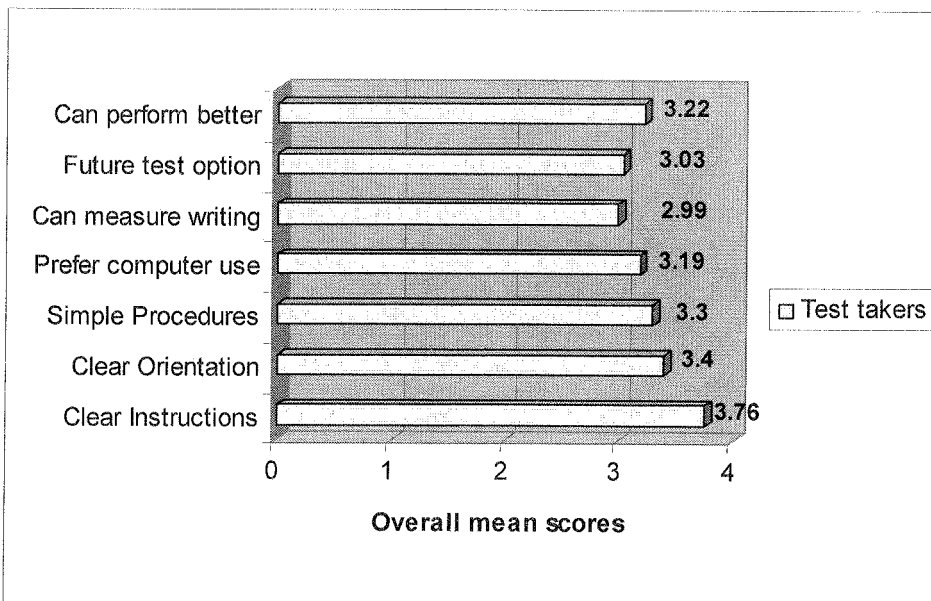
Results show that the most used facilitative feature is the spell-check tool. The second most utilized facilitative function is the grammar-check function, following a similar trend. The translation device was used by test takers of all three writing proficiency levels to a somewhat high degree. The thesaurus was used to a very low degree, about 6-10 times throughout the writing process. The least used facilitative function of all was the Self-reflective Reminder Questions (SRQ) with test takers of all writing proficiency levels consulting the SRQ only to a very low degree.

Research question five was investigated by studying the frequency counts derived from data collected from the retrospective questionnaire. Before the analyses were carried out, the following criteria was set. High scores indicated positive views toward the T-CBWT and low scores negative views. The mean scores are interpreted as follows:

- 1.00-1.49 Strong, negative view
- 1.50-2.49 Somewhat negative view
- 2.50-3.49 Somewhat positive view
- 3.50-4.00 Strong, positive view

Figure 3 presents overall opinions towards the T-CBWT.

Figure 3 Opinions towards the T-CBWT



From the figure, it can be concluded that test takers in general have positive views towards this type of test. Test takers for the most part thought the test instructions and orientation were clear, that the procedures were simple enough. Test takers in general felt positive towards using the computer to write during a test. Most felt the T-CBWT could measure their true writing ability. Overall, test takers felt they could write better on this type of test, and that they would opt for the T-CBWT again in the future.

4. Discussion & Implications

The Role of Facilitative Features

That all test conditions with facilitative features had an effect on mechanics scores for test takers of all writing proficiency levels clearly indicated that the spell-check and grammar-check functions in the MS Word were instrumental in helping test takers with lower level functions of the task, i.e. spelling and punctuation. Evidently, the instruments which played the most significant roles were not self-instigated, rather prompted by the computer. This suggests that test takers did not take full advantage of the remaining facilitative features, like the dictionary, thesaurus or the Self-reflective Reminder Questions.

Direct pedagogical implications would then point to the need to provide focused and on-going training to EFL learners in the accurate use of the dictionary. Such training take place for an extended period so that learners may be able to appreciate the benefits from being able to take full advantage of dictionary look-ups, for instance, to help them with part of speech or adjectives and adverbs, which will be useful during revision. With teacher encouragement, frequent practice of looking up words from the dictionary may potentially extend to the look-up of the thesaurus.

Other instructional implications are with respect to monitoring. That test takers made little use of self-reflective reminder questions during their writing would suggest that EFL learners may still lack training in the way of self-monitoring. This would require the instructor to familiarize learners with monitoring or evaluating their own work alongside the usual grammar lessons. Furthermore, instructors might want to find ways to instill the

notion that writing is an ongoing process involving evaluation at every stage. When teaching writing, evaluation can be given to assist students throughout the process and not only at the end when the final product is complete.

In terms of computer-assisted language learning, EFL learners might be exposed more to the computer for writing activities. Having learners frequently use the MS word processor and the tools that come with the program incorporates writing practice with typing skills, enhancing both English and typing skills that learners might require for future testing purposes.

The Role of Prescribed Drafts

That test condition had an effect on clarity & explicitness as well as topic development and supporting details scores, particularly for advanced level writers in the “with drafts” group, suggested that imposed draft writing may be instrumental in channeling the attention of the advanced level test takers’ in this test condition to content. That the same effect was not evident for the advanced level test takers of the ‘with facilitative functions and drafts’ group may be due to their preoccupation with utilizing facilitative functions to aid their writing. Had they been trained in utilizing the facilitative functions for a more extensive period and were more at ease with these features, the outcome may have been different.

Furthermore, that the required multiple drafts affected the content scores of only the advanced level test takers and not test takers of other writing proficiency levels in the same test condition might have been because the advanced level test takers were sufficiently skillful in the language enabling them to easily shift their attention to the development of their content. On the contrary, test takers of the intermediate and low-intermediate levels of the same test condition may have had to divide their attention to language as well as content, using up more time to structure their sentences and left with less time to develop their content.

The pedagogical implications would then signify that especially less skilled EFL learners necessitate more than simply writing lessons. These learners might also benefit from general

The Authors

Kanchana Prapphal, the corresponding author, is a professor at Chulalongkorn University Language Institute (CULI). She was President of Thai TESOL in 1986, Director of CULI from 1996-1999 and has been Director of the Academic Testing Center of Chulalongkorn University since 2000. Her main interests are language testing, classroom-based research and applications of IT in language teaching and testing.

Tanyaporn Arya received her M.A. in TESOL from San Jose State University. She is currently an instructor at Chulalongkorn University Language Institute and a Ph.D. student at Chulalongkorn University, specializing in language assessment. Her current research interests include language assessment, self-assessment, and impact studies.

References

- สุพัฒน์ สุกมลสันต์ (๒๕๔๗). การวิเคราะห์ข้อสอบแบบอัตนัยรายข้อ. *ภาษาปริทัศน์* ๒๑ : ๔๔-๕๕.
- ศิริชัย พงษ์วิชัย (๒๕๔๘). การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์ พิมพ์ครั้งที่ ๑๕. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- Arya, T. (2008). *The Effects of a Test Taker-Centered Computer-Based Writing Test on Thai First Year University Students' Writing Performance*. Unpublished doctoral dissertation. Chulalongkorn University.
- Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford: University Press.
- Chinnawongs, S (2001). In search of an optimal writing feedback strategy. *PASAA*, 31, 27-43.
- Cho, Y. (2003). Assessing Writing: Are We Bound by Only One Method? *Assessing Writing*, 8 (3), 165-191.
- East, M. (2006). The impact of bilingual dictionaries on lexical sophistication and lexical accuracy in tests of L2 writing proficiency: A quantitative analysis. *Assessing Writing*, 11 (2006), 179-197. Retrieved November 2007 from: www.sciencedirect.com
- Ebel, R.L. (1979). *Essentials of Educational Measurement*. Third edition. Englewood Cliffs: Prentice-Hall.
- Educational Testing Service (2005). TOEFL iBT Tips: How to Prepare for the Next Generation TOEFL Test and Communicate with Confidence. Retrieved July 2005 from: http://www.ets.org/Media/Tests/TOEFL/pdf/TOEFL_Tips.pdf

- Fraenkel, J.R. & Wallen, N.E. (2000). *How to design and evaluate research in education*. Fourth edition. Boston: McGraw-Hill.
- Fulcher, G. (2000). Computers in Language Testing. In Brett, P. & Motteram, G (Eds.). *A Special Interest in Computers: Learning and Teaching with Information and Communications Technology* (pp. 93-107). Manchester: IATEFL Publications.
- Ge, G. (2005). A strategic approach to teaching English writing. *Teaching English in China*, 28 (6), 110-116.
- Hair, R, Anderson, R. Tatham and W. Black (1995). *Multivariate Data Analysis*. New York: Macmillan Publishing.
- Hamp-Lyons, L. (1990). Second language writing: assessment issues. In B. Kroll (Ed.). *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press.
- Hinkle, D.E., Wiersma, W. & Jurs, S.G. (1988), *Applied Statistics for the Behavioral Sciences*. Second edition. Houghton Mifflin Co.
- Hinton, P.R., Brownlow, C., McMurray, I., & Cozens, B. (2004). *SPSS Explained*. Routledge.
- Horowitz, D. (1986). Process, not product: Less than meets the eye (in The Forum). *TESOL Quarterly*, 20, 141-144.
- Idaho State University Writing Center Holistic Rating Scale. Retrieved August 2006 from <http://www.sde.idaho.gov/sasa/docs/TCGuide%209-07-07.rtf>
- Isaac, S. & Michael, W. B. (1995). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in the planning, design and evaluation of studies in education and the behavioral sciences (3rd edition)*. San Diego, CA: EdITS/Educational and Industrial Testing Services.
- Kim, M. (2002). *Process and Product: An Investigation of the Writing of Non-native Speakers of English in a Computer-based Academic English Writing Test*. Unpublished doctoral dissertation. University of California at Los Angeles.
- Lee, Y.-J. (2006). The process-oriented ESL writing assessment: Promises and challenges. *Journal of Second Language Writing*, 15, 307-330.
- Leki, I. (1990). Coaching from the margins: Issues in written response. In Kroll (Eds.), *Second Language Writing: Research Insights for the Classroom*. Cambridge: Cambridge University Press.
- Lewkowicz, J. (1997). *Investigating Authenticity in Language Testing*. Unpublished doctoral dissertation, University of Lancaster.

- Moon, Y.I. (2000). The nature of peer revisions in an EFL writing class of a Korean university. *English Teaching*, 55 (1), 119-140.
- PSAE - A Writing Performance Definition. Retrieved August 2006 from http://www.isbe.net/assessment/htmls/writing_psaedef.htm
- Scannell, D.P. & Tracey, D.B. (1975). *Testing and Measurement in the Classroom*. Boston: Houghton Mifflin Company.
- TOEFL iBT Test – Independent Writing Rubrics retrieved 2005 from: <http://www.ets.org/Media/Tests/TOEFL/pdf/ibt%20scores.pdf>
- Trochim, W.M.K. (2002). *The Qualitative-Quantitative Debate*. Retrieved September 20, 2005 from: <http://www.socialresearchmethods.net/kb/qualdeb.htm>
- Vongpadungkiat, N. (2006). *A Study of Selected Predictors of English Reading Tests Performance in Consumer Product Marketing of Fourth-year University Students*. Unpublished doctoral dissertation. Chulalongkorn University.
- Weigle, S.C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Writing Competency Exam Holistic Rating Scale. Retrieved August 2006 from: <http://www.biola.edu/sics/downloads/2007/MACertificate-TESOLHandbook.pdf>

Appendix A

Self-reflective Reminder Questions (SRQ)

Self-reflective Reminder Questions

ชุดคำถามนี้ประกอบไปด้วยคำถามที่จะเตือนนิสิตว่าได้ทำในสิ่งที่ควรตลอดระยะเวลาการเขียนเรียงความหรือไม่ นิสิตสามารถกลับมาดูเพื่อเป็นการเตือนตัวเองได้ตลอดเวลา

วิธีทำ: ให้นิสิตทำเครื่องหมาย ✓ ในช่อง เฉพาะเวลาที่ได้ทำสิ่งเหล่านั้นแล้ว

ก่อนเขียน Pre-writing (Outline)

1. ฉันได้ทบทวนความคิดทั้งหมดลงไปบนกระดาษทด/คอมพิวเตอร์แล้วหรือยัง Yes
2. ฉันได้แยกความคิดออกเป็นประเด็นต่าง ๆ ที่ชัดเจนหรือยัง Yes
3. ฉันตัดสินใจได้หรือยังว่าจะเรียงเรียงลำดับความคิดของฉันอย่างไร Yes
4. ฉันนึกถึงตัวอย่างที่จะใช้สนับสนุนประเด็นต่าง ๆ หรือยัง Yes

ระหว่างเขียน During Writing (1st Draft)

5. ฉันเขียนรวบรวมรายละเอียดทุกอย่างตามที่ทด/คิดไว้หรือเปล่า Yes
6. ฉันมีประโยคหลักที่ตอบคำถามในโจทย์แล้วหรือยัง Yes
7. ฉันได้คิดถึงคำเชื่อมที่จะใช้ในการเชื่อมประเด็นความคิดในแต่ละส่วนแล้วหรือยัง Yes

การปรับเนื้อหาและเรียบเรียงใหม่ Rewriting (2nd Draft)

เนื้อหา Content:

8. ฉันมีประโยคความคิดหลักที่ตอบคำถามในโจทย์หรือยัง Yes
9. ฉันตอบคำถามได้อย่างถี่ถ้วนหรือยัง Yes
10. ฉันเขียนให้ผู้อ่านด้วยภาษาที่เหมาะสมหรือไม่ Yes
11. งานเขียนของฉันมีอย่างน้อย3ประเด็นย่อยไว้สนับสนุนความคิดหลัก Yes
12. แต่ละประเด็นย่อยมีตัวอย่างสนับสนุนอย่างแน่นหนาหรือไม่ Yes
- สามารถเพิ่มตัวอย่างได้อีกไหม Yes
13. ตัวอย่างและรายละเอียดเหมาะสม เกี่ยวข้อง และสมเหตุสมผลหรือเปล่า Yes
- ฉันสามารถตัดสิ่งที่ไม่เกี่ยวข้องออกได้ไหม Yes

14. ข้อถกเถียง/ประเด็นรายละเอียดต่าง ๆ หนักแน่นพอไหม Yes
สามารถเขียนให้มีน้ำหนักมากกว่านี้ได้ไหม Yes

การเรียบเรียง **Organization:**

15. งานเขียนของฉันมี คำนำ เนื้อหา บทสรุป ที่มีประสิทธิภาพแล้วหรือยัง Yes
16. งานเขียนของฉันแบ่งออกเป็นย่อหน้า ๆ ตามประเด็นย่อยที่มีอยู่หรือเปล่า Yes
17. ฉันได้จัดลำดับย่อหน้าต่าง ๆ ให้ต่อเนื่องกันอย่างสมเหตุสมผล Yes
18. แต่ละย่อหน้าเกี่ยวข้องกับประเด็นหลักหรือไม่ Yes
19. แต่ละย่อหน้ามีรายละเอียด มีน้ำหนักเพียงพอ และเท่า ๆ กันไหม Yes
20. ฉันได้ใช้การเชื่อมโยงความคิดให้มีความต่อเนื่องจากย่อหน้าหนึ่ง –
ไปอีกย่อหน้าหนึ่งหรือเปล่า Yes
21. ฉันได้ใช้คำเชื่อมเพื่อเชื่อมโยงความคิดตลอดทั้งเรื่อง Yes

การอ่านทบทวนเพื่อแก้ไขการใช้ภาษา **Editing and Proofreading (3rd Draft)**

22. ฉันได้อ่านเรื่องของฉันเพื่อตรวจสอบคุณภาพโดยรวมของงานเขียนหรือยัง Yes

ภาษา **Language**

23. ฉันใช้รูปประโยคที่หลากหลายหรือเปล่า Yes
24. ประโยคต่าง ๆ เขียนได้ดีหรือยัง Yes
ปรับแก้ให้เป็นประโยคที่สมบูรณ์ และมีประสิทธิภาพได้ไหม Yes
25. ฉันใช้คำเชื่อมต่าง ๆ ได้อย่างถูกต้องหรือเปล่า Yes
26. ฉันใช้คำเชื่อมที่หลากหลายหรือเปล่า Yes
27. ฉันใช้คำศัพท์ที่หลากหลายหรือเปล่า Yes
28. ฉันเลือกใช้คำที่เลือกสรรมาอย่างดี จำนวนที่เหมาะสม และ –
ศัพท์ที่ทำให้ผู้อ่านเห็นภาพ Yes
29. ฉันตรวจแก้ไวยากรณ์หรือยัง Yes
(Subject-verb/pronoun agreement, tense shift, parts of speech,
adjectives & adverbs, active/passive voice constructions, articles,
numbers (count and non-count), singular & plural forms, prepositions)
30. ฉันตรวจแก้ตัวสะกดหรือยัง Yes

31. ฉันตรวจแก้เครื่องหมายวรรคตอนหรือยัง Yes
32. ฉันตรวจแก้การเขียนตัวใหญ่ตัวเล็กหรือยัง Yes
33. ฉันตรวจแก้การจัดย่อหน้าหรือยัง Yes

Appendix B

The T-CBWT and CBWT contrasted

The T-CBWT and CBWT contrasted

Test taker-centered computer-based writing test (T-CBWT)	Computer-based writing test (CBWT)
The manipulated independent variable used as posttest	Used as pretest
<p>Description:</p> <ul style="list-style-type: none"> • Prompts parallel to CULI writing test & mirrors TOEFL's independent writing section • Employs the Microsoft Word program • Test takers' responses stored on computer • Administered, not scored via computer • Measures ability to write evaluative essays: one task type • Controlled topic & short input given • Approximately 350 words required • 90 minutes allowed • Either facilitative functions or required multiple drafts, or both, given • Track changes device required 	<p>Description:</p> <ul style="list-style-type: none"> • Prompts parallel to the T-CBWT • Employs the Microsoft Word program • Test takers' responses stored on computer • Administered, not scored via computer • Measures ability to write evaluative essays: one task type • Controlled topic & short input given • Approximately 350 words required • 90 minutes allowed • No facilitative functions or required multiple drafts • Track changes device not required

Appendix C

CBWT and T-CBWT Prompts

CBWT Pretest Prompt

A company has announced¹² that it wishes to build a large shopping center with movie theatres and a bowling alley right in your neighborhood¹³, very near your house. Do you support or oppose¹⁴ this plan for your community¹⁵? Why? Give at least **three** specific¹⁶ reasons to explain your opinion, including substantial¹⁷ examples and details¹⁸ in order to be convincing¹⁹. Make sure you address²⁰ all parts of the prompt.

T-CBWT Posttest Prompt:

It has recently²¹ been announced that street vendors²² selling food and all other items will from now on be permanently²³ banned²⁴ from selling on public streets and footpaths of Bangkok. Do you agree or disagree with this new restriction²⁵ on your community²⁶? Why? Give at least **three** specific reasons to explain your opinion, including substantial⁷ examples and details⁸ in order to be convincing⁹. Make sure you address¹⁰ all parts of the prompt.

¹² ประกาศ

¹³ หมู่บ้าน

¹⁴ ค้าน

¹⁵ ชุมชน

¹⁶ เฉพาะเจาะจง

¹⁷ แน่นหนา มากมาย

¹⁸ รายละเอียด

¹⁹หนักแน่นน่าเชื่อถือ

²⁰ ตอบ

²¹ เมื่อเร็ว ๆ นี้

²² พ่อค้าแม่ค้า รถเข็น หาบเร่ริมทาง

²³ อย่างถาวร

²⁴ ห้าม

²⁵ กฎข้อบังคับ

²⁶ ชุมชน

⁷ มากมาย แน่นหนา

⁸ รายละเอียด

⁹หนักแน่นน่าเชื่อถือ

¹⁰ ตอบ

Appendix D

Retrospective Questionnaire

Test Taker-Centered Computer-Based Writing Test Questionnaire

Part A. Facilitative Functions ในแบบสอบถามส่วนนี้ เราอยากทราบเกี่ยวกับลักษณะและวิธี การใช้ฟังก์ชันช่วยเหลือ ของนิสิต ในการสอบเขียนที่เพิ่งจบไป ขอขอบคุณสำหรับความร่วมมือ

วิธีทำ : กรุณาทำเครื่องหมาย (✓) ในช่องที่ตรงกับนิสิตที่สุด

Facilitative Functions Usage in the T-CBWT การใช้ฟังก์ชันช่วยในการสอบเขียนภาษาอังกฤษ	มากที่สุด > 16 ครั้ง	มาก 11-15 ครั้ง	น้อย 6-10 ครั้ง	น้อยที่สุด < 5 ครั้ง
1. ในการสอบครั้งนี้ นิสิตใช้ พจนานุกรม (translation) มากน้อยเพียงใด				
2. ในการสอบครั้งนี้ นิสิตใช้ พจนานุกรมคำพ้อง (thesaurus) มากน้อยเพียงใด				
3. ในการสอบครั้งนี้ นิสิตใช้ spell-check มากน้อยเพียงใด				
4. ในการสอบครั้งนี้ นิสิตใช้ grammar-check มากน้อยเพียงใด				
5. ในการสอบครั้งนี้ นิสิตใช้ Self-reflective Reminder Questions มากน้อยเพียงใด				

Part B. Opinions towards the T-CBWT

ในแบบสอบถามส่วนนี้ ผู้วิจัยอยากทราบเกี่ยวกับความเห็นโดยรวมของนิสิตที่มีต่อการสอบแบบนี้
วิธีทำ: นิสิตเห็นด้วยกับประโยคเหล่านี้มากน้อยแค่ไหน กรุณาทำเครื่องหมาย (✓) ในช่องที่ตรงกับนิสิตที่สุด

Opinions towards the T-CBWT ความเห็นต่อข้อสอบ T-CBWT	เห็นด้วย อย่างยิ่ง	เห็นด้วย	ไม่เห็นด้วย	ไม่เห็นด้วย อย่างยิ่ง
1. คำสั่งต่าง ๆ ของข้อสอบเขียนนี้ชัดเจนดี				
2. การสาธิตก่อนการลงมือทำข้อสอบเขียนนั้นชัดเจนดี				
3. ขั้นตอนต่าง ๆ ในการทำข้อสอบนี้ง่ายที่จะทำตาม				
4. ในการสอบข้อเขียนภาษาอังกฤษครั้งนี้ฉันชอบที่ได้ใช้คอมพิวเตอร์มากกว่าการใช้ปากกาเขียน				
5. ข้อสอบแบบนี้สามารถวัดความสามารถที่แท้จริงในการเขียนภาษาอังกฤษของฉันได้				
6. ในอนาคตฉันอยากจะสอบข้อเขียนภาษาอังกฤษในรูปแบบนี้				
7. ฉันคิดว่าฉันน่าจะเขียนได้ดีในการสอบข้อเขียนภาษาอังกฤษประเภทนี้ มากกว่าการสอบเขียนเขียนภาษาอังกฤษแบบดั้งเดิมที่ไม่มีฟังก์ชันหรือ การเขียนร่าง				

