CHAPTER IV
RESULTS

This chapter presents the results of the study in accordance with its research questions. The results are presented in 5 parts:

Part I: The effects of error treatments and language abilities of the students on the usage of English tenses. This part reports the results from the posttest concerning the effects of the two main factors and its interaction on the usage of tenses. Descriptive statistics and two-way ANOVA were employed to answer research questions 1, 2, and 3.

Part II: Retention. This part reports results from the comparison of the 3 observations– pretest, posttest, and delayed test in order to identify the retention in students. Descriptive statistics and two-way repeated measures ANOVA were utilized to provide the answer for research question 4.

Part III: The use of tenses by students with different abilities. This part presents the results from the pretest and the posttest in terms of the differences in the usage of English tenses by students with different language abilities. Which tenses are easy for low, moderate, and high achievers? What are the difficult ones? Are they the same or different? Percentage s were used to provide the additional answer for research question 2.

Part IV: Students' opinions. This part reports the students' opinions on the two types of error treatments that they received through the two CALL programs. Descriptive statistics, frequency, one-sample t-test, and Chi-square test of homogeneity (Fisher's exact test) were used to find the answer for research question 5.

Part V: Students' perceptions of their own learning. This part reports results from the self-rating scales. *Independent-samples t-test* was applied to test the differences between the SC and the OC groups. Then, *dependent-samples t-test* was used to examine the differences between pre- and post-treatment scales. The findings could be used to answer research question 5.

In this chapter, some abbreviations and statistical symbols are used as follows:

SC : self-correction

OC : overt correction

LA : low achievers

MA : moderate achievers

HA : high achievers
Min : minimum score
Max : maximum score
$\overline{X}$: Average or mean score
SD : standard deviation
df : degree of freedom
SS : sum of squares
p : critical value
ES : effect size
η_p^2 : partial Eta squared (to indicate the ES)
η^2 : Eta squared

Part I: The effects of error treatments and language abilities of the students on the usage of English tenses

After practicing with the CALL programs, the samples were posttested on the fifth week. First, basic statistics from the posttest are presented in Table 4.1. Mean, standard deviation, and range of scores are included.

Table 4.1: Means, standard deviations, and ranges of scores from the posttest

	Error treatment	n	Min	Max	Range	$\overline{X}$	SD
Low Achievers	Overt	35	5	34	29	15.51	5.43
(LA)	Self	35	8	31	23	15.71	5.16
	Total	70	5	34	29	15.61	5.26
Moderate Achievers	Overt	35	8	37	29	22.86	7.43
(MA)	Self	35	11	42	31	22.71	8.10
	Total	70	8	42	34	22.79	7.72
High Achievers	Overt	35	20	46	26	34.57	6.18
(HA)	Self	35	21	50	29	34.86	7.19
	Total	70	20	50	30	34.71	6.66

It can be seen from Table 4.1 that all groups had high standard deviation. This means that they had more variability from the central point in the distribution. One

significant point to be made here is that one of the students in the HA with SC group could reach the perfect 50 score from the posttest.

In order to further investigate the main effects and the interaction effect to answer research questions 1-3, two-way ANOVA would be employed. First, the researcher had to make sure that the basic assumptions of ANOVA were not violated. Results from the test of multivariate normality (Appendix M) showed that all groups had normal score distributions and the test of homogeneity of variances revealed a small but acceptable departure from homogeneity of variances (see results and discussions in Appendix N). Then, two-way ANOVA could be employed. Results from the Two-way ANOVA are presented in Table 4.2.

Table 4.2: Effects of error treatments and language abilities on the usage of tenses

Source	Sum of Squares	df	Mean Square	*F*
Between groups				
Error treatments	.686	1	.686	.015
Language abilities	13032.371	2	6516.186	146.525*
Error treatments x Language abilities	1.800	2	.900	.020
Within groups	9072.171	204	44.471	
Corrected Total	22107.029	209		

* $p < .05$

The main effects:

Error Treatments Factor: Results from the two-way ANOVA revealed that the *F* value of this factor was .015 and its *p* value was .45 (one-tailed). The *p* value from the calculation was greater than the critical value set at .05, indicating that the proposed directional hypothesis (hypothesis 1) was rejected. This means that students using self-correction program did not make a significantly higher mean score than those using overt correction program. In fact, they made similar gains overall (Table 4.1).

Language Abilities Factor: The *F* value for language abilities was 146.525. Its *p* value was equal .001 (two-tailed) which was lower than the critical value set at .05. Therefore, the null hypothesis for 'language ability' factor is rejected and the non-directional alternative hypothesis (hypothesis 2) was accepted. This means that language abilities factor had a significant effect on the usage of tenses in the posttest. In other words, language abilities were the significant source of the difference in the gain scores.

Since there were three ability groups in the study, a post-hoc comparison was computed to identify where the differences occurred. The researcher has applied Tukey's method because the number of subjects in each group was equal (n = 35). Results are presented in Table 4.3.

Table 4.3: Results from a post-hoc comparison of the three ability groups

	(I) Ability	(J) Ability	Mean Difference (I-J)	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Tukey HSD	Low	Moderate	-7.17*	1.127	-9.83	-4.51
		High	-19.10*	1.127	-21.76	-16.44
	Moderate	Low	7.17*	1.127	4.51	9.83
		High	-11.93*	1.127	-14.59	-9.27
	High	Low	19.10*	1.127	16.44	21.76
		Moderate	11.93*	1.127	9.27	14.59

* $p < .05$

Results from the post-hoc comparison point out that the subjects in different ability groups all performed differently. The high ability group performed differently from the moderate and the low ability ones. Also, the moderate ability group performed differently from the low group. To be more specific, the high group performed better than the moderate group and the moderate group performed better than the low group.

The interaction effect:

Theoretically, if an interaction is significant, it means that while one method of instruction did work better than the other this may be due to the second factor (Hatch & Farhady, 1982:158). An interaction effect exists when differences on one factor depend on the level you are on another factor. In this case, the *F* –ratio for the interaction of 'error treatments' and 'language abilities' on overall gain scores was .020. Its *p* value was equal to .98 (two-tailed) which was greater than the .05 critical value. As a result, the null hypothesis for the interaction effect was accepted, meaning that no significant interaction effect of 'error treatments' and 'language abilities' on the usage of English tenses existed. This can be shown graphically as in Figure 4.1.

Figure 4.1: Plots of mean scores from the posttest

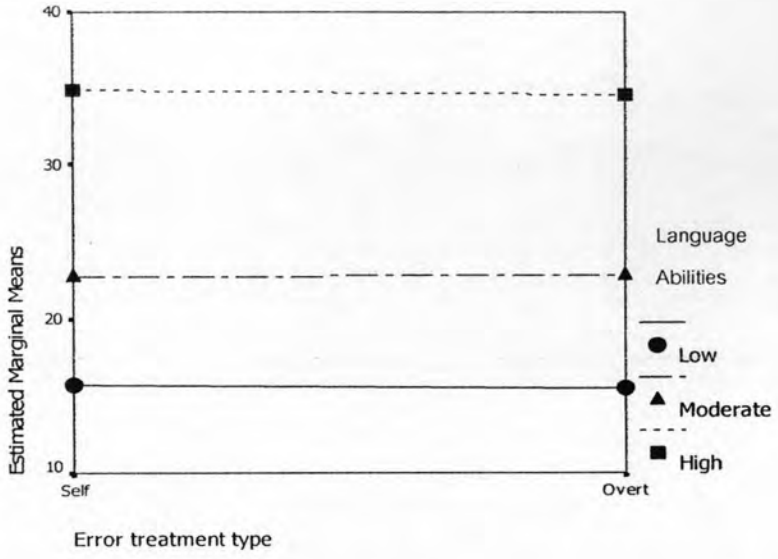


Figure 4.1 makes it easier to explain what has happened between the levels of the two main factors. Normally, if the interaction effect exists, there would be a cross between the lines plotted from the mean scores. The plot in Figure 4.1 showed that there was no cross between any pairs of the three lines. It was obvious that high achievers made greater gains than moderate achievers, and moderate achievers made greater gains than low achievers. Therefore, regardless of different types of error treatment they received, the subjects made similar improvement in terms of the posttest scores. This confirms that there was no interaction effect between 'error treatments' and 'language abilities' on the usage of English tenses in the posttest.

The effect size and the observed power

In addition to the *statistical significance* (judged by the *p* value) that was presented above, research should also report its *practical significance* in forms of the *effect size*. Effect size (ES) is a name given to a family of indices that measure the magnitude of a treatment effect (Becker, 2000a). There are many different types of ES measures that are suitable for different types of research. Measures of effect size in ANOVA are measures of the degree of association between an effect and the dependent variable (Becker, 2000b). Four of the commonly used measures of ES in ANOVA are:

Eta squared, partial Eta squared, omega squared, and the Intraclass correlation. SPSS package can display the partial Eta squared when requested.

Becker (2000b) explains that partial Eta squared (η_p^2) is the proportion of the effect + error variance that is attributable to the effect. Eta square (η^2) is, on the other hand, the proportion of the total variance that is attributed to an effect. The formulas for both estimates are as follows:

$$\eta_p^2 = SS_{effect} / (SS_{effect} + SS_{error})$$

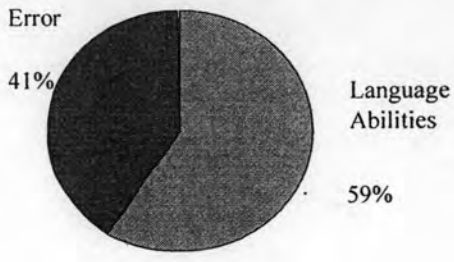
$$\eta^2 = SS_{effect} / SS_{total}$$

Table 4.4: Effect size and observed power of the error treatments and the language abilities factors

Effect	SS_{effect}	SS_{error}	SS_{total}	η^2	η_p^2	Observed power
Error treatments	.686	9072.171	22107.029	.001	.001	.052
Language abilities	13032.371	9072.171	22107.029	.590	.590	1.00
Error treatments x Language abilities	1.800	9072.171	22107.029	.001	.001	.053

Results from the calculation presented in Table 4.4 reveal that the 'error treatments' factor and the interaction did not have a significant effect size because they did not have significant effects on the posttest scores as mentioned earlier. Only the 'language abilities' showed the effect size at .590. Results from the Eta and the partial Eta formula were equal because the magnitudes of the other two effects (error treatments and interaction) were not significant. Graphic representation of the effect size is illustrated in Figure 4.2.

Figure 4.2: Graphic representation of the effect size



Neither 'error treatment' nor 'the interaction' had a significant effect on the posttest gains. The 'language abilities' factor was the only one significant factor that had an effect on the posttest gains. The magnitude of η^2 was 0.59, which means that the 'language abilities' factor by itself accounted for 59% of the overall variance. Cohen indicates the way to relate η^2 to Cohen's effect size as follows (Barnette, 2006: slide no. 89; Wuensch, 2005):

Eta-squared of .0099 relates to a Cohen "small effect" (.2)

Eta-squared of .0588 relates to a Cohen "medium effect" (.5)

Eta-squared of .1379 relates to a Cohen "large effect" (.8)

Then, the η^2 of .59 in the present study can be interpreted as a "very large effect" according to the above guidelines.

The Observed Power: Power is the ability to detect an effect if there is one (UCLA, 2007). The power ranges from 0 to1. Generally, the power of .80 is considered acceptable (UCLA, Ibid.). Table 4.4 shows that the observed power of 'language abilities' factor equals 1.00 which is the highest possible value. This means that there is no chance of failing to detect the effect. In other words, it means that the reliability of the findings is very high; so the findings can be trusted.

Part I summary: This part reports results from the posttest. The results provide answers for research questions 1-3 as follows:

1. Different types of error treatment given through the CALL did not have a significantly different effect on the usage of English tenses of the students in the posttest.

2. The 'language abilities' factor had a significant effect on the usage of English tenses in the posttest. Students in different ability groups all performed differently. The magnitude of the effect size for this factor was large. Fifty-nine percent of the overall variance could be explained by this factor.

3. There was no interaction effect between 'error treatments' and 'language abilities' on the usage of English tenses in the posttest. High ability students in both groups performed better than the moderate ability groups which performed better than the low ability groups, regardless of the types of error treatment they received.

Part II: Retention

This part reports results from the comparison of the 3 observations– pretest, posttest, and delayed test in order to identify the retention in students. This part will provide the answer to research question 4. Firstly, descriptive information from the three observations is presented in Table 4.5.

Table 4.5: Means and standard deviations of the two treatment groups from the three observations

	Error Treatments	n	$\overline{X}$	SD
PRETEST	Self	105	20.48	8.56
	Overt	105	20.39	8.39
	Total	210	20.43	8.46
POSTTEST	Self	105	24.43	10.50
	Overt	105	24.31	10.12
	Total	210	24.37	10.29
DELAYED POSTTEST	Self	103	23.80	9.72
	Overt	99	23.05	10.24
	Total	202	23.43	9.96

There were some dropouts of subjects after their midterm examination. Consequently, 202 students remained in the delayed test. Table 4.5 reveals that the students in both OC and SC groups did best in the posttest. They got relatively lower scores in the delayed test. The lowest mean scores were those from the pretest.

To find answers about the retention of the two error treatment types over time, two-way repeated measures ANOVA has been applied. The first step was to examine the 'sphericity' to determine if the homogeneity of variance assumption is met. The sphericity assumption is similar to the homogeneity of variance assumption with between subjects ANOVA. Normally, SPSS program automatically provides *Mauchly's test of Sphericity* value. The 'test' factor had *Mauchly's W* value of .878 ($p < .05$) which was significant (see results and discussions in Appendix O). Therefore, *Huynh-Fedlt* correction was the best choice when looking at the values from the 'tests of within-subjects effects' table. Table 4.6 presents results of two-way repeated measures ANOVA from the Huynh-Fedlt correction test.

Table 4.6 Results from the two-way repeated measures ANOVA

Source	Sum of Squares	df	Mean Square	*F*	η^2	η_p^2	Obs. Power
Between-subjects effects							
Error treatments (A)	.339	1	.339	.001	.001	.001	.050
Error (Subjects or S)	49435.254	200	247.176				
Within-subjects effects							
Tests (B)	1638.634	1.806	907.439	50.546*	.201	.202	1.000
Err Treatments x Tests (AB)	36.931	1.806	20.452	1.139	.005	.006	.239
Error (SB)	6483.670	361.156	17.953				

* $p < .05$

Between-subjects effects:

The main effect for 'error treatments' was not significant, $F(1, 200) = .001, p > .05$. The results means that types of error treatment the students received, as a standalone factor, were not the remarkable source of differences in the gain scores.

Within-subjects effects:

The tests of within-subjects effects indicated that the 'error treatments by tests' interaction effect was not significant, $F(1.806, 361.156) = 1.139, p > .05$. The findings reveal that there was not a significant interaction effect between 'error treatments' and 'tests' on the mean scores. In other words, high ability groups always performed better than the moderate groups which always performed better than the low groups regardless of the test time (whether in pretest, posttest, and delayed test).

Considering the effect of the 'tests' factor, results from Table 4.6 indicated a significant effect, $F(1.806, 361.156) = 50.546, p < .05$. It means that there was at least one pair of the mean scores from the pretest, posttest, and delayed test that was significantly different from one another. The factor had the effect size at .201 calculated by Eta-squared formula, meaning that this factor accounted for 20.1% of the overall variance. Relating the Eta-squared to Cohen's effect size, the magnitude can be interpreted as a 'large' effect. Its observed power at 1.00 indicates no chance of failing to detect the existing effect.

To further identify the differences among the three observations, the *pairwise comparisons* are utilized. The results are presented in Table 4.7.

Table 4.7: The pairwise comparisons

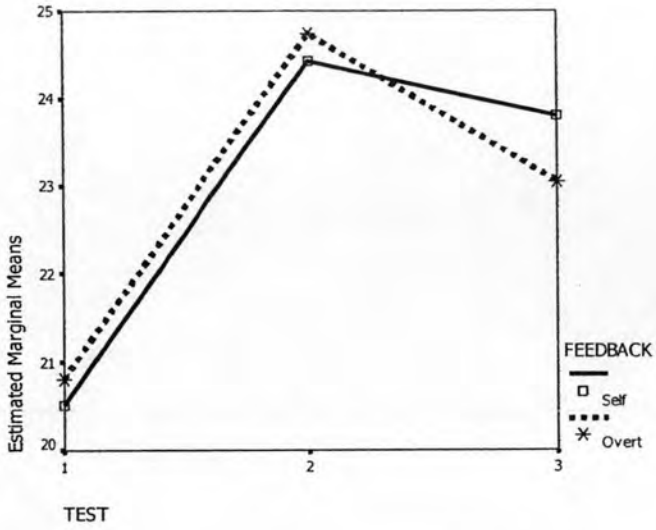
Error treatment	(I) TEST	(J) TEST	Mean Difference (I-J)	Std. Error
Overt	1	2	-3.929*	.627
		3	-2.242*	.678
	2	1	3.929*	.627
		3	1.687*	.480
	3	1	2.242*	.678
		2	-1.687*	.480
Self	1	2	-3.913*	.600
		3	-3.282*	.550
	2	1	3.913*	.600
		3	.631	.434
	3	1	3.282*	.550
		2	-.631	.434

* $p < .05$

For overt correction treatment, all mean differences differed significantly. Students' mean score of the pretest (Test 1) was remarkably different from that of the posttest (Test 2) and that of the delayed test (Test 3). Also, mean score of the posttest (Test 2) was significantly different from that of the delayed test (Test 3). This part was interesting because of the retention which was the difference between mean scores of test 2 ($\bar{x}_{t2} = 24.74$) and test 3 ($\bar{x}_{t3} = 23.05$). The mean difference was equal to -1.687 ($p < .05$), indicating that the difference was significant. Figure 4.3 illustrates the changes of mean scores from the three tests. It can be seen that the mean scores of the overt correction group dropped outstandingly from posttest to delayed test.

Unlike the overt correction, the self-correction treatment did not have a significant difference between mean score of the posttest ($\bar{x}_{t2} = 24.43$), and that of the delayed test ($\bar{x}_{t3} = 23.80$). The mean difference from Table 4.7 was - .631 ($p > .05$). It can be said that the mean score of self-correction group dropped insignificantly.

Figure 4.3: Plots of mean scores from the three observations



The plots show a large decrease in the mean scores for the overt correction treatment group from posttest (test2) to delayed test (test 3). The change for the self-correction treatment group from posttest (test2) to delayed test (test 3) was smaller and insignificant. It can be said from the results that students in the self-correction group could better maintain their knowledge than students in the overt correction group in the delayed test.

Part II summary: This part compares results from the pretest, posttest and delayed test to find an answer for research question 4 that asks about the retention by using two-way repeated measures ANOVA. The results could be summarized as follows:

1. There was no interaction effect between 'error treatments' and 'tests' found.

2. The between-subjects factor 'error treatments' did not have a significant effect on the gain scores while the within-subjects factor 'tests' did.

3. The decrease of scores for the self-correction treatment group from the posttest (test2) to the delayed test (test 3) was small and insignificant while the decrease of scores for the overt correction treatment group was greater and significant. It means that the self-correction treatment could promote better retention than overt correction.

Part III: The Use of Tenses by Students with Different Abilities

This part presents the results from the pretest, the posttest, and the delayed test in order to answer research question 2 about how students with different abilities used English tenses. One-way ANOVA, post-hoc comparisons (Sheffe test), and percentage were used to provide the answers.

Results from the pretest

To answer the questions about the differences in terms of the usage of English tenses by different ability groups; first, items that tested the same tense were put together (Table 4.8). Then mean scores of the three ability groups for each tense were calculated and then were compared by one-way ANOVA. Results are presented in Table 4.9.

Table 4.8: Test items grouped by tenses

Tense	Item No.
1. Simple Present Tense (11 items)	2,3,7,10,22,29,33,35,36,37,41
2. Simple Past Tense (9 items)	8,13,18,23,32,43,44,47,49
3. Present Continuous Tense (6 items)	6,17,19,26,31,40
4. Present Perfect Tense (9 items)	5,9,12,28,38,42,45,46,48
5. Future Tenses (6 items)	14,15,20,27,39,50
6. Past Continuous Tense (5 items)	1,4,25,30,34
7. Past Perfect Tense (4 items)	11,16,21,24

Table 4.9: Comparisons of mean scores made by different ability groups in the pretest

Tense	Source	Sum of Squares	df	Mean Square	*F*
Simple Present Tense	Between Groups	1013.898	2	506.949	173.541*
	Within Groups	590.082	202	2.921	
	Total	1603.980	204		
Simple Past Tense	Between Groups	614.109	2	307.054	162.076*
	Within Groups	382.691	202	1.895	
	Total	996.800	204		
Present Continuous Tense	Between Groups	118.328	2	59.164	39.211*
	Within Groups	304.794	202	1.509	
	Total	423.122	204		
Present Perfect Tense	Between Groups	112.779	2	56.389	23.521*
	Within Groups	484.265	202	2.397	
	Total	597.044	204		
Present Perfect Tense	Between Groups	194.429	2	97.215	71.240*
	Within Groups	275.649	202	1.365	
	Total	470.078	204		
Past Continuous Tense	Between Groups	174.234	2	87.117	57.403*
	Within Groups	306.566	202	1.518	
	Total	480.800	204		
Past Perfect Tense	Between Groups	33.965	2	16.982	17.177*
	Within Groups	199.713	202	.989	
	Total	233.678	204		

* $p < .05$

It should be noted that there were some students who did not answer some questions, so they were excluded by the SPSS program when computing for the one-way ANOVA. The numbers of students in the low, moderate, and high groups were 70, 67, and 68 respectively. This means that the results come from the group of 205 students.

Table 4.9 reveals that all of the *F* values for every tense were significant. This means that there was a significant difference among the ability groups in using all the tenses. The largest *F* values were from the simple present tense and the simple past tense ($F = 173.541$ and 162.076) indicating the largest differences of mean scores. To identify where the differences occurred, Scheffe Test which does not have the assumption of equal sample size was applied. Results are presented in Table 4.10.

Table 4.10 Mean scores of the three ability groups from the pretest categorized by Scheffe Test

Tense	No. of items in each tense	Subset		
		Group 1	Group 2	Group 3
Simple Present	11	LA (3.31)	MA (6.06)	HA (8.74)
Simple Past	9	LA (1.39)	MA (2.73)	HA (5.53)
Present Cont.	6	LA (2.34)	MA (3.15)	HA (4.19)
Present Perfect	9	LA (1.79) MA (2.28)	HA (3.54)	
Future Tenses	6	LA (1.19)	MA (2.42)	HA (3.56)
Past Cont.	5	LA (0.91) MA (1.43)	HA (3.07)	
Past Perfect	4	LA (0.73)	MA (1.24)	HA (1.72)

Figures in the parentheses were the mean scores. Results from Scheffe Test (Table 4.10) reveal that there were 5 tenses that the 3 ability groups all performed differently at the .05 significance level. They were: simple present, simple past, present continuous, future tenses, and past perfect. This means that students with different abilities were really different in terms of their ability to answer questions about these tenses. However, the mean scores of the LA and the MA groups for the remaining two tenses (present perfect and past continuous) were not significantly different; so they were grouped together, meaning that the LA and the MA groups could answer the questions about these two tenses similarly.

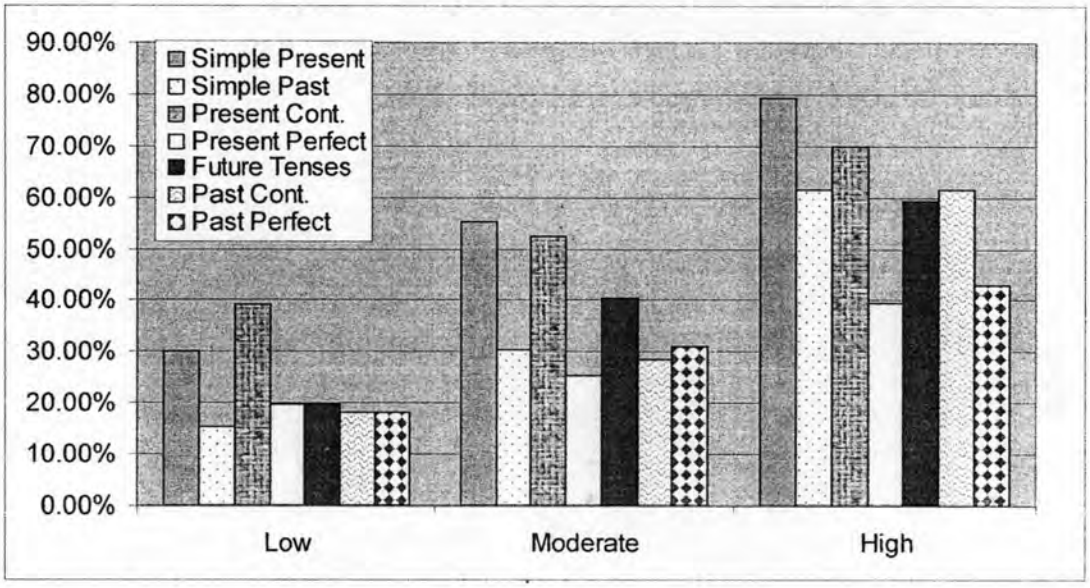
Next, the means of the seven tenses made by different ability groups were calculated. Since the number of items in each tense differs, the means were converted to percentages in order that they could be compared. Table 4.11 reports the percentages of the means.

Table 4.11: Percentages of the means for each tense from the pretest

Language Abilities	Simple Present	Simple Past	Present Cont.	Present Perfect	Future Tenses	Past Cont.	Past Perfect
Low	30.13 %	15.40 %	39.05 %	19.84 %	19.76 %	18.29 %	18.22 %
Moderate	55.09 %	30.35 %	52.49 %	25.37 %	40.30 %	28.66 %	30.97 %
High	79.41 %	61.44 %	69.85 %	39.38 %	59.31 %	61.47 %	43.02 %

The percentages from Table 4.11 reveal the information about how well the students in each ability group could answer the tests. In other words, it provides the information about the difficult as well as the easy tenses for each ability group. The higher the percentage, the easier the tense was. Results can be graphically illustrated as shown in Figure 4.4.

Figure 4.4: Percentages of the means for each tense from the pretest



From the pretest which is a natural setting, it was found that the two tenses that students in all ability groups could perform best were simple present and present continuous. MA and HA students did best in simple present and LA students did best in present continuous.

The most difficult tense for LA students is simple past, followed by past perfect and past continuous. MA students found present perfect the most difficult, followed by past continuous and simple past. As for the HA group, present perfect and past perfect were their weakness.

Results from the posttest

Data from the posttest were analyzed in the same way. The number of subjects who did not have any missing values and were then included in the calculation was 198 (LA = 60, MA =73, and HA = 65). First, one-way ANOVA was applied to examine whether the differences existed. Results are illustrated in Table 4.12.

Table 4.12: Comparisons of mean scores made by different ability groups in the posttest

Tense	Source	Sum of Squares	df	Mean Square	*F*
Simple Present Tense	Between Groups	769.373	2	384.687	68.790*
	Within Groups	1090.470	195	5.592	
	Total	1859.843	197		
Simple Past Tense	Between Groups	431.121	2	215.560	66.747*
	Within Groups	629.753	195	3.230	
	Total	1060.874	197		
Present Continuous Tense	Between Groups	88.652	2	44.326	25.248*
	Within Groups	342.343	195	1.756	
	Total	430.995	197		
Present Perfect Tense	Between Groups	239.486	2	119.743	26.660*
	Within Groups	875.828	195	4.491	
	Total	1115.313	197		
Present Perfect Tense	Between Groups	188.710	2	94.355	44.521*
	Within Groups	413.270	195	2.119	
	Total	601.980	197		
Past Continuous Tense	Between Groups	201.474	2	100.737	44.879*
	Within Groups	437.703	195	2.245	
	Total	639.177	197		
Past Perfect Tense	Between Groups	51.617	2	25.809	20.483*
	Within Groups	245.701	195	1.260	
	Total	297.318	197		

* $p < .05$

Table 4.12 reveals that all of the *F* values for every tense were significant. This means that there was a significant difference among the ability groups in using all the tenses in the posttest. Similar to the results from the pretest, the largest *F* values were from the simple present tense and the simple past tense ($F = 68.790$ and 66.747) indicating the largest differences of mean scores. However, when comparing the two sets of *F* values, one from the pretest and another from the posttest, it can be seen that they were quite different. *F* values from the pretest were much greater than those from the posttest. This means that, in the posttest, the differences of the mean scores of the three ability groups were smaller. When applying Scheffe Test to identify where the differences occurred, results were as presented in Table 4.13.

Table 4.13 Mean scores of the three ability groups from the posttest categorized by Scheffe Test

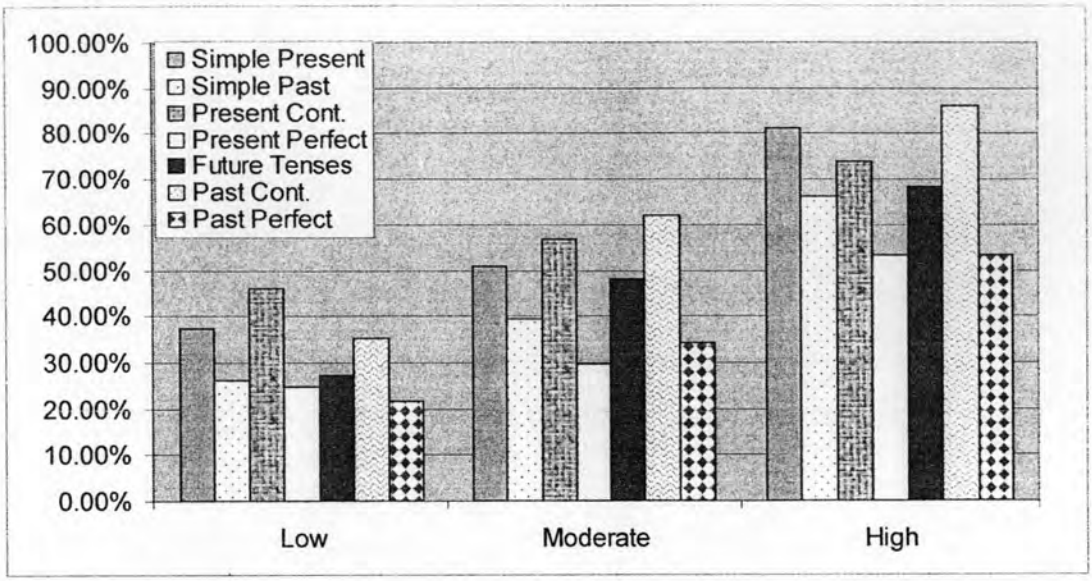
Tense	No. of items in each tense	Subset		
		Group 1	**Group 2**	**Group 3**
Simple Present	11	LA (4.12)	MA (5.59)	HA (8.92)
Simple Past	9	LA (2.35)	MA (3.53)	HA (5.97)
Present Cont.	6	LA (2.77)	MA (3.41)	HA (4.43)
Present Perfect	9	LA (2.23) MA (2.67)	HA (4.78)	
Future Tenses	6	LA (1.63)	MA (2.88)	HA (4.09)
Past Cont.	5	LA (1.77)	MA (3.11)	HA (4.31)
Past Perfect	4	LA (0.87)	MA (1.37)	HA (2.14)

It was found that in the posttest the three ability groups all performed differently at the .05 significance level in almost all tenses, except in the present perfect in which the LA and the MA groups performed similarly in terms of statistics. Table 4.14 reports the percentages of the items that the subjects answered correctly in each tense. Figure 4.5 depicts the percentages graphically.

Table 4.14: Percentages of the means for each tense from the posttest

Language Abilities	Simple Present	Simple Past	Present Cont.	Present Perfect	Future Tenses	Past Cont.	Past Perfect
Low	37.42 %	26.11 %	46.11 %	24.81 %	27.22 %	35.33 %	21.67 %
Moderate	50.81 %	39.27 %	56.85 %	29.68 %	47.95 %	62.19 %	34.25 %
High	81.12 %	66.32 %	73.85 %	53.16 %	68.21 %	86.15 %	53.46 %

Figure 4.5: Percentage of the means for each tense from the posttest



The first point that can be seen from Table 4.14 is that all figures from the posttest were greater that those of the pretest, except the performance of the MA group in simple present. It means that, in general, students in every ability group performed every tense better than they did in the pretest.

Overall, the bar chart of the posttest (Figure 4.5) did not look much different from that of the pretest (Figure 4.4). A notable change, however, was that past continuous became one of the top three in all ability groups. Present perfect and past perfect were still the problematic tenses for all groups, followed by simple past.

Results from the delayed test

As an attempt to figure out how students with different abilities used the tenses in the delayed test, the data were analyzed in the same way. The number of subjects who did not have any missing values and were then included in the calculation was 190 (LA = 60, MA =66, and HA = 64). First, one-way ANOVA was applied to examine whether the differences existed. Results are illustrated in Table 4.15.

Table 4.15: Comparisons of mean scores made by different ability groups in the delayed test

Tense	Source	Sum of Squares	df	Mean Square	*F*
Simple Present Tense	Between Groups	585.435	2	292.717	62.420*
	Within Groups	876.928	187	4.689	
	Total	1462.363	189		
Simple Past Tense	Between Groups	416.331	2	208.166	68.823*
	Within Groups	565.611	187	3.025	
	Total	981.942	189		
Present Continuous Tense	Between Groups	93.478	2	46.739	24.178*
	Within Groups	361.491	187	1.933	
	Total	454.968	189		
Present Perfect Tense	Between Groups	357.828	2	178.914	49.616*
	Within Groups	674.324	187	3.606	
	Total	1032.153	189		
Present Perfect Tense	Between Groups	192.050	2	96.025	43.966*
	Within Groups	408.418	187	2.184	
	Total	600.468	189		
Past Continuous Tense	Between Groups	148.823	2	74.412	39.899*
	Within Groups	348.756	187	1.865	
	Total	497.579	189		
Past Perfect Tense	Between Groups	18.405	2	9.202	9.132*
	Within Groups	188.438	187	1.008	
	Total	206.842	189		

* $p < .05$

Table 4.15 shows that all *F* values were significant, indicating that there was at least one pair of the means that was significantly different. Scheffe Test was utilized to identify such differences. Results from Scheffe test are presented in Table 4.16.

Table 4.16 Mean scores of the three ability groups from the delayed test categorized by Scheffe Test

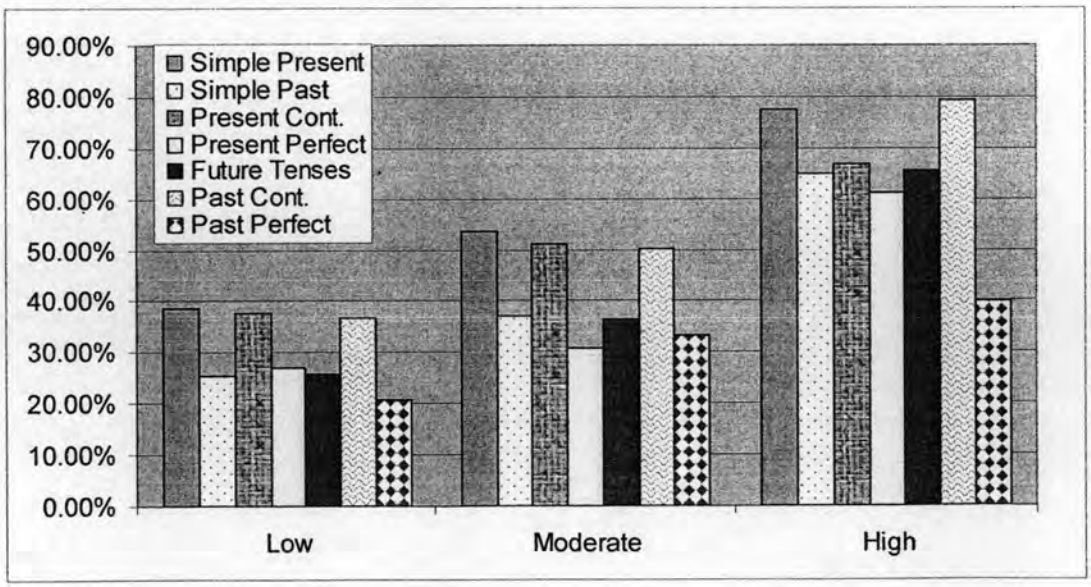
Tense	No. of items in each tense	Subset		
		Group 1	Group 2	Group 3
Simple Present	11	LA (4.23)	MA (5.89)	HA (8.53)
Simple Past	9	LA (2.28)	MA (3.32)	HA (5.83)
Present Cont.	6	LA (2.27)	MA (3.06)	HA (4.00)
Present Perfect	9	LA (2.43) MA (2.77)	HA (5.50)	
Future Tenses	6	LA (1.55) MA (2.18)	HA (3.94)	
Past Cont.	5	LA (1.83)	MA (2.52)	HA (3.97)
Past Perfect	4	LA (0.83)	MA (1.33) HA (1.59)	

Table 4.16 illustrates that there were 4 tenses that the 3 ability groups all performed differently at the .05 significance level. They were: simple present, simple past, present continuous, and past continuous. This means that students with different abilities were really different in terms of their ability to answer questions about these tenses. However, the mean scores of the LA and the MA groups for the present perfect and the future tenses were not significantly different; so they were grouped together, meaning that the LA and the MA groups could answer the questions about these two tenses similarly. Similarly, the ability to answer questions in the past perfect of the MA and the HA groups were not significantly different. Table 4.17 reports the percentages of the items that the subjects answered correctly in each tense. Figure 4.6 depicts the percentages graphically.

Table 4.17: Percentages of the means for each tense from the delayed test

Language Abilities	Simple Present	Simple Past	Present Cont.	Present Perfect	Future Tenses	Past Cont.	Past Perfect
Low	38.48%	25.37%	37.78%	27.04%	25.83%	36.67%	20.83%
Moderate	53.58%	36.87%	51.01%	30.81%	36.36%	50.30%	33.33%
High	77.56%	64.76%	66.67%	61.11%	65.63%	79.38%	39.85%

Figure 4.6: Percentages of the means for each tense from the delayed test



The graphic looked very similar to the results from the posttest. The first three prominent tenses that all ability groups performed best were simple present, present continuous, and past continuous. The most difficult tense for the LA and the HA groups was the past perfect tense while the most difficult tense for the MA group was the present perfect.

Part III summary: This part reports results from the pretest, the posttest and the delayed test to identify the differences in terms of an ability to use different tenses by different ability groups. The summary of the results is as follows:

1. The tenses that all ability groups used differently in all of the three tests were simple present, simple past, and present continuous. However, the LA group used some tenses without significant difference from the MA group; for example, the use of past continuous tense in the pretest, the use of future tenses in the delayed test, and the use of present perfect in all tests.

2. From the pretest, the notable tenses used correctly by the LA group were present continuous and simple present. The MA group could perform best in simple present, present continuous, and future tenses. HA students could prominently use the following 5 tenses in order: simple present, present continuous, past continuous, simple past, and future tenses.

3. From the posttest, results were not much different from the pretest, except that the past continuous became one of the top three tenses used correctly by all ability groups. Present perfect and past perfect were found to be the most difficult tenses for all groups, followed by simple past.

4. From the delayed test, the top three tenses correctly used by all ability groups were simple present, present continuous, and past continuous. The most difficult tense for the LA and the HA groups was the past perfect tense while the most difficult tense for the MA group was the present perfect

Part IV: Students' Opinions

There were 189 students who answered the opinionnaire (Appendix K); 93 were from the OC group and 96 from the SC group. By examining the mean scores from Table 4.19, it was found that, in general, both OC and SC groups had highly positive opinions towards the program they used with the means of 3.91 and 4.01 respectively. This

means that the SC group rated the opinionnaire slightly more positively than the OC group.

As the present study set criteria that a highly positive opinionnaire item should be rated ≥ 3.5 from the 5-point scales (see Chapter 3), the first step was to test whether the mean of each item met the criteria. To do that, one-sample t-test was applied. Results are presented in Table 4.18.

Table 4.18: Results from the one-sample t-test

Error treatments	Item No.	Test Value = 3.5				
		t	df	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Overt	Item 1	12.523*	92	.81	.68	.94
	Item 2	8.727*	92	.63	.49	.77
	Item 3	11.779*	92	.80	.67	.94
	Item 4	8.104*	92	.53	.40	.66
	Item 5	7.032*	92	.40	.29	.52
	Item 6	6.602*	92	.49	.34	.64
	Item 7	6.873*	92	.46	.32	.59
	Item 8	-.639	92	-.05	-.20	.10
	Item 9	.251	92	.02	-.11	.14
	Item 10	9.158*	92	.53	.42	.65
	Item 11	7.164*	92	.44	.31	.56
	Item 12	5.118*	92	.35	.21	.49
	Item 13	-.699	92	-.05	-.19	.09
Self	Item 1	14.198*	95	.91	.78	1.03
	Item 2	10.136*	95	.70	.56	.83
	Item 3	15.000*	95	.94	.81	1.06
	Item 4	9.302*	95	.65	.51	.78
	Item 5	8.196*	95	.54	.41	.67
	Item 6	6.474*	95	.48	.33	.63
	Item 7	7.111*	95	.51	.37	.65
	Item 8	.307	95	.02	-.11	.16
	Item 9	1.990*	93	.13	.00	.26
	Item 10	9.022*	95	.64	.50	.78
	Item 11	8.819*	95	.57	.44	.70
	Item 12	6.680*	95	.47	.33	.61
	Item 13	2.877*	95	.21	.06	.35

* $p < .05$

Table 4.18 reveals that 10 out of 13 items (76.92%) from the OC group could pass the criteria. The three items that OC students rated lower than the criteria were items 8, 9, and 13. Item 8 and item 13 asked about the students' attention when

practicing while item 9 was part of the questionnaire that asked about their attitudes towards the program. Like the OC group, the SC group also rated item 8 lower than the criteria. However, the remaining 12 items (92.31%) were rated higher and then passed the criteria set.

To identify the differences between the responses of the OC group and those of the SC group, a Chi-square test of homogeneity of proportion was employed. Since there were some cells that had the frequency of 5 or less, Fisher's Exact Test was applied for the correction (Tirakanant, 2003). Results from the Exact test are presented in Table 4.19.

Table 4.19: Frequency, means, and results from Chi-square test of the opinionnaire responses

No.	Traits	Opinions (Frequency)										χ^2	$\bar{X}$	
		5		4		3		2		1				
		OC	SC	OC	SC	OC	SC	OC	SC	OC	SC		OC	SC
	Contents and Design													
1	Easy to use	37	46	48	43	8	7	0	0	0	0	1.296	4.31	4.41
2	The instruction is comprehensible	28	33	50	49	14	14	1	0	0	0	1.330	4.13	4.20
3	Font size and color are appropriate	38	47	45	45	10	3	0	1	0	0	5.534	4.30	4.44
4	Number of items in each lesson is appropriate	20	30	56	50	17	16	0	0	0	0	2.320	4.03	4.15
5	The difficulty suits the learners' levels	10	22	64	56	19	18	0	0	0	0	5.003	3.90	4.04
6	The program is motivating.	23	24	46	46	24	26	0	0	0	0	0.076	3.99	3.98
7	The program helps me understand about 'tenses'.	17	23	55	52	21	20	0	1	0	0	1.903	3.96	4.01
8	I always pay attention when I practice with the program.	7	5	33	44	49	43	3	4	1	0	3.387	3.45	3.52
9	I think I know more about tenses after practicing with the program.	4	5	42	51	45	36	2	2	0	0	2.135	3.52	3.55
	Feedback													
10	The feedback is useful and helpful.	16	30	64	49	13	17	0	0	0	0	6.713*	4.03	4.14
11	The feedback is easy to understand.	12	23	64	57	16	16	1	0	0	0	4.736	3.94	4.07
12	The feedback makes me understand the contents better.	14	20	51	54	28	21	0	1	0	0	3.017	3.85	3.97
13	I always read the feedback and the explanation with full attention.	6	13	33	43	51	39	3	1	0	0	6.313	3.45	3.71
												Grand Mean	3.91	4.01

* $p < .05$

Results from Table 4.19 show that, overall, the proportions of the responses of the OC and the SC groups were not significantly different. There was only one item for which the proportions of answers were significantly different, that is item 10. The valid proportion of the SC group was 30:49:17 (approx.1:1.6:0.56) while that of the OC group was 16:64:13 (approx.1:4:0.8). Thirty students (31.25%) in the SC group 'strongly agree' with the statement of item 10 saying 'the feedback is useful and helpful' while only 16 students (17.20%) in the OC group do.

Part IV summary: This part reports results from the opinionnaire which was distributed to the students at the end of the study. There were 13 items in the opinionnaire. The results were analyzed by one-sample t-test, descriptive statistics, and Chi-square test (Fisher's Exact test) to provide answers for research question 5. The results could be summarized as follows:

1. Overall, students in both OC and SC groups had highly positive opinions towards the program that they used. When the means were tested by one-sample t-test, it was found that SC students had highly positive opinions towards the SC program in almost all items, except for one item that asked about their attention when practicing. On the other hand, the OC group had 3 items that they rated lower than the criteria of 3.5.

2. Results from Chi-square test by Fisher's Exact test revealed that the proportions of the answers of the OC and those of the SC groups were not significantly different, except item 10 that contained the statement 'the feedback is useful and helpful'. The raw data point out that the difference was due to the fact that the SC group rated the item more positively.

Part V: Students' Perceptions of Their Own Learning

This part reports results from the self-rating scales that were distributed twice– before and after the treatment. Figure 4.7 and Table 4.20 present the descriptive information about the pre-treatment scales and the post-treatment scales and results from *independent-samples t-test*.

Figure 4.7: Means of the pre- and post-treatment scales by error treatment type

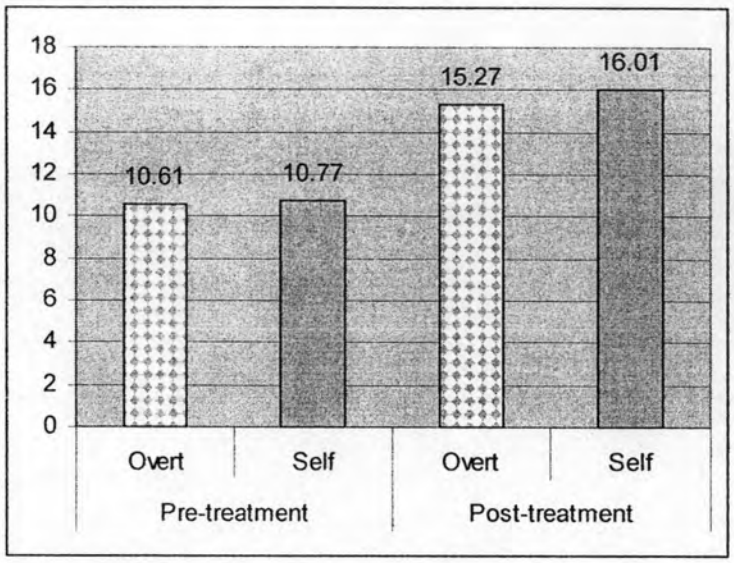


Figure 4.7 illustrates the means of the OC and the SC groups, before and after the treatment. The pre-treatment means of the OC and the SC groups were 10.61 and 10.77 respectively. The post-treatment mean of the OC group was 15.27 and that of the SC group was 16.01.

Differences between the OC and the SC groups

As an attempt to identify whether different error treatment types led to the differences in mean scores, *independent-samples t-test* was employed. First, mean scores of the OC and the SC group from the pre-treatment scales were compared in order to ensure that the two groups were equivalent. Results from Table 4.20 confirm that they were not significantly different (t = -.325, p > .05).

Table 4.20: Means, standard deviations, and results from the *independent-samples t-test* of the self-rating scales

Scales	Error Treatment	N	$\overline{X}$	S.D.	Mean Difference	df	*t*
Pre-treatment	Overt	105	10.61	3.62	-.16	208	-.325
	Self	105	10.77	3.59			
Post-treatment	Overt	103	15.27	3.84	-.74	206	-1.327
	Self	105	16.01	4.16			

Next, means from the post-treatment scales of the OC and the SC groups were compared. Raw data showed that the posttest mean of the SC group (16.01) was greater than that of the OC group (15.27). However, the difference was not large enough to yield a statistically significant result. From this it can be concluded that students receiving different error treatment types rated their knowledge similarly in the self-rating scales.

Differences between the pre- and the post-treatment

The data was further investigated to examine the differences between the answers before and after the treatment. *Dependent-samples t-test* was utilized to analyze the data. Table 4.21 shows results of the OC group and Table 4.22 presents results of the SC group. There were 7 pairs and each was asked about one of the 7 tenses. 'Post1' in the first column refers to the mean score of the post-treatment rating scales in item 1 (which asked about simple present tense).

Table 4.21: Paired differences of the overt correction group

Overt Correction		$\overline{X}$	SD	Std. Error Mean	df	*t*
Pair 1	Post 1 – Pre 1	.77	.782	.077	102	9.952*
Pair 2	Post 2 – Pre 2	.70	.790	.078	102	8.979*
Pair 3	Post 3 – Pre 3	.64	.790	.078	102	8.229*
Pair 4	Post 4 – Pre 4	.71	.749	.074	102	9.600*
Pair 5	Post 5 – Pre 5	.59	.678	.067	102	8.866*
Pair 6	Post 6 – Pre 6	.69	.701	.069	102	9.986*
Pair 7	Post 7 – Pre 7	.54	.711	.070	102	7.761*

* $p < .05$

Table 4.22: Paired differences of the self-correction group

Self-correction		$\overline{X}$	SD	Std. Error Mean	df	*T*
Pair 1	Post 1 – Pre 1	.80	.777	.076	104	10.549*
Pair 2	Post 2 – Pre 2	.87	.708	.069	104	12.543*
Pair 3	Post 3 – Pre 3	.77	.724	.071	104	10.918*
Pair 4	Post 4 – Pre 4	.67	.689	.067	104	9.919*
Pair 5	Post 5 – Pre 5	.70	.796	.078	104	9.076*
Pair 6	Post 6 – Pre 6	.73	.835	.081	104	8.998*
Pair 7	Post 7 – Pre 7	.70	.748	.073	104	9.519*

* $p < .05$

All *t*-values of both OC and SC groups were significant ($p < .05$). This means that both groups rated themselves at the post-treatment scales significantly higher than at the pre-treatment for every tense. It signifies that students in both groups perceived that they knew more about all 7 tenses than they did before using the program.

Part V summary: This part reports results from the self-rating scales. The data were analyzed by *t-test*. The findings could be summarized as follows:

1. Students in both SC and OC groups rated themselves at the post-treatment scales significantly higher than at the pre-treatment scales for every tense. It means that

they perceived they had more knowledge about the 7 tenses after using the CALL than they had before using the program.

2. Students who practiced with a different error treatment type did not have significantly different means in the post-treatment scales.

Chapter summary

This chapter reports the results and findings of the present study in response to the research questions. The summary is as follows:

Research question 1. *Do the types of error treatment given through CALL have an effect on the usage of English tenses of the students?*

Results from the posttest show that 'error treatment' factor did not have a significant effect on the usage of English tenses of the students. Students in the OC and the SC groups made similar gains. With the $F = .015$, $p > .05$ (one-tailed), **hypothesis 1** stating 'the mean score from the posttest of students receiving self-correction is significantly higher than that of the students receiving overt correction' was then rejected.

Research question 2. *Are there any differences among the usage of English tenses of undergraduate students with different language abilities: high achievers, moderate achievers, and low achievers? If so, how do students with different abilities use the tenses?*

The 'language ability' factor was found to be a significant source of the differences in the posttest gains ($F = 146.525$, $p < .05$, two-tailed) with a very large effect size. Fifty-nine percent of the overall variance could be explained by this factor. It means that students in different ability groups (LA, MA, and HA) performed significantly different in the posttest. HA students performed better than MA students and MA students performed better than LA students. Therefore, **hypothesis 2** saying 'on average, the gain scores from the posttest of the three ability groups are significantly different' was accepted.

Results from further investigation provide detailed information that the tenses that all ability groups used differently in all of the three tests were simple present, simple

past, and present continuous. However, the LA group used some tenses insignificantly different from the MA group; for example, the use of past continuous tense in the pretest, the use of the use of future tenses in the delayed test, and the use of present perfect in all tests.

Regarding the easy or difficult tenses for different ability groups, it was found that from the pretest, the notable tenses used correctly by the LA group were present continuous and simple present. The MA group could perform best in simple present, present continuous, and future tenses. HA students could prominently use the following 5 tenses in order: simple present, present continuous, past continuous, simple past, and future tenses. From the posttest and the delayed test, results were not much different from the pretest, except that the past continuous became one of the top three tenses used correctly by all ability groups. Present perfect and past perfect were found the most difficult tenses for all groups, followed by simple past.

Research question 3. *Is there an interaction effect between error treatments given through CALL and students' language abilities on their usage of English tenses?*

There was no significant interaction effect between types of error treatment and language abilities on the usage of English tenses found in the posttest ($F = .020, p > .05$). High achievers made greater gains than moderate achievers, and moderate achievers made greater gains than low achievers regardless of the error treatment type they received. Therefore, **hypothesis 3** proposing 'there is an interaction effect of error treatments and students' language abilities on the usage of English tenses at the 0.05 significance level' was rejected.

Research question 4. *Do the different types of error treatment given through CALL yield different retention rates in the delayed test?*

Results from the delayed test analyzed by two-way repeated measures ANOVA reveal that the decrease of scores of the SC group from posttest (test2) to delayed test (test 3) was small and insignificant while the decrease of scores of the OC group was greater and significant. This means that the self-correction treatment was better than the overt correction in terms of the retention. The SC students could maintain the knowledge in a higher percentage when tested by the delayed test. Consequently, **hypothesis 4** stating 'the decrease of scores from the posttest to the delayed test of the self-correction group is significantly lower than that of the overt correction group' was then accepted.

Research question 5. *What are the students' perceptions of their learning and their opinions on the CALL software programs?*

The students' perceptions of their own learning were obtained by analyzing data from the self-rating scales. It was found that students in both SC and OC groups rated themselves at the post-treatment scales significantly higher than at the pre-treatment scales for every tense. It means that they perceived they had more knowledge about the 7 tenses after using the CALL than they had been before using the program. However, students who practiced with a different error treatment type did not have significantly different means in the post-treatment scales. This means that the differences across OC and SC groups did not occur.

Results from the opinionnaire showed that, overall, students in both OC and SC groups had highly positive opinions towards the program that they used ($\overline{X}_{OC} = 3.91$, $\overline{X}_{SC} = 4.01$). When tested the means by one-sample t-test, it was found that SC students had highly positive opinions towards the SC program in almost all items, except for item 8 that asked about their attention when practicing. On the other hand, the OC group had 3 items that they rated lower than the criteria of 3.5.

Results from Chi-square test by Fisher's Exact test revealed that in general the OC and the SC groups rated the opinionnaire insignificantly different. However, the SC group rated the materials significantly higher than the OC in terms of the 'feedback' (item 10).

The next chapter, chapter V, deals with the summary of the study, discussions of the findings, implications, and recommendations for further study.