
Theses

2006

The Effectiveness of a Literature-Driven English Programme in Improving the English Language Skills of Secondary One Students in Singapore

Mary Delfin Pereira
University of Notre Dame Australia

Follow this and additional works at: <http://researchonline.nd.edu.au/theses>



COMMONWEALTH OF AUSTRALIA
Copyright Regulations 1969

WARNING

The material in this communication may be subject to copyright under the Act. Any further copying or communication of this material by you may be the subject of copyright protection under the Act.

Do not remove this notice.

Publication Details

Pereira, M. D. (2006). The Effectiveness of a Literature-Driven English Programme in Improving the English Language Skills of Secondary One Students in Singapore (Doctor of Education (Ded)). University of Notre Dame Australia.
<http://researchonline.nd.edu.au/theses/5>

This dissertation/thesis is brought to you by ResearchOnline@ND. It has been accepted for inclusion in Theses by an authorized administrator of ResearchOnline@ND. For more information, please contact researchonline@nd.edu.au.



CHAPTER 4

The Quantitative Findings and Discussion

Introduction

In this chapter the findings of the quantitative data are presented. These findings are derived from the pre-test and post-test scores of the writing and reading comprehension skills assessments. The findings from the exit survey that the students in the experimental group completed at the conclusion of the Literature-Driven English Programme (LDEP) are also presented. These findings are discussed in this chapter because of the nature of the analysis during which mixed methods were employed. Chapter 6 integrates the findings and discussions from the quantitative data in this chapter and the qualitative data in Chapter 5.

Two forms of pre-tests and post-tests were given to students in the experimental and control groups. Each group was given a set of reading comprehension and writing skills tests before the experimental group was taught the experimental curriculum. At the end of the teaching sequence of the LDEP, the control and experimental groups were given additional sets of reading comprehension and writing skills tests to complete. The results of the tests from the two groups were analysed and compared to discover if there was a difference in the improvement, if any, in the reading comprehension and writing skills of the control and experimental groups, as well as within sub-groups. The students in the experimental group were also given exit survey questionnaires to complete. It was hoped that the questionnaires would lead to a better understanding of the reading preferences of the students as well as give them an avenue to offer their feedback and opinions about the experimental curriculum.

In this chapter, the findings of the pre-test and post-test results of the writing and reading comprehension tests are presented and discussed to ascertain if

there was any measurable improvement in the performance of the students in the experimental group after being taught under the LDEP. Additionally, comparisons were made between the experimental group and control group as well as between the sub-groups. These sub-groups were created to study the impact of mediating variables such as gender, different levels of fluency in the English language, and varying school and class cultures.

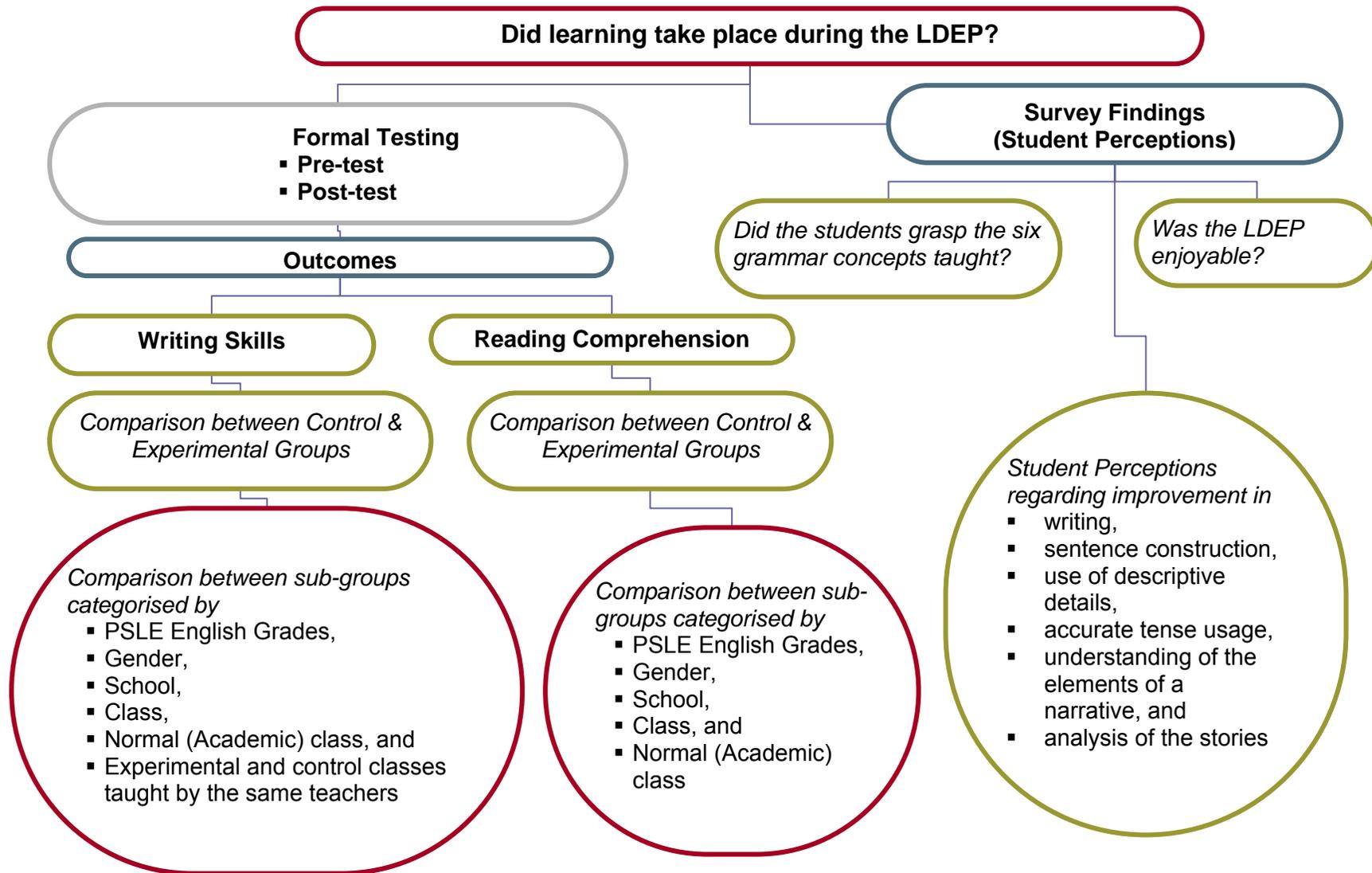
The pre-test and post-test results of the writing skills and reading comprehension tests of the students in both the experimental and control groups were analysed through the use of descriptive statistics. The intent of the analysis was to discover:

- If there was an improvement in the performance of the students in the experimental group after the LDEP,
- If there was an improvement in the performance of the control group of students who did not go through the LDEP,
- If there was an improvement, was there a difference between the improvement shown by the experimental group and the control groups?
- Lastly, if there was an improvement, did all the sub-groups, based on the mediating variables, display the same degree of improvement?

Findings of the survey that the students answered and returned at the conclusion of the LDEP were also used to ascertain if the students themselves thought that they had improved in writing and in the comprehension of the stories that they had read during the period. The questionnaire contained multiple-choice questions as well as open-ended questions. These multiple-choice and open-ended survey questions were analysed through manifest content analysis during which 'the data... [were] coded or grouped into categories' (Wilkinson & Birmingham 2003, p.69) which were then subjected to frequency counts.

A schematic outline of the content and sequence of this chapter is provided in Figure 4.1.

Figure 4.1: Schematic Outline of Chapter 4



Pre-Test and Post-Test Findings

Pre-test and post-test scores from 16 experimental classes and 8 control classes involving Secondary One students from 4 schools were analysed. Schools 1 and 4 were co-educational schools while School 2 was an all-boys' school and School 3 was an all-girls' school. A comparison of the English language ability of the students across the four schools was made on the basis of the English grades that the students received in the 2003 Primary School Leaving Examinations (PSLE), for which most of the Secondary One students had to sit before gaining admission into secondary schools. The percentages of students, in the experimental and control groups, with the different PSLE English grades in the four schools are given in Table 4.1:

Table 4.1: PSLE English Grades of Students by School and Treatment Groups

		A* (%)	A (%)	B (%)	C (%)	Missing# (%)
School 1##	Express	13.9	81	1.8	0	3.3
	Normal (Academic)	0	5.0	72.5	20	2.5
School 2	Experimental	84.5	14.5	0.9	0	0
	Control	76.6	22.5	0	0	0.9
School 3	Experimental	5.1	59.8	25.6	2.6	6.8
	Control	2.5	72.2	25.3	0	0
School 4	Experimental	1.3	56.3	42.5	0	0
	Control	1.7	59.7	37.8	0.8	0
Overall		26.06	49.8	21.17	1.46	1.5

#: A small number of students did not sit for the PSLE.

##: In School 1, there were no control groups. PSLE grades of the one Normal (Academic) class included in the analysis and those of the seven Express classes are given.

In Schools 2, 3 and 4 only the Express classes participated in the research.

In Schools 1 and 2, 94.9% and 99% of their Express students achieved Grades A* and A in the 2003 PSLE respectively (Table 4.1). In Schools 3 and 4, 69.8% and 59.5% of their Secondary One Express cohorts achieved Grades A* and A respectively (Table 4.1). In 2003 about 44% of the 49, 867 students who completed the PSLE in Singapore achieved Grades A* and A (Bukit Timah Primary School n.d.) and out of 49, 867 students, 62.2% entered the Special and Express Streams while 34.9% entered the Normal

(Academic) and Normal (Technical) streams (Ministry of Education Singapore 2003b). On comparing the PSLE English grades of the students in the four schools, it may be safe to assume that Schools 3 and 4 may be more representative of the general school population whereas Schools 1 and 2 may be more representative of the top secondary schools. Therefore, in analysing the difference between the pre-test and post-test scores among the students of the four schools, a distinction was also made on the basis of the schools, gender and the PSLE grades of the students so that possible implications of the findings to other similar schools in Singapore may be detected.

The scores of the experimental Class 9, a Normal (Academic) class, were excluded from all analyses for a number of reasons. When the exit survey questionnaires that the students completed were examined, inconsistencies between the self-reported information and the information provided by the school were discovered. Due to the inconsistencies, it was felt that the reporting could not be accurately relied upon. In addition, during class observations and through conversations with the English teacher, in general the class was observed to be unmotivated and disruptive. Moreover, no feedback from the Literature teacher of this class was received since she declined to meet or be interviewed by the researcher. For these reasons, Class 9 was dropped from all quantitative analysis as it was felt that an accurate analysis of the class performance and an assessment of the class could not be made due to the unreliable or inadequate data collected.

Before presenting the analysis of the writing skills and reading comprehension test results, it is noted that the writing skills test formed the major component of the test analysis with the reading comprehension assuming a minor role. Unequal importance was accorded to the different tests because in the experimental curriculum more time was spent on teaching narrative writing skills. Even the literary analysis was linked to the acquisition of writing skills. In trying to balance the limited time schools were able to offer for this research and the need to gain as much understanding as possible, it was decided that the focus would be more on the writing component than reading comprehension. It was hoped that during the process of reading the literature

text, analysing it and discussing the language in relation to writing skills, the students would begin to acquire better reading comprehension skills as well. Thus more attention was paid to the writing test results than the reading comprehension results since there was more concentration on the acquisition of writing skills than reading comprehension skills.

Both the written and reading comprehension test results were analysed using SPSS™ version 12 and Microsoft Excel™ 2003. The improvement scores of the students in the experimental and control groups were analysed through the calculation of Frequencies, Percentiles, Mean Differences, Effect Size, Confidence Intervals and statistical significance of the difference between the post-tests and pre-tests. The Frequencies, Percentiles, Confidence Intervals and statistical significance figures were computed through SPSS™ whereas the Effect Size was computed using Microsoft Excel™. SPSS™ does not offer Cohen's d calculation of Effect Size within the paired t-test routine and so Excel™ was used to calculate Cohen's d based on the SPSS™ output.

It must also be noted that, in the analysis, the pre-test and post-test results of individual students were taken as co-related variables. As such, the analysis was done on the basis of a correlated design, which is not only 'inherently balanced' but may also 'reduce error variance and increase statistical power' (Kline 2004, p.20).

Additionally, a series of univariate analyses was performed on the outcomes of the different sub-groups. In the analysis, the dependent variable was the mean difference score between the post-test and pre-test. There were two different dependant variables, writing skill improvement and reading comprehension skill improvement. The independent variables were the control and experimental groups; students with PSLE English grades A*, A, B and C; gender; school; class; Normal (Academic) class. In relation to the writing skills assessment, an additional independent variable included the experimental and control classes taught by the same teachers.

Since 'univariate analysis *describe* the units of analysis' and 'bivariate and multivariate analyse are aimed primarily at *explanation*' (Rubin & Babbie 1997, p.473, italicised by Rubin & Babbie), univariate analysis was selected. The aim of the research was to describe the effect of the curriculum or the size of the dependent variables with respect to each of the independent variables and the sub-groups within each independent variable. The purpose was not to concentrate on the variables or explain 'the relationships between the variables themselves' (Rubin & Babbie 1997, p.475).

Moreover, since the study was based on a naturalistic design, the group sizes were unpredictable. The number of students between groups (and within sub-groups) varied substantially. Take for instance the two independent variables: gender and PSLE English grades. As a top boys' school was included, there were 96 boys with PSLE grade A* as opposed to 27 girls with grade A*. There were 88 boys with grade A and 166 girls with grade A, 31 boys with grade B and 53 girls with grade B, and 6 boys with Grade C to 4 girls with Grade C. Multi-way analysis of multivariate interaction effects may be more meaningful if the study is 'a completely independent groups design with different subjects in each group and the number of subjects in each group... [is] equal' (Brown, K. W. et al. 1999, p.347). In this study independent groups of equal sample sizes could not be achieved and that fact reinforced the choice of univariate analyses.

To examine the effect of the curriculum on each independent variable t-tests were utilised. The use of inferential statistics such as t-tests would 'provide information regarding the magnitude of the effect, or the relationship' (Tashakkori & Teddlie 1998, p.116). Thus, the t-test was deemed appropriate to study the extent of the effect that the curriculum would have on the different groups. However 't tests thrive more when sample sizes are not very different for any given total *N*' (Rosenthal & Rosnow 1991, pp.304-305, italicised by Rosenthal & Rosnow), and uneven sample sizes in the different groups could make any interaction study between variables unreliable. The individual t-tests with the degrees of freedom (df) and standard deviation reported would make it possible for the reader to decide for himself or herself how much

importance he or she would like to place on any comparison between the test results of the different independent variables or sub-groups within these independent variables.

In the writing assessment, comparisons were made between the performances of students from experimental and control groups, and within the experimental group, between the performances of sub-groups of students: from the various schools; of different gender; and with differing PSLE English grades. The intent was to compare the performances among the sub-groups within each of the independent variables. Where these comparisons were made of the outcomes of the dependent variable among the sub-groups within an independent variable, the One-Way Anova was mainly utilized. It was necessary to compare the pre-test results of the sub-groups within the independent variable so that a more meaningful comparison may be made of the final outcomes of the sub-groups in the post-test. In this instance, the varying sample sizes between groups is of less importance since two One-way Anova tests were carried out on the same sub-groups with the same sample sizes.

Moreover, univariate analyses would serve the purpose just as well as a multivariate analysis since 'multivariate analysis often wind up as a series of univariate analyses conducted with individual outcomes' (Kline 2004, p.22). Thus, t-tests and One-Way Anovas were preferred to Two-Way Anovas.

Computing the Effect Sizes

In order to find out the magnitude of the difference between the post-test and pre-test scores, which may indicate improvement, the Effect Size of this difference or improvement was computed. The Cohen's d estimate of Effect Size (ES) was calculated with the aid of Microsoft Excel using the formula (Burns 2000, p.203):

$$\text{Cohen's } d = t / \sqrt{N}$$

It was decided that computing the ES from the t value would be appropriate since the t value includes in its computation the 'observed average difference

score and its standard error' (Kline 2004, p.45). In an experimental research of this sort, in which the aim is to find differences between two tests, it is logical to base the computation of the ES of the performance of the experimental and control groups on the mean difference scores between the post-tests and the pre-tests and standard error of these mean difference scores.

Traditionally, an ES of 0.2 is considered to be small, 0.5 is medium, and an ES of 0.8 is a large effect size (Burns 2000; Kline 2004; Smithson 2000). With respect to education, the Joint Dissemination Review Panel of the National Institute of Education in the United States requires an ES of 0.33, though it will accept an ES of 0.25 as an indicator of educational significance (Wolf 1986). Thus, there are no clear guidelines that can be applied 'across all behavioural research areas. This is because what may be considered a large effect in one area may be modest in another' (Kline 2004, p. 133). In the case of this research, it was decided to base the evaluation of the ES on the requirement of the National Institute of Education of an ES of 0.33 though an ES of 0.25 may be sufficient to indicate effectiveness since the experimental curriculum ran for only five to six weeks in each school, a short period in terms of language instruction. It is reasonable to argue that 'longer, more intense intervention may potentially have a larger effect than a shorter, less intense intervention' (Kline 2004, p.42). Therefore, an ES of 0.5 and above is taken to indicate that the curriculum was highly effective in imparting the writing and reading comprehension skills.

Analysis of the Writing Test Scores of the Pre-test and Post-test

All classes in the study, whether in the control or experimental groups, were assessed using a writing skills test. As part of the writing skills test, the students were instructed to write narratives for the pre-test and post-test, based on different story-starters. The pre-test was administered before the curriculum was introduced, no more than one week prior to the commencement of the curriculum. The post-test was administered within a week of the conclusion of the curriculum instruction. Copies of the pre and post writing skills tests are included in Appendix 3b.

The pre-test and post-test writing skills instruments were marked using a predetermined marking scheme (please refer to Appendix 5a). The researcher marked the tests after all the classes in the four schools had completed the post-tests.

The marking was conducted in a manner that would minimise expectancy effects. The pre-test scripts were all marked before the post-test scripts were marked. In total, there were close to a thousand pre-test scripts and a thousand post-test scripts, so it would have been difficult to track the marks any particular group or student had scored in the pre-test when the post-test scripts were being marked. In addition, marks were only recorded on Excel™ spreadsheets after the pre-test and post-test scripts were all marked.

Another step was taken to ensure that the marking was reliable. A request was made that each of the teachers in the participating schools mark the pre-test and post-test scripts of five students. The score differences between the post-test and the pre-test were then compared to the score differences awarded by the researcher for the same scripts. A paired t-test comparison of the post-pre scores awarded by the teachers and the researcher yielded a negligible mean difference score of -0.125 with a low statistical significance of 0.719 (2-tailed). The regression model on the same scores resulted in a regression beta of 0.199, and R^2 was only 1.7%. There appears to be negligible disparity in the post-pre scores awarded by the teachers and the researcher for the same scripts. Thus, it may be safe to assume that the marking of the researcher was reliable.

One discrepancy needs to be recorded. When the pre-test scripts of the school with the largest number of classes (nine) were marked, initially it was decided not to award half marks. However, since the maximum mark awarded for each category was only five, later it was decided that in order to have more levels of differentiation, it would be necessary to include half marks. However, the school had requested to view the scripts before the scripts could be re-marked or even before the marks could be recorded (since the post-test scripts were not yet marked). When the scripts were returned,

only scripts from six classes remained intact. Scripts from two classes, Classes 3 and 8, were missing, and a small number of scripts from another class was also missing. Therefore, it was decided to exclude the two classes and results of the students in the third class with missing scripts from the analysis for the following reasons:

- It was felt that a fair analysis cannot be made without marking the scripts again to take into account the inclusion of half marks in the assessment.
- Even though the teachers provided the pre-test marks, the comparison of results between the pre-test and post-test would be distorted since the post-test scripts were marked to reflect the inclusion of half marks in the assessment of the writing under the different categories.
- There could be decreased reliability of a comparison between the outcomes of these classes and those of the other classes in the same school and other schools since the marking of the scripts of the other classes would have included more levels of differentiation in the marking.

It must be noted at this point that the returned scripts were in their original condition; there were no alterations made to the scripts by the students or teachers before they were returned.

In analysing the data, not only was the difference in outcomes between the control and experimental groups assessed but the outcomes of sub-groups within the experimental group were analysed as well in order to ascertain whether mediating variables had an effect on the outcomes. These sub-groups were categorised based on: the PSLE grades of the students, gender, school, class, and Normal (Academic) stream. In addition, the results of a sub-group of experimental and control classes taught by the same teachers were also analysed.

Comparison of Improvement Scores in Writing between the Experimental and Control Groups

The control group of students in the three schools were taught narrative writing skills through the regular curriculum. The regular, that is control, curriculum also included the teaching of text-based grammar. Extracts were used to teach grammar explicitly and the grammar concepts taught were linked to writing tasks. In two of the schools the teaching of these narrative writing techniques and grammar concepts was not linked to the literature text. In the third school, writing tasks that were linked to the literature text were more in the nature of literary analysis. Narrative elements and narrative writing techniques were often taught separately without an explicit link being made to the literature text. Grammar was seldom explicitly taught. Similar narrative writing techniques and narrative elements were also taught during the LDEP. However, the teaching of these narrative writing skills and the explicit teaching of grammar were linked to the literature text. The grammar concepts taught were also pre-selected to aid in the writing of narratives. Therefore, while the control group was taught similar grammar concepts, the experimental, that is the LDEP, group learnt how the use of these concepts could have an impact on the narrative.

In the initial analysis a comparison was made between the experimental and control groups based on the difference between the post-test and the pre-test results in the writing assessment of the individual students. The valid scores of 482 students from the experimental group and 290 students from the control group were analysed. The invalid scores comprised of scores from students who had only sat for one of the tests, either the pre-test or the post-test, and those students who had returned scripts that were less than half a page long or who had very similar scripts. The former meant that no valid comparison could be made since comparison was made between the post-test and the pre-test of individual students. Likewise, similar or incomplete scripts that were very brief were deemed inadequate for an accurate assessment to be made of the narrative writing abilities of the students.

Figure 4.2 shows the percentage of students in the experimental and control groups with the different improvement scores. The improvement scores are the differences between the post-test scores and the pre-test scores.

Figure 4.2: Improvement Scores for Writing Skills Assessment for Experimental and Control Groups

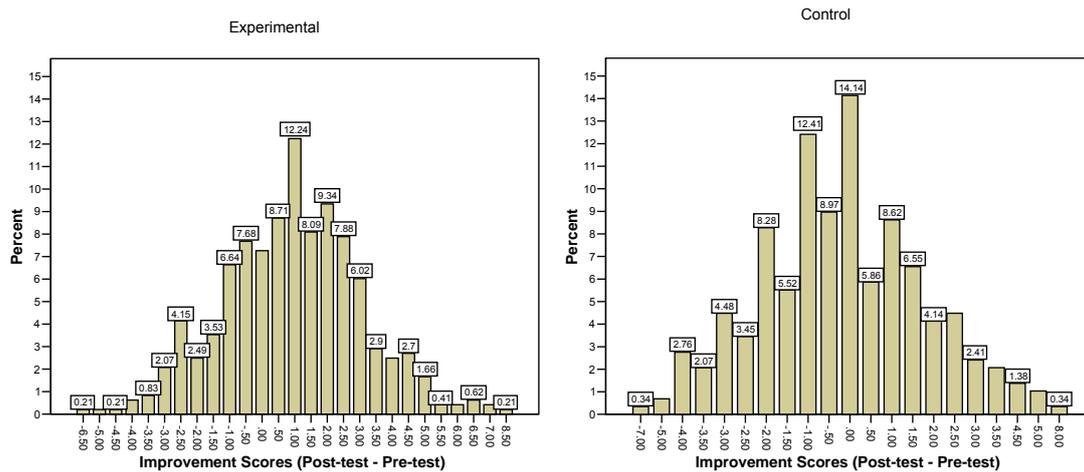


Figure 4.2 above suggests that more students in the experimental group than in the control group improved. More students in the experimental group attained positive improvement scores. From a total of 482 students in the experimental group, approximately 64% received positive improvement marks. In contrast, about 37% of the 290 students in the control group achieved positive improvement marks. The frequency charts suggest that most of the students in the control group did not improve while the opposite is true of the experimental group as the majority of them had improved.

The mean scores of the pre-test and post-test of the students in the experimental and control groups were examined since more students in the experimental group demonstrated greater improvement in the post-test than the students in the control group. The mean scores and the 95% confidence limits for the results of the pre-test and the post-test of the experimental and control groups are displayed in Figure 4.3.

Figure 4.3: Mean Scores of the Pre-Test and Post-Test Writing Assessment for the Experimental and Control Groups

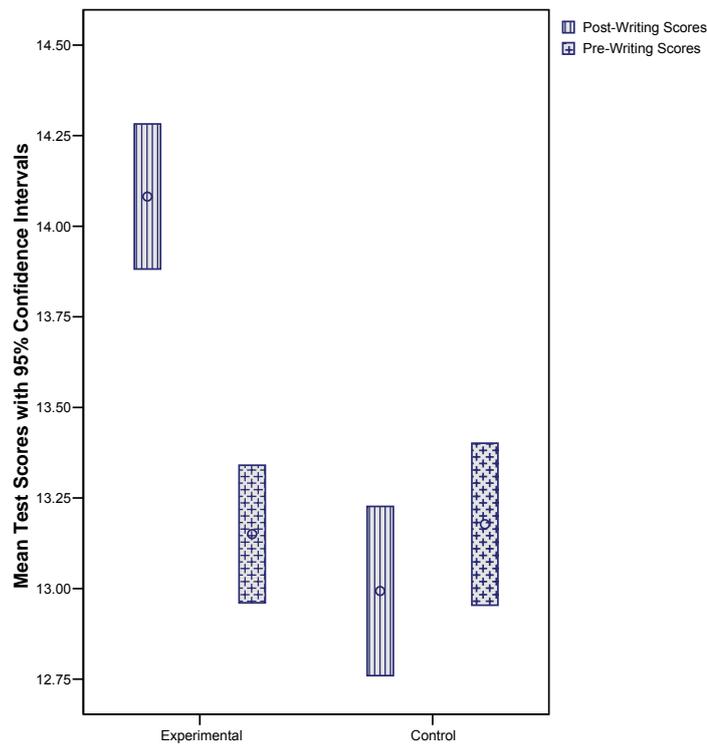


Figure 4.3 suggests that the pre-test scores of the control and experimental groups are comparable as the mean scores of the two groups are similar. However, the post-test scores present a very different story. The mean of the post-test scores of the control group is actually lower than the mean of the pre-test scores, though the difference is slight (Figure 4.3).

On the other hand, the mean of the post-test results of the experimental group is much higher than the mean of the pre-test scores or the mean of the post-test scores of the control group. Moreover, the 95% Confidence Intervals of the pre-test and post-test scores of the control group overlap (Figure 4.3). However, the experimental group has distinctly different pre-test and post-test scores such that their 95% Confidence Intervals do not overlap, indicating that in 95% of the cases, the mean would be much higher (Figure 4.3). It appears that many of the students in the experimental group had improved.

The One-Way Anova test results of the pre-test scores of the experimental and control groups revealed no significant difference between the results of the two groups. Next the post-test results of the experimental and control

groups were examined to discover whether the post-test results of these groups were significantly different from each other. Table 4.2 presents the outcome of the One-Way Anova test of the post-test results of the experimental and control groups.

Table 4.2: The One-Way Anova test results of the Post-test Scores of the Control and Experimental Groups

Post Writing Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	214.665	1	214.665	46.055	.000
Within Groups	3588.999	770	4.661		
Total	3803.664	771			

The post-test results of the One-Way Anova reveal that there is a highly significant difference between the two groups. There is a probability of less than 0.001% that the difference between the post-test scores of the experimental and control groups had occurred by chance alone (Table 4.2).

On the other hand, there was no statistically significant difference between the pre-test results of the experimental and control groups. Therefore, it can be assumed that the students in the experimental and control groups performed similarly in the pre-test. In other words, the students appeared to be of similar abilities. However, in the post-test, there seems to be a significant difference (Table 4.2). If the students were of similar abilities, why was there a significant difference in the post-test results of the two groups? The variables, including chance, which may explain the difference in the post-test, were also present when the students completed the pre-test. The low probability in the post-test, in contrast to the higher probability in the pre-test, which can be attributed to chance alone, may indicate an appreciable improvement in the performance of the experimental group due to the independent variable, the LDEP.

In order to get a clearer picture of the improvement, or lack of, achieved by either group, a paired t-test of the differences between the test scores within each group was conducted to discover the degree of the significance of these differences.

In short,

- Was there a significant difference between the pre-test and post-test results within the control and experimental groups, and
- Was there an improvement in the post-test results of either, or both, groups?

The paired t-test is the appropriate test because for each student in the experimental and control groups, there is a single pre-test score and a single post-test score.

From the paired t-tests on the pre-test and post-test writing results of the control and the experimental groups, the following results were obtained (Table 4.3). The ES was also computed from the t value.

Table 4.3: Paired t-test results of the Mean Difference Scores between the Post-test and the Pre-test Writing Assessment by Treatment Groups

	Paired Differences					t	df	Sig. (2-tailed)	ES
	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower	Upper				
Experimental	0.93154	2.13901	0.09743	0.74010	1.12297	9.561	481	.000	0.44
Control	-0.18448	2.03973	0.11978	-0.42023	0.05126	-1.540	289	.125	N.A.

**Table adapted to include paired t-test results from the SPSSTM output and the ES.*

Table 4.3 indicates that the mean difference score between the post-test and pre-test scores of the experimental group is positive, showing that on the average, the experimental group had improved. The mean improvement score is significantly different from zero at $p < 0.001$. However, the mean difference score of the control group is negative, though at -0.18 it is very slight (Table 4.3). The mean difference score of the control group is also not significantly different from zero, thereby indicating that the scores of the post-test and pre-test were similar. Thus, on the average, the students in the experimental group improved by 0.93 marks (computed to two decimal places) whereas the students in the control group may be considered to have scored similarly on both the pre-test and the post-test (Table 4.3).

Overall Effect Size of the Mean Improvement Score of the Experimental Group

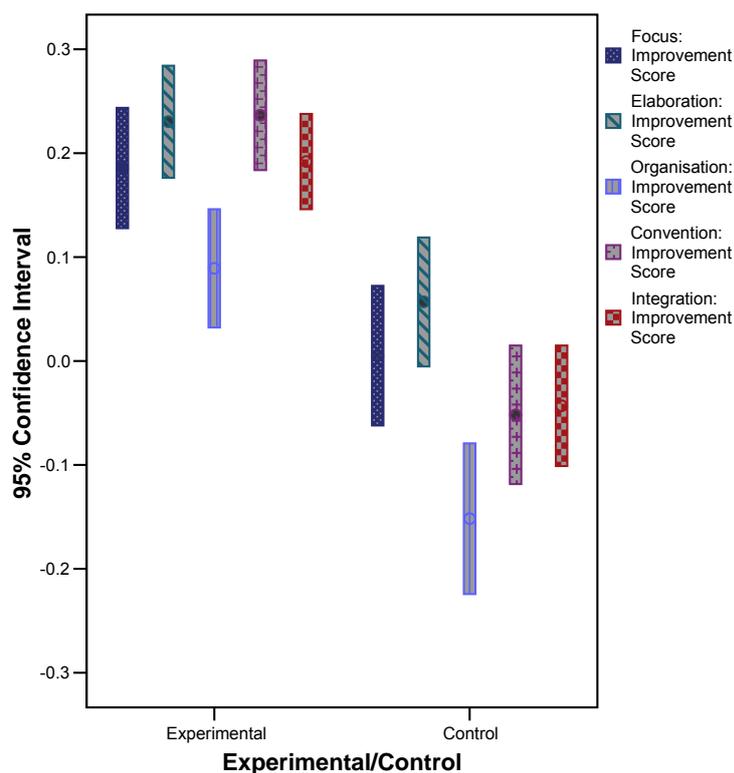
The next step taken was to calculate the Effect Size (ES) of the mean difference between the post-test and pre-test results, or mean improvement score, of the experimental group only. As there is no statistically significant difference between the pre-test and the post-test which the students in the control group completed and there is, in fact, a low negative “t” value, it can be said that there was no change in performance, or “effect”. Thus, the ES for the control group was not calculated.

From Table 4.3, it can be seen that the overall ES of the mean difference score between the post-test and the pre-test of the experimental group was calculated to be 0.44 (to 2 decimal places). This is above the 0.33 standard set by the American Joint Dissemination Review Panel of the National Institute of Education (Wolf 1986). Generally, the curriculum appears to have been effective in improving the writing skills of many of the students in the experimental group.

There were five sub-categories on which the students were assessed in the writing tests. The five categories were *Focus*, *Elaboration*, *Organisation*, *Convention* and *Integration*. In each sub-category, the maximum marks allotted were five marks. The improvement scores of the students in the experimental and control groups in the different categories were then analysed.

Figure 4.4 displays the improvement scores as well as the 95% Confidence Intervals in each of the five categories for the experimental and control groups.

Figure 4.4: Mean Improvement Scores in the Five Sub-Categories in the Writing Test for the Experimental and Control Groups



The graph in Figure 4.4 indicates that the experimental group had improved in all the categories, attaining positive mean improvement scores in the five categories. On the other hand, the students in the control group achieved positive improvement score in only one category, namely “elaboration”. In the other categories, the mean improvement scores were either negative or zero (Figure 4.4.). Since the lower bounds of the Confidence Intervals of the improvement scores for the experimental group are all above zero, there is a 95% chance that the mean improvement score would not be zero, but be positive. Moreover, in each of the sub-categories none of the Confidence Intervals of the experimental group overlaps with the corresponding Confidence Intervals of the control group. In fact, the lower bounds of the Confidence Intervals of the mean improvement scores of the experimental group are higher than the upper bounds of the Confidence Intervals of the mean improvement scores of the control group (Figure 4.4).

Most of the students in the experimental group appeared to have improved more than the students in the control group in all the five measured areas.

To confirm that indeed the experimental group had improved more than the control group in the five measures areas, independent sample t-tests on the improvement scores of the experimental and control groups in each of the five sub-categories were performed. Table 4.4 on p.117 includes the independent t-test results from the SPSS™ output with the addition of the ES.

The independent samples test results in the adapted Table 4.4 indicate that there is a statistically significant difference between the improvement scores of the experimental group and the control group. The positive t values in all the five sub-categories indicate a positive difference between the improvement scores of the experimental group and the control group, thereby signifying that there was a greater improvement in the students from the experimental group than the control group in all the five sub-categories (c.f. Table 4.4, p.117).

The Levene's test for equality of variances indicates whether the experimental and control groups have approximately equal variances on each of the dependent variables given in the first column in Table 4.4 (c.f. p.117). The Levene's test indicates that the variances are not considered to be equal in four out the five categories, namely "focus", "elaboration", "convention" and "integration". In one of the categories, "organisation", since $p > 0.05$, it indicates that the variances are approximately equal. As such, under "organisation" the t-value where the variance is assumed to be equal is taken for the computation of the ES of the difference in improvement score between the experimental and control groups whereas in the other four categories, the t-values where the variances are not assumed to be equal are taken for the computation of the Effect Sizes of the difference in improvement scores. The appropriate t value is printed in bold in Table 4.4 (c.f. p.117).

The positive Effect Sizes indicate there is a degree of difference in the improvement scores of the experimental and control groups with the experimental group performing better than the control group. The greatest degree of difference can be found under the categories "convention" and "integration" (c.f. Table 4.4, p.117). The students in the experimental group,

who were explicitly taught selected grammatical concepts pertaining to narrative writing through the literature text, appeared to have improved most (ES = 0.27) in that area, which was assessed under the category “convention”. The control group was taught grammar concepts explicitly as well but these were not linked to the literature text. In addition, in the category “integration”, which assesses the narratives of the students as a whole, the degree of improvement registered by the experimental group over the control group was also greater than the improvement demonstrated in the three other categories of “focus”, “organisation” and “elaboration” (c.f. Table 4.4, p.117).

Not only were there positive degrees of difference in the improvement scores of the experimental and control groups in all the five sub-categories but the experimental group had also achieved positive mean improvement scores in all the categories. On the other hand, the control group achieved close to zero or very slight positive mean improvement scores in two sub-categories (“focus” and “elaboration”) and negative mean improvement scores in three other sub-categories (c.f. Figure 4.4, p.114). The experimental curriculum had included the explicit teaching of narrative writing skills and grammar with the aid of the literature text. Therefore, the explicit teaching of the relevant skills using a literature text as a tool appears to have been effective in improving the skills of the students in the experimental group in all the five measured areas.

The ES of each of these five measured categories is relatively modest, ranging from 0.16 to 0.27. The maximum score awarded for each of these categories was five points and the variance in each category is comparatively small. Hence the mean differences are restricted by the small range of the marking scale. The intent of the research was to investigate the effect of the LDEP and the writing test amply achieves that aim. Later research may readily address components or categories of the writing skills to develop a finer grained assessment of the improvement that the LDEP offers writing skills in English.

Table 4.4: Independent samples t-test results of the Mean Improvement Scores in the 5 Sub-Categories in the Writing Assessment

		Levene's Test for Equality of Variances		t-test for Equality of Means					ES
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	
Focus: Improvement Score	Equal variances assumed Equal variances not assumed	15.455	0.000	3.899 4.003	770 659.586	0.000 0.000	0.18051 0.18051	0.04630 0.04510	0.16
Elaboration: Improvement Score	Equal variances assumed Equal variances not assumed	17.093	0.000	4.033 4.149	770 663.627	0.000 0.000	0.17339 0.17339	0.04300 0.04179	0.16
Organisation: Improvement Score	Equal variances assumed Equal variances not assumed	0.140	0.708	5.122 5.139	770 615.462	0.000 0.000	0.24094 0.24094	0.04704 0.04688	0.18
Convention: Improvement Score	Equal variances assumed Equal variances not assumed	6.009	0.014	6.627 6.664	770 619.767	0.000 0.000	0.28824 0.28824	0.04350 0.04325	0.27
Integration: Improvement Scores	Equal variances assumed Equal variances not assumed	4.909	0.027	6.217 6.252	770 619.577	0.000 0.000	0.23501 0.23501	0.03780 0.03759	0.25

**Table of SPSSTM independent t-test results output adapted to include the ES.*

The findings indicate that the experimental group had improved more than the control group in the writing assessment; however, that information alone is insufficient. It is more important to discover the extent to which they had improved and whether that improvement had occurred irrespective of the mediating variables. If there is an appreciable improvement among the students of different abilities in the experimental group across the four schools, then the findings would have practical significance.

The students in the experimental group were sub-divided based on mediating variables. The mediating variables considered in the analysis included:

- PSLE English grades used to ascertain the prior English Language ability of the students,
- gender,
- schools,
- classes,
- Normal (Academic) and Express classes, and
- experimental and control classes that were taught by the same teachers.

The sub-groups were analysed so as to discover if there were differences among the groups with respect to the improvement shown in the post-test writing assessment.

PSLE English Grades and the Improvement Scores of the Students

The PSLE is a state-wide examination conducted for four subjects, namely English, Mother Tongue Language, Mathematics and English. The examination is externally set and marked. The research design included the collection of the previous PSLE performance results of the students. These English PSLE grades are recorded from A* to C with C being the lowest. These grades are accepted as indicators of ability levels at the commencement of Secondary One and the Singapore system uses PSLE grades in order to determine the academic

streams of students in Secondary One. It is reasonable to investigate whether the curriculum was equally successful for students with different levels of ability as indicated by the PSLE English grades of the students. In addition, if the students had improved, it would be pertinent to discover the extent of this improvement.

In analysing the writing results of the students, the results of students who had no PSLE English grades, because they did not sit for the PSLE, were not analysed. Since the intent was to discover the effect that the different PSLE grades might have on the writing test scores of the students, the inclusion of students with missing grades would serve no purpose.

Before analysing the improvement scores of the students with different PSLE English grades, a One-Way Anova test of the pre-test results was performed to discover if there were any statistically significant differences between the four groups (A*, A, B and C). The results are displayed in Table 4.5.

Table 4.5: One-Way Anova test of the Writing Assessment Pre-test Scores of the Students with Different PSLE English Grades

ANOVA

Pre-Writing Scores

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	514.069	3	171.356	50.048	.000
Within Groups	1598.928	467	3.424		
Total	2112.998	470			

Table 4.5 indicates that statistically there is a significant difference between the groups ($p < 0.001$). The different groups had performed differently in the pre-test. Therefore, a One-Way Anova was performed on the improvement scores to discover if this difference was maintained in the post-test. The test results are displayed in Table 4.6:

Table 4.6: One-Way Anova test results of the Writing Assessment Improvement Scores of the Students with Different PSLE English Grades

ANOVA

Improvement Scores (Post-test - Pre-test)

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.395	3	3.798	.827	.479
Within Groups	2143.755	467	4.590		
Total	2155.151	470			

Table 4.6 demonstrates that there are no statistically significant differences between the groups with the different PSLE English grades since $p > 0.05$. In other words, the groups appeared to have improved at the same rate. This result is taken to indicate that the programme overcame some of the differences in achievement since the students with different PSLE English grades improved similarly.

In the next stage, the 95% Confidence Intervals of the means of the difference scores between the post-test and the pre-test writing assessment of students with the different PSLE English grades were computed.

Figure 4.5: 95% Confidence Intervals of the Means of the Improvement Scores by PSLE English Grades

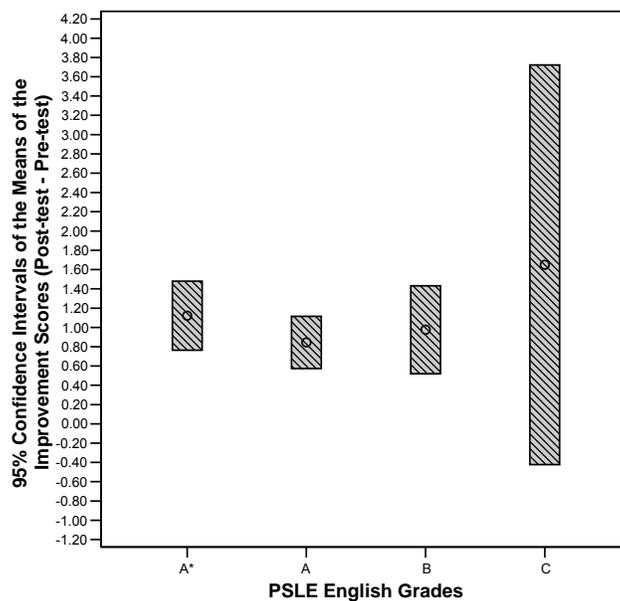


Figure 4.5 illustrates that the means of the improvement scores of students with Grades A*, A, and B are quite similar. In addition, the 95% Confidence Intervals are also very similar. In the case of these three groups, there is a 95% likelihood that the means in repeated tests would be above 0.5. In the case of students with Grade C, the mean of the improvement score is higher than those with Grades A*, A, and B (Figure 4.5). However, the Confidence Interval of the mean of the improvement score of the students with Grade C indicates a wide variation in the marks as the standard error of the mean is rather large. Within the 95% Confidence Interval, there is also the possibility of the mean improvement score dropping to below zero (Figure 4.5). In part, the large standard error of the mean may be due to the relatively low number of students with a PSLE English grade of C (Table 4.7). A more detailed analysis by gender later in this chapter will also shed some light on the large difference between the performances of the students with Grade C and those with the other grades.

Table 4.7: Mean Improvement Scores (Post-test – Pre-test) in the Writing Assessment by PSLE English Grades

PSLE English Grades	Mean Improvement Scores	No. of Students	Std. Deviation
A*	1.1220	123	2.00649
A	0.8445	254	2.18747
B	0.9762	84	2.10121
C	1.6500	10	2.89684
Total	0.9575	471	2.14136

Table 4.7 shows the mean improvement scores of the students with the different PSLE English grades. The greatest mean improvement score of 1.65 was achieved by students with Grade C. The students with Grades A*, A and B attained similar mean improvement scores of 1.12, 0.84 and 0.98 (Table 4.7). Students with Grades A*, A, and B appear to have improved at a similar rate whereas the mean improvement of the students with Grade C is higher though with a wider variation in the individual marks.

The ES of the improvement gained by each group of students with different PSLE English grades was next calculated in order to ascertain the extent of the improvement that the curriculum effected for all students, regardless of the PSLE English grade. Paired t-tests of the mean difference scores between the post-test and pre-test for each group of students, with respect to their PSLE English grades, were performed. The results and the ES, calculated from the t value, are displayed in Table 4.8.

Table 4.8: Paired t-test Results and the ES of the Mean Difference Scores between the Post-test and Pre-test Writing Assessment by PSLE English Grades

PSLE Grade	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
				Lower	Upper				
A* (n=123)	1.12195	2.00649	0.18092	0.76380	1.48010	6.201	122	.000	0.56
A (n=254)	0.84449	2.18747	0.13725	0.57418	1.11479	6.153	253	.000	0.39
B (n=84)	0.97619	2.10121	0.22926	0.52020	1.43218	4.258	83	.000	0.46
C (n=10)	1.65000	2.89684	0.91606	-0.42227	3.72227	1.801	9	.105	0.57

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

Table 4.8 suggests that there is a statistically highly significant difference between the post-test and the pre-test results of students with Grades A*, A and B, since $p < 0.001$. Though the results of the students with Grade C are not as significant, since probability is greater than 0.05, the ES of the improvement achieved by these students is the largest (Table 4.8). However, a cautionary note should be included here. The different sample sizes could influence the significance level and the ES. The small sample size of students with Grade C could account for the low statistical significance and the large ES. Power is usually lower in small sample sizes (Kline 2004) unless the variance changes and the small sample size can also lead to low significance level (Muijs 2004). There were double the number of students who had Grade A than Grade A*, and three times as many students who had Grade A than Grade B. On the other hand, there were only ten students with Grade C who participated in the research, about 25 times fewer than the students with Grade A (Table 4.8).

Nevertheless, the results of all the students across the grade levels show a moderate or greater than moderate ES, thereby indicating that the curriculum was effective for many of the students, regardless of their PSLE English grades.

The next mediating variable that was taken into account was gender. Did the boys and girls register similar improvements in the post-test? If they had, to what extent did they improve? Additionally, was the degree of improvement the same for students of either gender, regardless of their PSLE English grades?

Mean Scores and the Effect Sizes of Improvements in Relation to Gender and PSLE English Grades

The mean scores and the 95% Confidence Intervals of the pre-test and post-test writing assessment by gender is displayed in Figure 4.6.

Figure 4.6: Mean Scores of the Pre-test and Post-test Writing Assessment and 95% Confidence Intervals by Gender

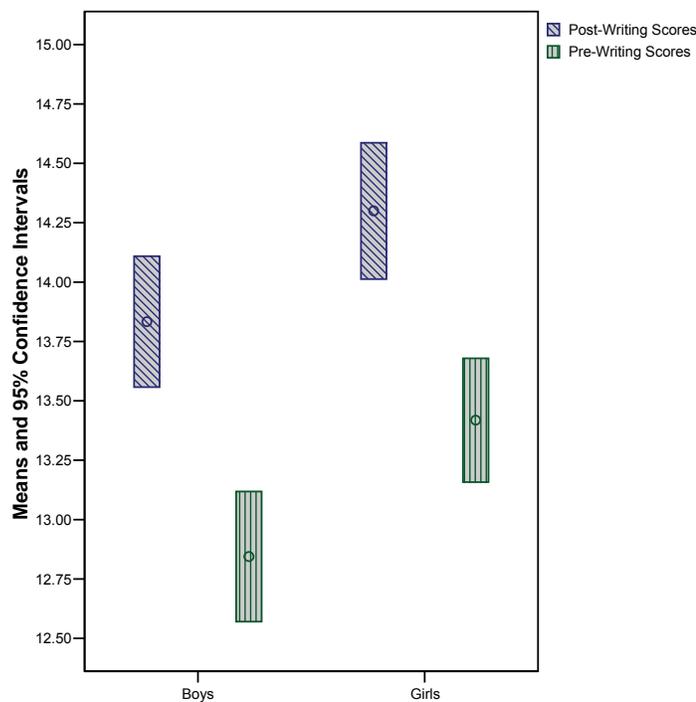


Figure 4.6 reveals that both boys and girls improved on their mean scores in the post-test from the pre-test. However, though boys did not perform as well as the

girls in the post-test, the boys appear to have improved slightly more than the girls. A comparison of the 95% Confidence Intervals between the pre-test and post-test results of the boys and girls indicate that the disparity between the performance of the girls and the boys may have been reduced in the post-test. In the post-test, there is a narrowing of the gap between the lower bound of the Confidence Interval of the results of the girls and the upper bound of the results of the boys. The gap in the pre-test appears to be larger (Figure 4.6).

Table 4.9 below displays the mean differences between the post-test and pre-test results of the writing assessment for boys and girls.

Table 4.9: Mean Scores of the Pre-test and Post-test Scores in the Writing Assessment by Gender

	Boys/Girls	N	Mean	Std. Deviation	Std. Error Mean
Post-Writing Scores	Boys	225	13.8333	2.09591	.13973
	Girls	257	14.2996	2.33681	.14577
Pre-Writing Scores	Boys	225	12.8444	2.08583	.13906
	Girls	257	13.4183	2.11974	.13223

The mean difference score between the post-test and the pre-test in the writing assessment is about one mark for the boys and 0.9 for the girls (Table 4.9). In real terms, the boys improved more than the girls by a margin of about 0.1 of a mark. To discover the significance of this margin of difference, the Effect Sizes of the improvement gained by the boys and girls were computed. The results of the paired t-test between the post-test scores and the pre-test scores for the boys and girls as well as the ES computed from the t value are displayed in the adapted Table 4.10 below.

Table 4.10: Paired t-test results of the Post-test and Pre-test Scores in the Writing Assessment by Gender

	Paired Differences					t	df	Sig. (2-tailed)	ES
	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference					
				Lower	Upper				
Boys (n=225)	0.98889	1.99298	0.13287	0.72706	1.25071	7.443	224	.000	0.50
Girls (n=257)	0.88132	2.26180	0.14109	0.60348	1.15916	6.247	256	.000	0.39

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

From the paired t-test results in Table 4.10, it can be ascertained that the statistical difference between the post-test and pre-test results is highly significant ($p < 0.001$) for the boys as well as the girls. From the computation of the improvement score ES, it can be ascertained that the ES for the boys is larger at 0.5 whereas for the girls, the ES stands at 0.39 (Table 4.10). However, both these Effect Sizes are above the ES of 0.33 set by the American Dissemination Review Panel of the National Institute of Education as the standard of educational significance (Wolf 1986). The curriculum appears to have been effective in improving the narrative writing skill of the boys and girls. The curriculum also appears to have benefited the boys more with the curriculum being highly effective in improving their writing skills (c.f. p.105 for the discussion on the ES and the benchmarks for indicating the effectiveness of the curriculum).

The next question asked was whether the degree of improvement was the same for students of either gender, regardless of their PSLE grades. To discover the answer to that question, the Effect Sizes of the various groups, based on gender and PSLE English grade, were calculated. However, before computing the effect sizes, the paired t-test results of the post-test and pre-test writing scores based on gender and PSLE English grades were examined. Table 4.11 shows the results of the paired t-tests and the ES for the different groups.

Table 4.11: Boys' Paired t-test results of the Post-test and Pre-test Scores in the Writing Assessment by PSLE English Grades

PSLE Grade	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
				Lower	Upper				
A* (n=96)	1.07292	1.98412	0.20250	0.67090	1.47494	5.298	95	.000	0.54
A (n=88)	0.85227	2.00310	0.21353	0.42786	1.27669	3.991	87	.000	0.43
B (n=31)	0.82258	1.80069	0.32341	0.16208	1.48308	2.543	30	.016	0.46
C (n=6)	3.08333	2.20038	0.89830	0.77418	5.39249	3.432	5	.019	1.40

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

Table 4.11 demonstrates that there is a statistically significant difference between the post-test and pre-test results that the boys of all grade levels had taken, with the boys getting a mean improvement score ranging from 0.82 (to 2 decimal

places) to 3.08 (to 2 decimal places) (Table 4.11). The improvement score Effect Sizes range from 0.43 for boys with Grade A to a large 1.40 for boys with Grade C, though it must be noted that the analysis included only six boys with Grade C (Table 4.11). From the ES, it can be ascertained that there is significant evidence to show that the curriculum was highly effective in imparting writing skills to most of the boys, regardless of their PSLE grades. The writing test scores of the girls were then scrutinised.

Table 4.12: Girls' Paired t-test results of the Post-test and Pre-test Scores in the Writing Assessment by PSLE English Grades

PSLE Grade	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
				Lower	Upper				
A* (n=27)	1.29630	2.11342	0.40673	0.46026	2.13234	3.187	26	.004	0.61
A (n=166)	0.84036	2.28502	0.17735	0.49019	1.19053	4.738	165	.000	0.37
B (n=53)	1.06604	2.27029	0.31185	0.44027	1.69181	3.418	52	.001	0.47
C (n=4)	-0.50000	2.61406	1.30703	-4.65956	3.65956	-0.383	3	.728	-0.19

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

Table 4.12 shows that there is a statistically significant difference between the post-test and pre-test results in the writing assessment that the girls with grades A* to B had taken, with the mean improvement score ranging from 0.84 (to 2 decimal places) to 1.30 (to 2 decimal places). The ES of between 0.37 and 0.61 of the improvement scores of these girls (Table 4.12) are educationally significant (c.f. p.105), thereby indicating that the curriculum was highly effective in improving the writing skills of the girls with Grades A* to B. However, there is no significant difference with regard to the performance of the girls with grade C, and the mean difference between the post-test and the pre-test is -0.50. The effective difference is slight, with the ES showing a negative 0.19 value (Table 4.12), thus demonstrating that the girls had performed similarly in the pre-test and post-test. The curriculum may not have delivered measurable benefits to the girls with Grade C. The findings with regard to the girls with Grade C and their apparent lack of improvement in writing will be further discussed in the next section when comparison is made to the performance of the boys in the writing test.

The graph below better illustrates the differences, if any, between the genders in relation to improvement achieved in the post-test. The Effect Sizes of the improvement demonstrated by the boys, girls and for all the students with regard to PSLE English grades are graphically represented in Figure 4.7.

Figure 4.7: Effect Sizes of the Improvement in the Writing Skills Assessment by Gender and PSLE English Grades

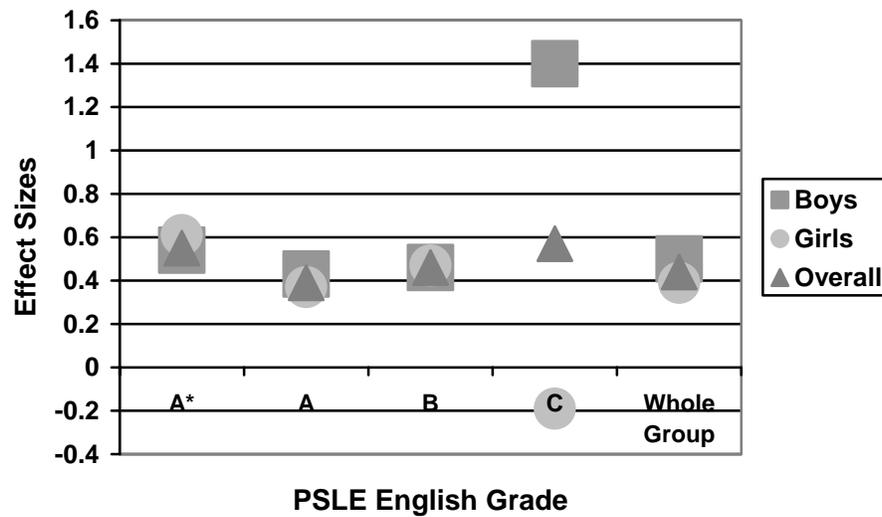


Figure 4.7 indicates that the overall improvement score ES and the improvement score Effect Sizes of boys and girls are similar. The overall ES is 0.44 (c.f. p.112) and the ES for the boys and girls are 0.5 and 0.39 respectively (c.f. p.124). Thus, overall, the curriculum appears to have been effective for all students though the boys appear to have benefited more than the girls.

The Effect Size of the improvement displayed by the boys with Grade A is slightly larger than that of the girls with the same grade, and the girls with Grade A* attained a slightly larger ES than the boys with the same grade (Figure 4.7). The Effect Sizes of the improvements attained by the boys and girls with Grade B are almost alike. Generally, the curriculum appears to have benefited both boys and girls with Grades A*, A, and B.

The negative mean difference score between the post-test and the pre-test for the girls and the large positive mean difference for the boys in the writing

assessment would have skewed the Confidence Interval of the improvement scores of the students with Grade C (c.f. Figure 4.5 on p.120). Figure 4.7 indicates that the improvement score ES for the boys is large whereas the ES for the girls is a negative value. Why the curriculum appears to have been effective for the boys with Grade C but not for girls with the same grade is unclear. Further research needs to be conducted to discover the reasons for it. However, it must also be noted that the sample sizes for students with Grade C were very small. Writing results of only six boys and four girls with Grade C were analysed. In addition, the standard error of the means are also large; for the boys, it is 0.89830 (c.f. Table 4.11, p.125) and for the girls, it is 1.30703 (c.f. Table 4.12, p.126). This would indicate that there are wide variations in the mean difference scores between the post-test and the pre-test results. As such, the findings would not be as dependable as the findings of the students with Grades A*, A, and B.

From the analysis of the data, it may be implied that the experimental curriculum seemed to have had a positive effect on most students, regardless of gender or level of fluency in the language (as can be ascertained from the different PSLE English grades). The boys experienced moderate to large ES improvements in their writing skill scores and only one sub-group of girls with a lower PSLE grade did not display any improvement in their writing skills scores.

Comparison of the Improvements in the Writing Assessment Gained by the Students in the Four Schools

A One-Way Anova was performed on the pre-test and improvement scores of the schools to discover if there were differences in the scores among the four participating schools. The intent was to discover if the students from each school improved at the same rate or if there were differences in the improvements achieved by the students of different schools. The One-Way Anova test results are displayed in Table 4.13.

Table 4.13: The One-Way Anova of the Pre-Test and Improvement Scores in the Writing Assessment in the 4 Schools

		Sum of Squares	df	Mean Square	F	Sig.
Pre-Writing Scores	Between Groups	23.815	3	7.938	1.773	.151
	Within Groups	2140.530	478	4.478		
	Total	2164.345	481			
Improvement Scores (Post-test - Pre-test)	Between Groups	34.848	3	11.616	2.564	.054
	Within Groups	2165.892	478	4.531		
	Total	2200.741	481			

Table 4.13 demonstrates that there are no statistically significant differences among the pre-test scores of the four schools ($F \approx 1.8$, $p=0.151$). In other words, the students in the four schools appear to have performed similarly in the writing assessment pre-test. However, in the improvement scores, there appears to be a difference in the mean improvement scores of the students from the four schools (Table 4.13), since $p=0.054$, thus indicating that there might be greater improvement in the performances of the students in some schools than in others. The means of the improvement scores and the Confidence Intervals of those means were then computed. The findings are displayed in Figure 4.8.

Figure 4.8: Means and Confidence Intervals of the Improvement Scores in the Writing Assessment in the 4 Schools

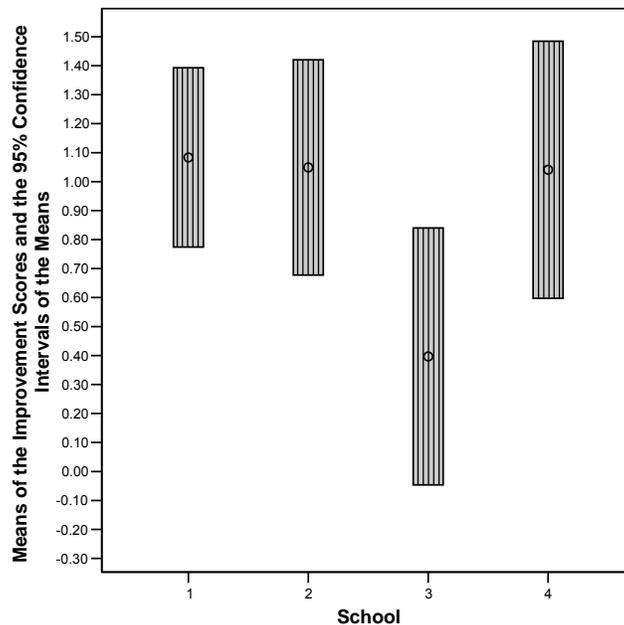


Figure 4.8 suggests that the mean improvement scores of three of the schools, 1, 2 and 4 are very similar. In addition, the Confidence Intervals of these schools indicate that, in repeated experiments, there is a 95% probability that the mean improvement score would be positive since the lower bounds are above zero. On the other hand, the mean improvement score of School 3 is not only much lower, the lower bound of the confidence interval is just below zero (Figure 4.8). There is a probability that the students might get a negative improvement mark in repeated tests. Thus, it appears that after the experimental curriculum, the other three schools had improved much more than School 3. In order to substantiate if this is true, paired t-tests of the post-test and pre-test writing scores were performed. The Effect Sizes, derived from the t values, were also computed. The results are displayed in the adapted Table 4.14 below.

Table 4.14: Paired t-test of the Mean Difference Scores (Post-test – Pre-test) in the Writing Assessment attained by the 4 Schools

School	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
				Lower	Upper				
1 (n=210)	1.083333	2.27404	0.15692	0.77398	1.39269	6.904	209	.000	0.48
2 (n=102)	1.04902	1.88993	0.18713	0.67780	1.42024	5.606	101	.000	0.56
3 (n=97)	0.39691	2.19960	0.22334	-0.04641	0.84022	1.777	96	.079	0.18
4 (n=73)	1.04110	1.89984	0.22236	0.59783	1.48436	4.682	72	.000	0.55

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

Table 4.14 indicates that in all the schools, the students achieved positive mean improvement scores though the mean improvement score of School 3 is much lower than the three other schools. The mean improvement scores of Schools 1, 2 and 4, range from around 1.04 to 1.08 (Table 4.14). In School 3, an all-girls' school, the mean improvement score is comparatively much less, with the students achieving a mean improvement score of only 0.40 (to 2 decimal places). There is also no statistically significant difference between the post-test and the pre-test results of the students from School 3. The mean improvement scores of the students from the other three schools are however statistically significant (Table 4.14). The students from Schools 1, 2 and 4 appear to have improved significantly more than the students from School 3 who may not have improved.

The Effect Sizes of the improvement achieved by the students in the four schools are shown in the last column in Table 4.14. The mean improvement scores of students in School 2, an all-boys' school, and Schools 1 and 4, which are co-educational schools, achieved similar Effect Sizes of 0.56, 0.48 and 0.55 respectively. In contrast, the ES of the improvement attained by students of School 3, an all-girls' school, was a small effect of 0.18 (Table 4.14). The curriculum appears to have been highly effective in Schools 1, 2 and 4, but less effective in School 3. A possible explanation for these differences may rest with gender differences and with PSLE English grades seeming to play a role in the outcome. It was presented earlier in this chapter that the Effect Sizes of the improvement scores for boys and girls with Grades A*, A, and B were similar, though the boys appear to have benefited slightly more (c.f. Table 4.10, p.124). Moreover, boys with Grade C attained a very large ES whereas girls with Grade C achieved a negative ES. Table 4.15 below displays the percentages of the number of students with the different PSLE English grades in the four schools.

Table 4.15: Percentage of Students with the Different PSLE English Grades in the 4 Schools

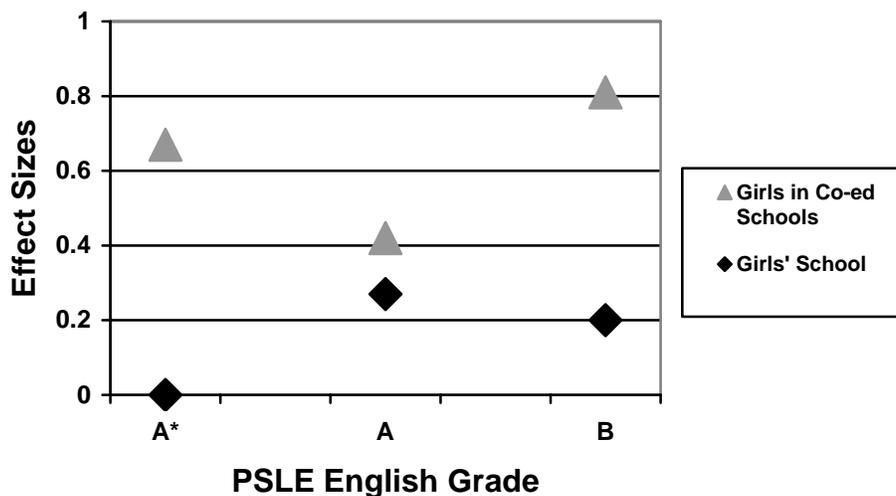
PSLE English Grades	School 1	School 2	School 3	School 4
A*	12.5	84.5	5.5	1.3
A	73.7	14.5	64.2	56.3
B	11.2	0.9	27.5	42.5
C	2.6	0	2.8	0
Total	100	99.9 [#]	100	100.1 [#]

[#] rounding error

Three of the four girls with Grade C came from School 3 while the 2.6% of students with Grade C in School 1 (Table 4.15) included six boys and a girl. The boys with Grade C achieved a high ES (c.f. Table 4.11, p.125) while the girls with the same grade achieved a negative ES (c.f. Table 4.12, p.126). Since almost all the girls with Grade C came from School 3 the poor performance of the girls with Grade C would have affected the overall ES of the performance of the school. In addition, most of the students in School 3 attained PSLE English Grade A (Table 4.15). Girls with Grade A attained a lower improvement ES of 0.37 in

comparison to the improvement ES of 0.43 of their male counterparts (c.f. Table 4.11, p.125 and Table 4.12, p.126). However, it could also be possible that the girls in School 3 may not have performed as well as the other girls in the co-educational schools 1 and 4. If girls with Grade A from the all girls' school did not perform well, their performance could affect the overall ES of the performance of all the girls with Grade A. Therefore, the improvement score ES of the girls in the two co-educational schools as well as the ES for School 3 were calculated. Only the improvements of students with Grades A*, A, and B were calculated; the ES of students with Grade C was not computed because only one of girls came from another school, School 1. The Effect Sizes of the improvements achieved by the respective groups of students are displayed in Figure 4.9.

Figure 4.9: Effect Sizes of the Improvement in the Writing Skills Assessment attained by Girls in the Coeducational Schools and Girls' School



The five girls with Grade A* in School 3 attained a mean improvement score of zero while the 21 girls in School 1 achieved an improvement score ES of 0.67 (Figure 4.9). In School 4, there were no girls with Grade A*. Similarly, the performances of the 108 girls with Grade A in the co-educational schools, Schools 1 and 4, achieved a larger ES of 0.42 compared to their 58 counterparts in School 3 who attained an ES of 0.27. Figure 4.9 also reveals that there is a greater disparity between the improvements achieved by students with Grade B.

There were 27 girls in the co-educational schools with Grade B and a similar number of girls (27) in School 3, an all girls' school. The improvement score ES that the girls from the co-educational schools achieved was 0.81 compared to the small ES of 0.20 achieved by the girls in School 3. Though there are differences in the number of students with the different grades in the respective schools, the evidence points to the likelihood that the curriculum was not as effective in School 3 as it was among the girls in the co-educational schools, Schools 1 and 4. The reason for the lower ES attained by the girls in School 3 may lie in factors other than gender, including the manner in which the programme was implemented in School 3. An attempt is made in the next chapter on Qualitative Findings to discover the possible reason for the disparity (c.f. pp.195-197).

In summary, it appears that the curriculum was highly effective in the all-boys' school, School 2 and in the co-educational schools, Schools 1 and 4. While the curriculum seems to be not as effective in the all-girls' school, School 3, the girls nevertheless showed some sign of improvement in the writing skills post-test even if that improvement was not statistically significant.

At this stage, it must be noted that School 2 had an integrated English and Literature Programme whereas the other three schools did not. However, there were important differences between the LDEP and Literature-based English Programme in School 2 since there was a greater concentration on the teaching of language skills during the LDEP. Therefore, the curriculum in School 2 was accepted as different from the LDEP just like the curricula used in the other schools. Though the control group in School 2, unlike the other control group of students in Schools 3 and 4, improved in the writing skills assessment, the improvement was not as large as the improvement registered by the experimental group of students in School 2. The mean improvement score ES of the experimental group of students in School 2 was 0.56 ($t = 5.606$, $df = 101$) whereas the control group in the same school achieved an ES of 0.23 ($t = 2.365$, $df = 103$).

In the following section, the improvement attained by the students in the various classes in the experimental group and in experimental and control classes taught by the same teachers were analysed to discover if:

- Class culture would have an influence on the outcomes,
- Normal(Academic) class performance matched that of the Express classes, and
- Teacher differences would affect the outcome.

Analysis of the Improvement Achieved by the Different Classes in the Experimental Groups

In Table 4.16, the paired t-test results of the differences between the post-test and the pre-tests as well as the Effect Sizes of these differences attained by the individual classes are presented. Class 5N(A) was the Normal (Academic) class.

Table 4.16: Paired t-test of Improvement in the Writing Skills Assessment by Class

School	Class	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
					Lower	Upper				
1	1	0.85526	2.19307	0.35576	0.13442	1.57611	2.404	37	0.21	0.39
	2	1.47143	2.31346	0.39105	0.67673	2.26613	3.763	34	0.0001	0.64
	4	0.17308	2.15380	0.42239	-0.69686	1.04301	0.410	25	0.685	0.08
	5N(A)	1.81429	2.07253	0.35032	1.10235	2.52622	5.179	34	0.000	0.88
	6	1.12821	2.27896	0.36493	0.38945	1.86696	3.092	38	0.004	0.50
	7	0.85135	2.42632	0.39888	0.04238	1.66033	2.134	36	0.040	0.35
2	11	0.56250	1.93753	0.34251	-0.13605	1.26105	1.642	31	0.111	0.29
	13	1.18571	1.67194	0.28261	0.61138	1.76005	4.196	34	0.000	0.71
	15	1.35714	2.01309	0.34027	0.66562	2.04866	3.988	34	0.000	0.67
3	16	0.38889	2.23642	0.37274	-0.36781	1.14559	1.043	35	0.304	0.17
	18	0.08621	2.08767	0.38767	-0.70790	0.88031	0.222	28	0.826	0.04
	20	0.68750	2.28512	0.40396	-0.13637	1.51137	1.702	31	0.099	0.30
4	23	1.25641	1.85980	0.29781	0.65353	1.85929	4.219	38	0.000	0.68
	24	0.79412	1.94281	0.33319	0.11624	1.47200	2.383	33	0.023	0.41

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

An examination of the Effect Sizes of the performances of the students in the different classes in Schools 1, 2 and 4 reveals that of the 11 experimental classes, only one class, Class 4, attained a slight ES (Table 4.16). The Effect

Sizes with regard to the other classes were moderate to large. The Normal (Academic) Class recorded a large ES of 0.88. Generally, the curriculum appears to have been effective in most of the classes, irrespective of the different teaching styles, class or school cultures, or whether the class was Express or Normal (Academic). With regard to School 3, Class 20 achieved a moderate ES improvement score of 0.30 (Table 4.16). The ES of improvement score attained by Class 16 was 0.17 but if the scores of the three girls with Grade C were excluded, the ES goes up to 0.24. In the light of the fact that the two control classes in School 3 attained negative improvement score Effect Sizes (-0.16 and -0.20 respectively), the curriculum appears to have helped the girls in two of the experimental classes to improve. (For one control class, $t = -0.944$, $df = 36$; for the other control class, $t = -1.250$, $df = 39$). Students in Class 18 appear not to have improved much. The very slight improvement score ES of 0.04 that students from Class 18 achieved would affect the ES of School 3 (Table 4.16). Possible reasons as to why Class 4 in School 1 and Class 18 in School 3 achieved slight Effect Sizes will be explored in the next chapter on Qualitative Findings (c.f. pp.217-220).

Analysis of the Difference Scores of the Experimental and Control Classes Taught by the Same Teachers

In School 4, the experimental and control classes were taught by the same teachers. The following Table 4.17 compares the performances of these classes.

Table 4.17: Paired t-test of the Improvement in the Writing Skills Assessment of the Individual Classes taught by the same teachers in School 4

School	Class	Mean	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
					Lower	Upper				
4	23(E*)	1.25641	1.85980	0.29781	0.65353	1.85929	4.219	38	0.000	0.68
	21(C*)	-0.05556	1.91527	0.31921	-0.70359	0.59248	-0.174	35	0.863	-0.03
	24(E)	0.79412	1.94281	0.33319	0.11624	1.47200	2.383	33	0.023	0.41
	25(C)	-0.42424	1.81194	0.31542	-1.06673	0.21825	-1.345	32	0.188	-0.23

*(E) stands for Experimental class and (C) stands for Control class

**Table adapted to include paired t-test results from the SPSSTM output and the ES.

Classes 23 and 21 were taught by the same teacher. Class 23 was the experimental class and Class 21 was the control class. The experimental class achieved a large ES of 0.68 whereas the control class displayed no improvement since the very slight ES is negative (Table 4.17). Experimental Class 24 and Control Class 25 were taught by the same two teachers, one of whom taught English and the other taught Literature. The ES attained by Class 24 indicates that the experimental curriculum was effective in improving the writing skills of the students but the negative ES achieved by Class 25 (Table 4.17) illustrates that the students did not improve. Since the teacher effect remained the same for the experimental and control classes taught by the same teachers, the contrast between the Effect Sizes achieved by the experimental and control classes demonstrates that the experimental curriculum may have been effective. In the writing assessment, most of the students in the experimental group improved whereas most of the students in the control group maintained their pre-test scores. In addition, there appears to be general improvement in the performance of the students in the experimental group even across the following mediating variables: differing language abilities, gender, school, Normal (Academic) or Express stream, class or varying teaching styles. In conclusion, it may be presumed that the experimental curriculum was effective in improving the writing skills of many of the students.

Analysis of the Reading Comprehension Pre-Test and Post-Test Scores

In these multiple site case studies, there were seventeen experimental classes which were taught the experimental curriculum and eight control classes which were taught under the usual curriculum in four separate schools. However, for reasons already stated at the beginning of this chapter (c.f. p.101), one of the experimental classes, a Normal (Academic) class, had to be excluded from the analysis. Thus, in the final analysis, only the results from sixteen experimental and eight control classes were included. It is once more noted that the reading comprehension skills was a minor component of the LDEP since a greater

emphasis was given to the writing skills. Due to the short duration of the programme more time was spent on the explicit and implicit teaching of writing skills. However, through the literary analysis of the text and the teaching of writing skills it was hoped that the comprehension skills of the students would improve as well. Therefore, the evidence offered in this section of any improvement in the reading comprehension skills of the students in the experimental group is in support of evidence presented of the improvement in the writing skills displayed by the students in the experimental group to indicate that the LDEP was effective in improving the language skills of the students. In the experimental curriculum, reading comprehension skills were taught differently from the control classes. In the control classes in all the schools, the students had the usual lessons on comprehension skills that included completing comprehension exercises with extracts as passages. These lessons would normally involve the students completing the comprehension exercises and the teachers going over the answers. In some schools, the students were also taught techniques on how to answer comprehension questions. The experimental group, on the other hand, was not given comprehension exercises unrelated to the literature text; rather, they were guided through the analysis of the content and language of the literature text. Therefore, in analysing the outcomes of the reading comprehension pre-test and post-test, the intention was to discover if the experimental programme would lead to an improvement in reading comprehension skills through the analysis of the literature text. In addition, it was hoped that a comparison with the control group would yield answers to whether the teaching of reading comprehension skills through linking literary analysis with writing would aid just as well in the improvement of these skills as the usual custom of teaching comprehension skills through the use of distinct comprehension exercises.

In one school (School 1), since the school wanted all the Secondary One Express and Normal (Academic) classes to participate, for practical reasons the classes were taught under the experimental curriculum in two phases. In the first

phase, five classes, including a Normal (Academic) class, were taught the experimental curriculum while the other four classes in the level followed the usual curriculum. These four classes then took part in the second phase during which the first group of five classes returned to their usual curriculum. One of the four classes in Phase 2 was the Normal (Academic) class that was excluded from the final analysis (c.f. p.101). Since the post-test that the students in Phase 1 completed coincided with the pre-test that the students in Phase 2 completed, the post-test of the Phase 1 classes became the pre-test of the Phase 2 classes. Subsequently, the pre-test reading comprehension assessment of the classes in Phase 1 became the post-test of the classes taking part in Phase 2. Please refer to Table 3.2, p.73 for a tabular representation of the implementation in phases. Two sets of different comprehension passages, Passage A and Passage B, extracted from the same story, were used for the pre-test and post-test (c.f. Appendices 3a and 4a). In Schools 1 (Phase 1), 2, 3 and 4 Passage A was the pre-test and Passage B was the post-test and in School 1 (Phase 2) Passage B was the pre-test and Passage A was the post-test. Lists of acceptable and unacceptable answers were created (c.f. Appendix 5b). These lists were open to scrutiny, and teachers who taught the experimental and control classes had access to them. As with the writing skills assessment, all the pre-test and all the post-test scripts were marked before the scores were recorded. This was to ensure that the pre-test and post-test marks of any particular individual or group would be difficult to track, thereby acting as a safeguard against unreliable or inconsistent marking by the researcher.

An additional safeguard was taken to ensure that the marking was reliable. Teachers from each of the participating schools marked the pre-test and post-test scripts of five students. In Schools 2, 3 and 4, Passage A was the pre-test and Passage B was the post-test. In School 1, since the scripts of students in Phase 2 were given to the teachers to be marked, Passage B was the pre-test and Passage A, the post-test. The difference scores between the post-test and the pre-test awarded by the teachers were recorded, and then compared to the

difference scores awarded by the researcher for the same scripts. The difference scores were used because in the analysis of the test scores of the students, the focus was on the difference scores between the post-test and pre-test as these indicate improvement or lack of improvement in the performance of the students.

The paired t-test between difference scores awarded by the teachers and the researcher yielded a mean of zero with a two-tailed significance of 1. The computation of the regression between the scores resulted in a constant of -0.005 and $\beta = 0.817$. The disparity between the difference scores that the teachers and the researcher awarded for the same scripts appears to be negligible.

On completing the marking of the reading skills test papers, it was found that on the average, students received fewer marks in the test with Passage B than in the test with Passage A. It is important that the tests with Passages A and B are similar in terms of level of difficulty for two reasons. While the test with Passage A was the pre-test and the test with Passage B was the post-test for many of the students, the reverse was true for other students. The students in School 1 (Phase 2) completed the test with Passage B as the pre-test and the test with Passage A as the post-test. There can be no comparability between the two groups if the tests are of different difficulty levels. It is also important to ensure that the tests are similar so that assessment of any improvement in the reading comprehension skill of the students would be reliable.

In order to ascertain if there is a difference in the level of difficulty between Passage A and Passage B, a third group of 179 Secondary One students from two schools were given the tests to complete. This group was outside the sample of the four schools. These students came from four Express classes and one Normal (Academic) class and they completed the reading comprehension tests on the same day. The reading comprehension test with Passage A is

termed as Test A and the test with Passage B is termed as Test B. After marking the test papers of these students, the scores of each student were kept as paired scores. On comparing these paired scores, it was found that they scored on average, about 1.79 marks less in Test B than in Test A (Table 4.18). The paired t-test results in Table 4.18 indicate that there is a significant difference between Test B and Test A ($t = -6.7, p < 0.05$). In other words, since the tests were completed on the same day, it would be safe to assume that Test B was more difficult than Test A.

Table 4.18: Paired t-test of Test B and Test A in the Reading Comprehension Assessment

	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Test B Scores – Test A Scores	-1.78771	3.55213	0.26550	-2.31164	-1.26378	-6.733	178	.000

It was then decided to adjust the Test B scores using a regression model so that Test A would be equivalent to Test B. The results are displayed in Table 4.19.

Table 4.19: Results derived from the Regression Model

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.672 ^a	.451	.448	3.36773

a. Predictors: (Constant), Test B

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1652.421	1	1652.421	145.695	.000 ^a
	Residual	2007.468	177	11.342		
	Total	3659.888	178			

a. Predictors: (Constant), Test B

b. Dependent Variable: Test A

Table 4.20: Coefficients from the Regression Model

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	5.517	.851		6.480	.000
Test B	.725	.060	.672	12.070	.000

a. Dependent Variable: Test A

Using the information from Table 4.20, Test B results were adjusted using the formula:

$$\begin{aligned} \text{Test B adjusted} &= \text{Constant} + (\text{coefficient} \times \text{Test B original}) \\ &= 5.517 + (0.725 \times \text{Test B original}) \end{aligned}$$

The adjusted Test B results and the Test A scores were then paired in a paired t-test to determine if the regression approach would work in levelling the difference between the pre-test and the post-test. The following results were obtained:

Table 4.21: Paired t-test of the Difference in the Reading Comprehension Scores attained in Adjusted Test B and Test A

	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Lower	Upper			
Adjusted Test B Scores – Test A Scores	0.00373	3.35826	0.25101	-0.49160	0.49907	0.015	178	.988

With Test B scores adjusted, there is no statistically significant difference between the Test A and Adjusted Test B scores since p is greater than 0.05 and the mean difference is a negligible 0.004 (to 3 decimal places). Through the adjustment made by following the regression model, the t-test shows that the mean score marks that the students achieved from Test A and the Adjusted Test B are very similar. Consequently, all Test B scores of the participating students in Schools 1, 2, 3 and 4 were adjusted using the formula:

$$\text{Test B adjusted} = 5.517 + (0.725 \times \text{Test B Original})$$

It is once again noted that Test B was the post-test for the students in Schools 1 (Phase 1), 2, 3 and 4 and the pre-test for the students in Phase 2, School 1.

Comparison of the Reading Comprehension Results of the Experimental Classes and the Control Classes in Schools 3 and 4

On analysing the results it was found that the top students in the experimental and control groups displayed little improvement in the post-test. Figure 4.10 below displays the mean improvement scores of all the students in the experimental and control groups with the different PSLE English grades.

Figure 4.10: Mean Improvement Scores and their Confidence Intervals in the Reading Comprehension Assessment by PSLE English Grades

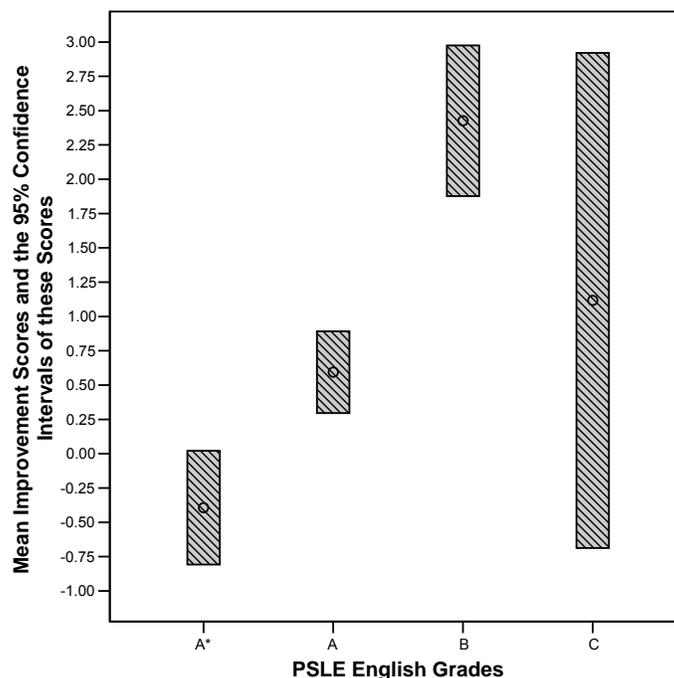


Figure 4.10 demonstrates that students with A* showed no improvement. The mean improvement score is less than zero and the upper and lower bounds of the Confidence Interval are at and below zero (Figure 4.10). In addition, the following Table 4.22 shows that more than 65% of the students with Grade A* had scored 68% or more in the pre-test. In the post-test, about 54% of the same group of students scored 68% or more.

Table 4.22: Percentage of Students with Grade A* scoring 68% or more in the Pre-test and Post-test

Range of Marks (% of Test Score)	Percentage of Students	
	Pre-test	Post-test
20 – 23 (80% – 92%)	23.1%	10.8%
17 – <20 (68% – <80%)	42.5%	43.6%
Overall	65.6%	54.4%

Table 4.22 reveals that more than three-fifths of the students with Grade A* were already achieving high scores in the pre-test, thereby indicating that these students were already performing at a high level on this type of test. In contrast, 30.6% of the students with Grade A, 1.8% of students with Grade B and none of the students with Grade C scored within the range of 17 to 23 marks in the pre-test. Thus, the post-test may not have been challenging for most of the students with Grade A*, and that may be the reason why the percentage of students scoring the same range of marks dropped to just over half of them (Table 4.22).

Regression towards the mean suggests that a group of students scoring highly in the first test might have lower mean scores in the second test (Shaughnessy & Zechmeister 1990). This phenomenon might explain why as a group the students with Grade A* did not do as well in the post-test. Since the mean improvement score for the students from the control and experimental groups with Grade A* was below zero, it might be assumed that for these groups of students there was no change in performance in the second test. Moreover, as over half of the students with Grade A* still managed to score 68% and more in the second reading comprehension test, it is probable that these students needed a more challenging test than the uniform one that was given to all students. It must be noted, however, that with regard to the lowest-scoring students, there were only three students with PSLE Grade C in the experimental group (c.f. Table 4.26, p.148). Since no comparison could be made with the control group, the performance of this group is not included in the discussion of the findings.

As students with PSLE English Grade A* in both the control and experimental groups displayed either no or little improvement, it was decided that these students would be excluded from the analysis. Thus, the students with A* from Schools 3 and 4 and all the Express students from Schools 1 and 2 were excluded from the analysis. All the students from the Express stream from Schools 1 and 2 were excluded because most of the Express students in Schools 1 and 2 who attained PSLE Grade A would have scores that were close to that needed to achieve Grade A*. This conclusion was reached based on the PSLE English grades of the students and the aggregate PSLE scores of the students in these two schools when compared with the grades and scores of the students in Schools 3 and 4. More than 90% of the students in the Special/Express classes in Schools 1 and 2 had obtained Grades A* and A (c.f. Table 4.1, p.100) in the PSLE. Additionally, in 2004, the PSLE aggregate scores of the Express students who entered Secondary One in Schools 1 and 2 were much higher than the aggregate scores of the students who entered Schools 3 and 4. The aggregate scores are given in Table 4.23.

Table 4.23: Mean of the Aggregate PSLE Scores* of Express Students who entered Schools 1, 2, 3 and 4 in 2004

School	Non-Affiliation	Affiliation [#]
	Mean	Mean
School 1	250	
School 2	267	
School 3	230	210
School 4	220	

[#] Under 'Affiliation', the aggregate score of students who entered the secondary school from an affiliated primary school is given. Only one school, School 3, has an affiliated primary school.

*Source: The respective school information sites, <http://app.sis.moe.gov.sg/schinfo/index.asp>

In the PSLE, students are tested on English Language (100 marks), Mother Tongue (100 marks), Mathematics (50 marks) and Science (50 marks). For each student, the Ministry of Education reports the grades of the individual subjects and an aggregate score. The maximum aggregate score for the four subjects is 300. Thus, from the mean aggregate scores in Table 4.23, it may be deduced

that, on the average, in the respective schools, each Express student would obtain the following marks in English:

Table 4.24: Calculation of the Mean PSLE English Scores from the aggregate scores

Schools	Calculation	Mean PSLE English Score
1	$(250/300) \times 100$	83%
2	$(267/300) \times 100$	89%
3	$(230/300) \times 100^{\#}$; $(210/300) \times 100^{\#\#}$	77% [#] ; 70% ^{##}
4	$(220/300) \times 100$	73%

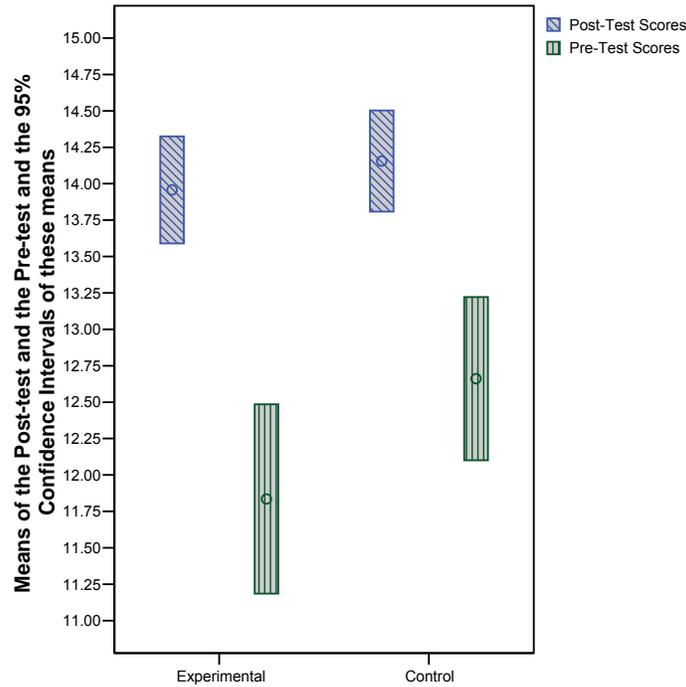
[#] Of students from non-affiliated primary schools

^{##} Of students from affiliated primary schools

In order to get a grade of A*, the students would have to obtain a score of 90% or more. The range of scores for Grade A is wide, going from a minimum of 75% to 89%. Since the average PSLE English scores of students in Schools 1 and 2, being just 7% and 1% less than the 90% (Table 4.24), were close to scores needed for Grade A*, it was decided that these schools would be excluded from the analysis. Students from Schools 3 and 4 had mean English scores that were much lower than the score needed for Grade A* (Table 4.24). Most of the students from these schools with Grade A would have attained scores well below that needed for an A*. Thus, the scores of the students with PSLE Grade A, as well as those with Grade B, from these schools were analysed.

The means and the Confidence Intervals of the means of the pre-test and the post-test of the experimental and control groups in Schools 3 and 4 were computed, and these are displayed in Figure 4.11.

Figure 4.11: Means and the Confidence Intervals of the Post-test and Pre-test Scores in the Reading Comprehension Assessment by Treatment Groups



Though the mean of the post-test of the experimental group in the assessment of the reading comprehension is lower than the post-test mean of the control group, the mean of pre-test scores of the experimental group is much lower (Figure 4.11). The experimental group appears to have narrowed the gap and caught up with the control group, thereby indicating that the experimental group had improved more than the control group. Paired t-tests of the reading comprehension post-test and pre-test scores of the experimental and control groups were next analysed. The results and the Effect Sizes are presented in Table 4.25.

Table 4.25: Paired t-tests of the Improvement Scores in the Reading Comprehension Assessment by Treatment Groups in Schools 3 and 4

	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
				Lower	Upper				
Experimental	2.12226	3.33722	0.25155	1.62579	2.61872	8.437	175	0.000	0.64
Control	1.49425	3.46401	0.25197	0.99720	1.99130	5.930	188	0.000	0.43

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

There is a statistically highly significant difference between the post-test and the pre-test for both the experimental and control groups, since $p < 0.001$ (Table 4.25). However, the mean difference score of the experimental group is greater than the mean difference score of the control group. Moreover, the ES of the mean improvement score of the experimental group at 0.64 is greater than the ES of the mean improvement score of the control group which was 0.43. The experimental group in Schools 3 and 4 appears to have performed better than the control group in the same schools even though the experimental group did not work on distinct reading comprehension exercises on which the control group worked. The students in the experimental group may have improved through the process of doing literary analysis and through exploring the use of language in the literature text. It may also be pertinent to note that, in all the schools, the control classes would have continued with same form of instruction in comprehension skills that they had in the first half of the school year. Though the experimental classes had the benefit of the same form of instruction during the first two terms in the year, in Term 3, the control classes continued to have the same form of instruction while the experimental classes did not. The control classes would have had the benefit of having the comprehension skills reinforced whereas the experimental classes did not have these skills reinforced during the duration of the programme.

In the next section, the improvement that the students achieved in the post-test in the reading comprehension assessment was scrutinised to ascertain if differing language fluency, as determined through the PSLE English grades of the students, may have had an influence on the outcomes.

PSLE Grades and Improvement Scores in the Reading Comprehension Assessment by Treatment Groups in Schools 3 and 4

In Table 4.26, the results of the paired t-tests of the mean difference scores between the post-test and pre-test and the Effect Sizes of the improvements achieved by the Experimental and Control Groups are displayed.

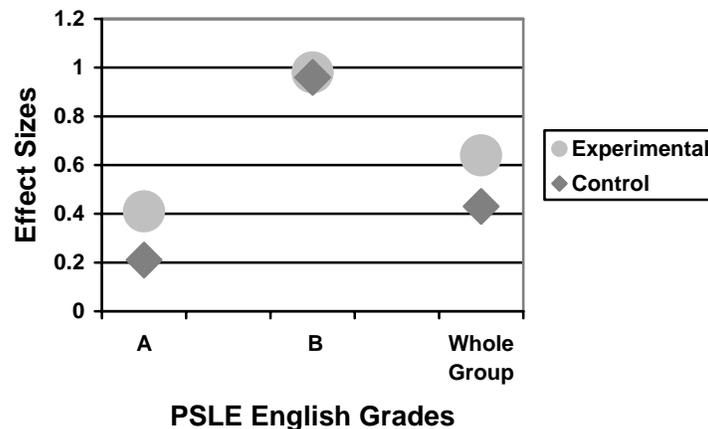
Table 4.26: Paired t-tests of Improvement Scores by PSLE English Grades and Treatment Groups in Schools 3 and 4

PSLE Grade		Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
					Lower	Upper				
A	Experimental	1.26490	3.08659	0.29839	0.67331	1.85649	4.239	106	0.000	0.41
	Control	0.71938	3.35994	0.29933	0.12698	1.31179	2.403	125	0.018	0.21
B	Experimental	3.21743	3.29826	0.43308	2.35020	4.08466	7.429	57	0.000	0.98
	Control	3.06372	3.19770	0.40942	2.24475	3.88269	7.483	60	0.000	0.96
C	Experimental	4.89200	2.97027	1.71489	-2.48656	12.27056	2.853	2	0.104	1.65

*Table adapted to include paired t-test results from the SPSS™ output and the ES.

With respect to students with Grade A, the mean improvement score and the ES of the experimental group are greater than those of the control group. The ES of the mean improvement score is 0.41 for the experimental group and 0.21 for the control group (Table 4.26). In the case of the students with Grade B, statistically there is a highly significant difference between the mean improvement scores of both the experimental and control groups since $p < 0.001$. The mean difference score and the ES of this group of students in the experimental group are greater than those of their counterparts in the control group. The margin is however slight with the improvement score ES of the experimental group being 0.98 and the ES of the control group being 0.96 (Table 4.26). Figure 4.12 shows a comparison of the Effect Sizes between the experimental and control groups as a whole and on the basis of the PSLE English grades of the students.

Figure 4.12: Effect Sizes of the Improvement Scores by Treatment Groups and PSLE English Grades



Students with Grade A from the experimental group achieved a larger ES than the students in the control group (Figure 4.12). However, with regard to the students with Grade B, the gap is marginal (Figure 4.12) with both the experimental and control groups having similarly large Effect Sizes. Both groups improved greatly in the post-test over the pre-test though the experimental group improved slightly more than the control group (Figure 4.12). Overall Effect Sizes of the experimental and control groups show that generally the experimental group, irrespective of PSLE English Grades, had registered greater improvements in the reading comprehension assessment than the control group.

Gender and PSLE English Grades and Improvement Scores in the Reading Comprehension Assessment by Treatment Groups

The improvement scores of the students in the experimental as well as control groups in Schools 3 and 4 were analysed with respect to gender as well as the PSLE English Grades of the boys and girls. All the boys came from School 4, a co-educational school, since School 3 was an all-girls' school.

The paired t-test results of the post-test and pre-test of the reading comprehension assessment that the boys and girls completed and the ES of the improvements gained are displayed in Table 4.27.

Table 4.27: Paired t-tests of Improvement Scores in the Reading Comprehension Assessment by Gender and Treatment Groups in Schools 3 and 4

		Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
					Lower	Upper				
Boys	Experimental	2.09771	2.27599	0.38471	1.31589	2.87954	5.453	34	0.000	0.92
	Control	1.68589	3.41302	0.50878	0.66050	2.71127	3.314	44	0.002	0.49
Girls	Experimental	2.12835	3.55852	0.29968	1.53586	2.72083	7.102	140	0.000	0.60
	Control	1.43436	3.48941	0.29078	0.85957	2.00915	4.933	143	0.000	0.41

**Table adapted to include paired t-test results from the SPSSTM output and the ES.*

The paired t-test results in Table 4.27 indicate a statistically high significance between the post-test and pre-test in the reading comprehension assessment though the significance is slightly lower for the control group of boys. The mean

improvement scores are greater for both the boys and girls in the experimental group than in the control group. The Effect Sizes attained by both boys and girls in the experimental group are larger than those achieved by the control group of boys and girls. In the experimental group, the boys attained a large improvement ES of 0.92 as opposed to the much lower ES of 0.49 of the control group of boys (Table 4.27). With regard to the girls, the experimental group achieved an improvement ES of 0.6 in contrast to the lower ES of 0.41 attained by the control group (Table 4.27). In both cases, the boys and girls in the experimental group improved more than those in the control groups.

The next step taken was to compare the performances of the boys with PSLE English Grades A and B in the experimental and control groups. There was only one boy in the control group with a Grade C and so no comparison could be made.

The results of the paired t-tests of the post-test and the pre-test as well as the ES of the improvement achieved by the male students in School 4 are presented below in Table 4.28.

Table 4.28: Paired t-tests of Improvement Scores of Boys in the Reading Comprehension Test by PSLE English Grades and Treatment Groups in School 4

PSLE Grade		Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
					Lower	Upper				
A	Experimental	1.55152	2.28175	0.49792	0.51288	2.59017	3.116	20	0.005	0.68
	Control	0.98218	3.49241	0.66000	-0.37204	2.33640	1.488	27	0.148	0.28
B	Experimental	2.91700	2.08167	0.55635	1.71508	4.11892	5.243	13	0.000	1.40
	Control	2.78731	3.11576	0.77894	1.12704	4.44758	3.578	15	0.003	0.66

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

Table 4.28 indicates that in all cases, except for the control group of students with Grade A, there are statistically significant differences between the pre-test and post-test since $p < 0.05$. The experimental group of students with Grades A and B and the control group of students with Grade B appear to have improved significantly more than the control group of students with Grade A. The mean

improvement scores of the experimental group of students with Grades A and B are greater than the mean improvement scores of the control group of students with the corresponding grades (Table 4.28). The experimental group of students appears to have improved more than their counterparts in the control group.

Table 4.28 reveals that the boys with Grades A and B in the experimental group attained much larger improvement score Effect Sizes of 0.68 and 1.4 than their counterparts in the control group with improvement Effect Sizes of 0.28 and 0.66 respectively. This indicates that there is a greater likelihood that the improvement of the boys in the experimental group would be more educationally significant than the improvement displayed by the control group of boys.

In Table 4.29, the ES of the mean improvement scores and the results of the paired t-tests of the reading comprehension post-test and pre-test scores of the girls in Schools 3 and 4 are presented. Since there were students with Grade C in only the experimental group, their scores were excluded from the analysis as no comparison could be made with the control group.

Table 4.29: Paired t-tests of Improvement Scores of Girls in the Reading Comprehension Assessment by PSLE English Grades and Treatment Groups in Schools 3 & 4

PSLE Grade		Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
					Lower	Upper				
A	Experimental	1.19491	3.26044	0.35158	0.49587	1.89395	3.399	85	0.001	0.37
	Control	0.64430	3.33575	0.33696	-0.02448	1.31307	1.912	97	0.059	0.19
B	Experimental	3.31302	3.61545	0.54505	2.21382	4.41222	6.078	43	0.000	0.92
	Control	3.09257	3.25896	0.49131	2.10175	4.08338	6.295	43	0.000	0.95

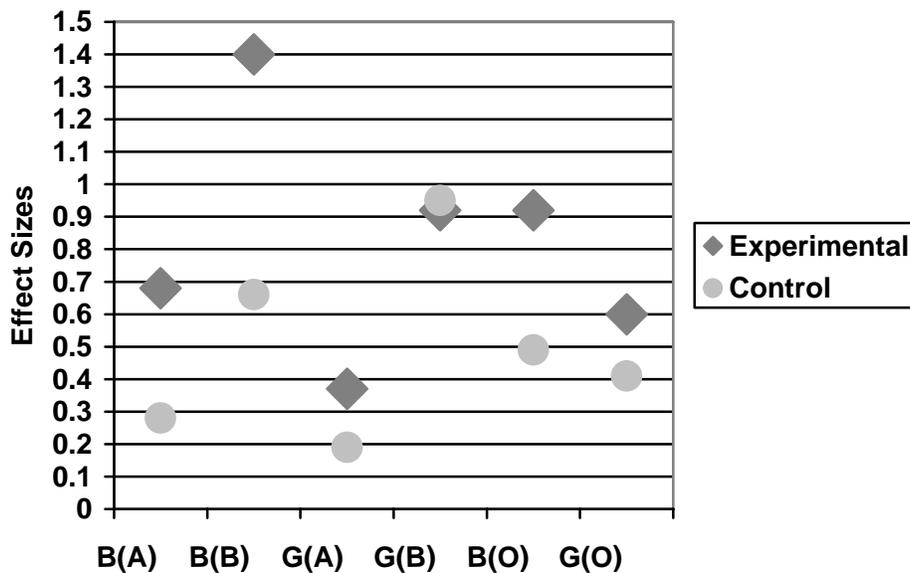
**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

The results of the paired t-tests, as displayed in Table 4.29, reveal some resemblance to the results attained by the boys in School 4. There is a statistically significant difference between the pre-test and the post-test scores for students in the control group with Grade B and for students in the experimental group with Grades A and B since $p < 0.05$ (Table 4.29). The mean improvement scores registered by the girls in the experimental group are greater

than those of the control group though the difference between the girls with Grade B is negligible. The Effect Sizes indicate that the girls with Grade A in the experimental group achieved a moderate improvement score ES whereas their counterparts in the control group attained a small ES of 0.19 (Table 4.29). The Effect Sizes of the improvement scores of the girls with Grade B in the experimental and control groups are similarly large. The improvements achieved by the girls in the control and experimental groups appear to be of similarly important increments.

A more detailed study of the Effect Sizes of the improvements achieved by the boys and girls in the experimental and control groups is undertaken in the next section so as to discover the relationship between gender and the LDEP and the usual mode of teaching reading comprehension skills adopted in the schools. Figure 4.13 displays the Effect Sizes of the improvements achieved by the boys and girls as a whole and by PSLE English grades.

Figure 4.13: Effect Sizes of the Improvement by Gender, PSLE English Grades and Treatment Groups



B(A) stands for Boys with Grade A; B(B) – Boys with Grade B; G(A) – Girls with Grade A; G(B) – Girls with Grade B; B(O) – Boys' Overall ES; G(O) – Girls' Overall ES

The Effect Sizes of the various groups of boys and girls with different PSLE English grades in Figure 4.13 reveal that in general the boys and girls from the experimental group appear to have improved more than the boys and girls from the control group. As illustrated by Figure 4.13, only one group of girls with Grade B in the control group has a slightly larger ES (0.95) than the girls with the same grade in the experimental group (0.92). However, the difference is slight and both groups of girls appear to have benefited, whether they had learnt reading comprehension skills through the LDEP or through the usual curricula of the schools. However, in the case of the boys, there is a large difference between the experimental group and control group. The experimental curriculum appears to have been highly effective for boys with Grades A and B in the experimental group with the boys registering much larger Effect Sizes than their counterparts in the control group (Figure 4.13).

There are similarities in the results of the boys and girls when a comparison is made between the experimental and control groups. There are also marked differences in the Effect Sizes of the improvement scores achieved by the boys and girls. Tables 4.27 to 4.29 and Figure 4.13 (c.f. pp.149-151 and 152) reveal that the boys in the experimental group displayed greater improvement than the experimental and control groups of girls, irrespective of PSLE English grades. These boys displayed greater improvement than the boys in the control group as well. The results may imply that with boys reading comprehension skills may be enhanced through the contextual and literary analysis of the literature text. The gap between the Effect Sizes attained by the boys in the experimental and control groups (Figure 4.13) is also wide. It may imply that the LDEP is better than the traditional mode of teaching comprehension skills in improving the reading comprehension skills of the boys.

In the final section, the findings from the results of the individual schools, School 3 and School 4, are presented. Overall, the experimental group appears to have done better than the control group (c.f. Table 4.25, p.146). However, would this

mean that in the individual schools the experimental group performed better than the control group as well? The outcomes of the different groups in Schools 3 and 4 were studied to ascertain if in each of these schools, the experimental group of students also registered greater improvement than the control group of students.

Reading Comprehension Assessment of the Experimental and Control Groups in School 3

The paired t-test results of the reading comprehension post-test and the pre-test and the ES of the improvement achieved by the students in the experimental and control groups in School 3, an all-girls' school, are displayed in Table 4.30.

Table 4.30: Paired t-tests of Improvement Scores (Post-test – Pre-test) in the Reading Comprehension Assessment in School 3 by Treatment Groups

	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	Df	Sig. (2-tailed)	ES
				Lower	Upper				
Experimental	2.27926	3.61386	0.35437	1.57645	2.98207	6.432	103	0.000	0.63
Control	1.39567	3.56791	0.41199	0.57476	2.21657	3.388	74	0.001	0.39

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

Both the experimental and control groups in School 3 improved (Table 4.30). However, the experimental group achieved a larger mean improvement score ES of 0.63 compared to the improvement score ES of 0.39 of the control group (Table 4.30). In spite of having a focused instruction on comprehension skills, the control group of classes displayed less improvement than the experimental group. The LDEP appears to have been highly effective in improving the reading comprehension skills of the students in the experimental group.

Reading Comprehension Assessment of the Experimental and Control Groups in School 4

School 4 is a co-educational school and the results of School 4 mirrors that of School 3 in that the experimental group displayed greater improvement in the post-test than the control group. The results of the paired t-tests of the reading

comprehension post-test and the pre-test and the improvement score ES are presented in Table 4.31.

Table 4.31: Paired t-tests of the Improvement Scores (Post-test – Pre-test) in the Reading Comprehension Assessment in School 4 by Treatment Groups

	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	Df	Sig. (2-tailed)	ES
				Lower	Upper				
Experimental	1.89547	2.90103	0.34189	1.21376	2.57718	5.544	71	0.000	0.65
Control	1.55911	3.40828	0.31922	0.92668	2.19153	4.884	113	0.000	0.46

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

The experimental group attained a larger improvement ES of 0.65 compared to the improvement ES of 0.46 achieved by the control group (Table 4.31), indicating that the experimental group of students had improved more than the control group of students. The Effect Sizes of the improvement registered by the students in the experimental groups in Schools 3 and 4 are similar since they are around 0.60 (Tables 4.30 & 4.31). In summary, it may be stated that the experimental curriculum was highly effective in enhancing the comprehension skills of the students.

Since the experimental group showed greater improvement, the outcomes of the individual classes in the experimental group in the two schools were examined in order to see if differences between classes had an impact on the outcomes. The paired t-test results of the reading comprehension post-test and pre-test and the Effect Sizes of the improvements shown by the students in the individual classes are displayed in Table 4.32 below.

Table 4.32: Paired t-tests of the Improvement Scores of the Experimental Classes in the Reading Comprehension Assessment in Schools 3 and 4

School	Classes	Mean Difference	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		T	df	Sig. (2-tailed)	ES
					Lower	Upper				
3	16	4.29762	3.76972	0.59604	3.09201	5.50324	7.210	39	0.000	1.14
	18	0.69986	3.00581	0.50808	-0.33268	1.73239	1.377	34	0.177	0.23
	20	1.40148	2.74622	0.50996	0.35688	2.44609	2.748	28	0.010	0.51
4	23	1.85416	2.89399	0.47577	0.88926	2.81907	3.897	37	0.000	0.64
	24	1.93914	2.95006	0.49865	0.92576	2.95252	3.889	35	0.000	0.66

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

A comparison of the Effect Sizes of the improvement registered by the individual experimental classes in Schools 3 and 4 in Table 4.32 reveals that the LDEP was highly effective in improving reading comprehension skills in all the classes except Class 18 in School 3. Though the mean improvement score ES achieved by Class 18 is smaller by comparison, the ES of 0.23 still shows that the curriculum had some positive effect. Therefore, the experimental curriculum was effective irrespective of class differences.

With the removal of one Normal (Academic) class from analysis, only the performance of one Normal (Academic) class, Class 5, remained to be analysed. Even though the number of students, at forty, is small, and there were no control Normal (Academic) class, it was felt that analysis of the performance of the students in Class 5 would help in ascertaining if the experimental curriculum could be as successful with Normal (Academic) students as with the Express students. In the following section, the outcomes of the Normal (Academic) class and the sub-groups within that class are presented. The paired t-test results between the reading comprehension post-test and the pre-test and the Effect Sizes of the improvement achieved by the students in the Normal (Academic) Class 5 are displayed in Table 4.33.

Table 4.33: Paired t-tests of the Improvement Scores (Post-test – Pre-test) in the Reading Comprehension Assessment of Class 5

	Improvement Scores	Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)	ES
				Lower	Upper				
Overall	3.52387	3.21716	0.50868	2.49498	4.55277	6.928	39	0.000	1.10
Boys	3.83724	3.32322	0.72519	2.32453	5.34995	5.291	20	0.000	1.15
Girls	3.17753	3.148530	0.72232	1.65999	4.69507	4.399	18	0.000	1.01
ALL									
Grade A	1.12950	3.72999	2.63750	-32.38311	34.64211	0.428	1	0.742	0.30
Grade B	3.21355	3.23462	0.60065	1.98317	4.44393	5.350	28	0.000	0.99
Grade C	5.27325	2.87851	1.01771	2.86675	7.67975	5.181	7	0.001	1.83
BOYS									
Grade B	3.45986	3.53701	0.94531	1.41765	5.50206	3.660	13	0.003	0.98
Grade C	4.80450	3.17371	1.29566	1.47390	8.13510	3.708	5	0.014	1.51
GIRLS									
Grade A	1.12950	3.72999	2.63750	-32.38311	34.64211	0.428	1	0.742	0.30
Grade B	2.98367	3.03173	0.78279	1.30475	4.66258	3.812	14	0.002	0.98
Grade C	6.67950	1.53796	1.08750	-7.13850	20.49750	6.142	1	0.103	4.34

**Table adapted to include paired t-test results from the SPSS™ output and the ES.*

Table 4.33 demonstrates that the Normal (Academic) students in Class 5 had improved substantially by a relatively large mean improvement ES of 1.10. In addition, the boys and girls appeared to have improved to a similar extent with the improvement ES for boys being slightly higher at 1.15 and the girls attaining an ES of 1.01 (Table 4.33). Boys and girls with Grades B improved greatly, attaining the same improvement ES of 0.98 (Table 4.33). With regard to the girls with Grades A and C, and boys with Grade C, since the sample sizes are small, little importance can be placed on the results of the students with these grades.

In conclusion, the curriculum appears to have benefited the students with Grades A and B in the Express and Normal (Academic) classes in Schools 3, 4 and 1. Imparting comprehension skills contextually and through literary analysis during the LDEP seems to have been highly effective in most cases.

Though the teaching of writing skills and reading comprehension formed the major and minor components of the LDEP respectively, the students in the experimental group appear to have improved in both the writing as well as the reading comprehension skills assessments. These students in general displayed greater improvement than the control group of students. In the writing skills assessment, the students in the experimental group improved irrespective of differences in PSLE English Grades, schools, classes and gender. In the reading comprehension skills assessment, only the results of the students with Grades A and B in Schools 3 and 4 were analysed since it was felt that the test instruments were unchallenging for the students with Grade A*. There were too few students with Grade C. Students with Grades A and B in the experimental group improved more than the students with similar grades in the control group. These students generally improved more than the control group of students in the different sub-groups of analysis based on gender, PSLE English Grades and school.

Having analysed the performances of the students in the pre-test and post-test, it would now be useful to study the perceptions of the students regarding the experimental programme through the survey they were asked to complete at the end of the programme.

SURVEY FINDINGS

All seventeen experimental classes completed a survey (c.f. Appendix 8). However, only the responses of sixteen classes (n = 604) were analysed since it was decided to exclude one of the classes, Normal (Academic) Class 9, from all analysis of the data collected for reasons stated at the beginning of this chapter on p.101.

In the survey that the students were requested to complete, they responded to:

- The number of grammar concepts they had grasped during the LDEP (Literature-Driven English Programme), and
- Whether they felt they had acquired specific writing skills or had improved in their writing in general.

Figure 4.14 shows the percentages of students who reported grasping none, or one or more of the grammar concepts taught.

Figure 4.14: Students' Self-Report of Grammar Concept Acquisition

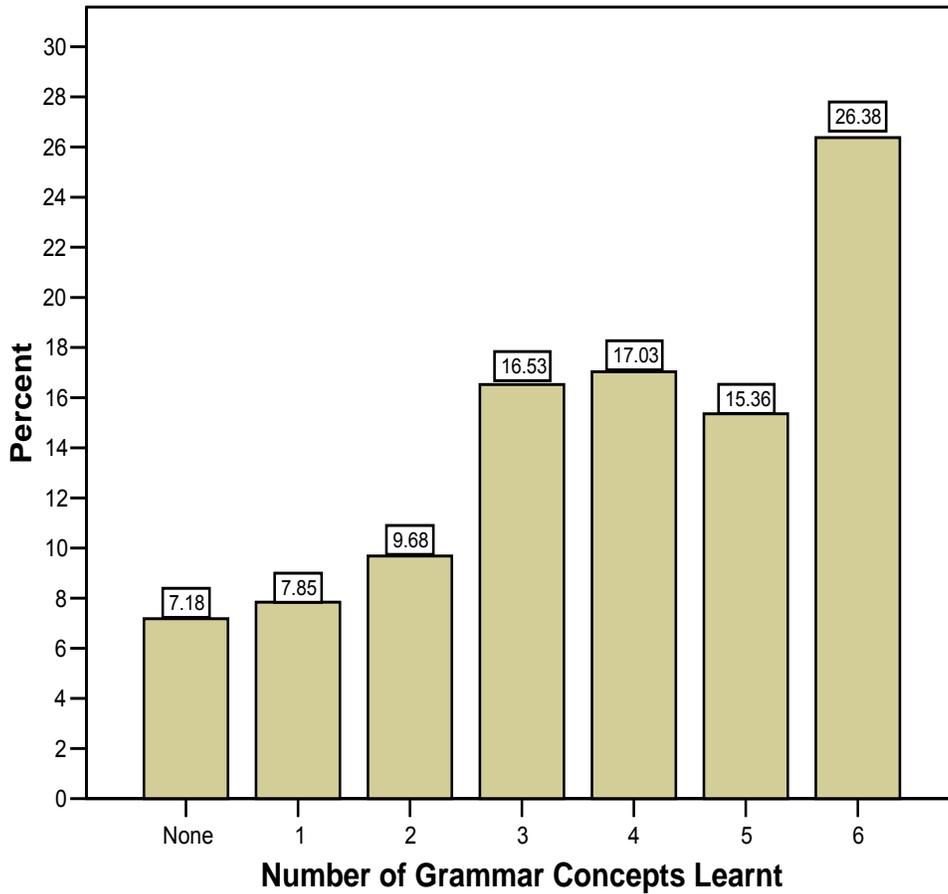


Figure 4.14, reveals that of a total of 599 students who responded to the question of how many of the grammar concepts they had managed to grasp, about 7.2% of them reported that they had grasped none. 17.5% of the students understood one to two of the concepts taught while the majority, about 75.3% managed to grasp three or more of the grammar concepts. Of the 75.3%, about 26.4% had comprehended all six concepts taught. The LDEP appears to have succeeded in transmitting the learning of some of the concepts to many of the students. Table 4.34 explores to what extent the students had consciously tried to transfer the concepts they had learnt into their own pieces of writing.

Table 4.34: Students' Self-Report on the Outcomes of the Programme

Questions	Yes (%)	No (%)	n
▪ Do you try to create appropriate sentence types to suit the content of your writing after the Programme?	64.2	35.8	600
▪ After the Programme, where appropriate, do you include more adjectives, adverbs or other descriptive words when you write?	76.8	23.2	604
▪ Are you more aware of the appropriate tenses to use in your writing after the Programme?	79.8	20.2	603
▪ Do you find that you are writing better now than you were at the end of Term 2?	55.2	44.8	601
▪ Does the literature text help you to understand the elements of a story better?	66.1	33.9	602
▪ Were the lessons useful in helping you to analyse the stories?	71.4	28.6	597

Around 64% of the students made an effort to vary their sentence structures to suit the writing. Over 75% of the students reported that they were adding more elaborative details and were more careful with their tenses after the LDEP (Table 4.34). With regard to literary analysis, about 70% of the students felt the lessons helped them in the analysis and 66% of them comprehended the elements of a story better through the use of the literature text. Though only 55.2% perceived that they had become better writers, the higher percentages in previous responses, showing that students were conscious of some of the conventions of writing and were making an effort to apply them, suggest that more may have improved in their writing potential than they themselves think.

The explicit teaching of grammar concepts appears to have resulted in learning taking place. Not only did their writing skills improve as evidenced by their mean improved scores in the post writing test (c.f. p.110), but most of the students also perceived that they had grasped at least one of the concepts. In addition, many of them also viewed themselves as improving in one or more areas in their writing.

In the survey, the students were also asked about their reading preferences, and their liking for the English and Literature lessons before the Programme and for the Programme itself.

Table 4.35: Percentages of the Approval Responses of the Students

Questions	Yes (%)	No (%)	n
▪ Do you enjoy reading?	80.7	19.3	602
▪ Did you enjoy reading Literature Text A?	62.5	37.5	603
▪ Did you enjoy reading Literature Text B?	54.7	45.3	382*
▪ Before the start of the Literature-Driven English Programme (LDEP), did you enjoy the English lessons?	62.9	37.1	598
▪ Before the start of the LDEP, did you enjoy the Literature lessons?	62.3	37.7	509 [#]
▪ Did you enjoy the LDEP?	53.6	46.4	593

*In Schools 1 and 4, the students were required to read two short stories, Text A and Text B.

[#]The students in School 4 were taught Literature for the first time in Term 3 when the LDEP was started.

With reference to Table 4.35, it may be established that a high percentage of students (around 81%) enjoyed reading, though the question was broad enough for the students to include different types of reading materials, including non-fiction and self-help books. Before the LDEP, 62% of the students enjoyed the Literature lessons, and a similar percentage of students enjoyed reading at least one of the literature texts that was used during the Programme. A similar percentage of students (about 63%) also enjoyed the English lessons before the LDEP. About 54% of the students reported enjoying the LDEP.

The students were also asked open-ended questions about what kinds of books they preferred reading and their reactions to the LDEP. From these responses, categories were created. Through their responses, it was hoped that a better understanding of their reading preferences as well as their reactions to the programme may be gained than from a general “yes” or “no” response to questions found in Table 4.35. The categories and the percentages of male and female responses to each category are presented in Table 4.36. For example, in School 1, 69% of the males and 62% of the females surveyed reported that they preferred reading mystery, adventure or fantasy books (Table 4.36). The percentages of the responses from either the males or females total more than 100% because some students provided multiple responses to the question requesting them to state their reading preferences. The overall number of responses provided by males and females are also offered in Table 4.36 and

from these responses it may be seen that the responses are more than the number of respondents. For example, in School 1, the 86 males offered 109 responses while the 128 females offered 181 responses.

Table 4.36: Student Response to Reading Preferences

School 1 (Respondents: 128 females; 86 males)			School 2 (Respondents: 90 males)	
Reading Preferences	Male	Female	Reading Preferences	Male
Mystery, Adventure, Fantasy	69%	62%	Mystery, Adventure, Fantasy	86%
Science Fiction	1%	5%	Science Fiction	24%
Supernatural/Horror	22%	16%	Supernatural/Horror	8%
Teenage Fiction	0%	11%	Teenage Fiction	0%
Romance	3%	17%	Romance	0%
Comics	6%	4%	Comics	6%
Social Issues (Novels)	10%	13%	Social Issues (Novels)	0%
War	1%	0%	War	7%
Non-Fiction	14%	13%	Non-Fiction	13%
<i>No. of responses</i>	<i>109</i>	<i>181</i>	<i>No. of responses</i>	<i>129</i>

School 4 (Respondents: 31 females; 34 males)			School 3 (Respondents: 79 females)	
Reading Preferences	Male	Female	Reading Preferences	Female
Mystery, Adventure, Fantasy	85%	97%	Mystery, Adventure, Fantasy	76%
Science Fiction	3%	0%	Science Fiction	0%
Supernatural/Horror	18%	10%	Supernatural/Horror	20%
Teenage Fiction	6%	19%	Teenage Fiction	14%
Romance	3%	16%	Romance	22%
Comics	9%	3%	Comics	5%
Social Issues (Novels)	0%	6%	Social Issues (Novels)	3%
War	3%	0%	War	24%
Non-Fiction	0%	3%	Non-Fiction	0%
<i>No. of responses</i>	<i>43</i>	<i>48</i>	<i>No. of responses</i>	<i>129</i>

According to Table 4.36, the majority of the students, irrespective of gender, reported enjoying books on mysteries, adventures and fantasy. There were marked differences in three categories of reading preferences. Over a fourth of the girls in the two co-educational and in the girls' schools enjoyed reading books on Teenage Fiction and Romance (Table 4.36). While below 10% of boys in the co-educational schools mentioned liking these kinds of books, none in the boys' school reported liking reading Teenage Fiction or Romance (Table 4.36). On the other hand, close to a fourth of the boys in the boys' school enjoyed reading science fiction and none of the girls in either the girls' school or co-educational

School 4 mentioned preferring science fiction. In another co-educational school, School 1, more girls than boys mentioned liking science fiction (Table 4.36).

Since a literature book was used to teach language skills, it is important to evaluate whether the choice of literature text had an impact on the learning of the students and their enjoyment of the programme. Table 4.37 displays the number of students who reported liking the story or stories they had to read under literature. These literature texts were prescribed by the schools.

Table 4.37: Positive Reaction to Text studied under Literature

School 1

Description of Text	Male %	Female %
Text A – Story of a man romancing and fooling a much older woman	49.7	56.2
Text B – Story of a young woman who married a much older man for money	61.9	60.8

School 4

Description of Text	Male %	Female %
Text A – Story of a young boy befriending an underprivileged boy and later suspecting him of cheating	52.9	78.1
Text B – Story of a young boy being visited by his dying grandmother	25.1	34.4

School 2

Description of Text	Male %
Text – Science Fiction: Collection of short stories about man’s relationship with the robots that he had created	74

School 3

Description of Text	Female %
Text – Adapted Shakespearean play about love, racism and friendship	75.2

In the co-educational schools, there were mixed receptions to the chosen short stories. In School 1, the percentages of the boys and girls who liked the stories were little different (Table 4.37) even though only 3% of the boys reported liking romance stories (c.f. Table 4.36, p.162). After the LDEP, the boys managed to achieve an improvement score ES of 0.4 ($t = 3.709$, $df = 85$) in the writing assessment and the male Normal (Academic) students in School 1 improved by an ES of 1.15 in the reading comprehension assessment (c.f. p.156). (The results of the students in the Express stream were not analysed, c.f. pp.142-145.)

In School 4, however, the first story chosen appealed to more girls than boys and the second story, in general, was not well-liked by either the girls or boys (Table 4.37). In spite of the fact that in School 4 few boys liked the stories, in the writing test they improved with an ES of 0.59 ($t = 3.574$, $df = 36$). In the reading comprehension, the ES of the improvement score was 0.92 (c.f. p.149). In the single-sex schools, about three-fourths of the students liked the assigned texts (Table 4.37). The results of the reading comprehension assessment of the students from School 2 were not analysed (c.f. pp.142-145) but their writing had improved, with this experimental group of students achieving an improvement score ES of 0.56 (c.f. Table 4.14, p.130). The ES of the improvement score in writing achieved by the girls in School 3 was marginally lower at 0.18 (Table 4.14, p.130) but the improvement score ES attained in the reading comprehension was a moderately large 0.63 (c.f. Table 4.30, p.154).

In Table 4.38, the reactions of the students to the LDEP are displayed. The categories found in Table 4.38 were created from the responses that the students gave to open-ended questions about what they did or did not like about the LDEP. There were students who provided multiple responses to each question, giving more than one positive or negative reaction. For each reaction, sub-groups of male and female responses in each school were created. Within each male or female sub-group, each reaction was calculated as a percentage of the total male or total female respondents from that school. For example, In School 1 (Table 4.38) only 0.8% of boys liked the novelty of the LDEP whereas 1.2% of the girls in School1 liked the novel aspects of the program. The aim was to discover the percentage of males and females reporting each positive or negative reaction. The percentages of positive reactions in each school for each sub-group total more than 100% because of the multiple responses provided by some of the students, as indicated by the overall positive responses of the males and females. For example, in School 1, the 126 males offered 169 positive responses. The percentages of negative reactions in three schools for each sub-group total less than 100% because some students did not report any negative

reactions to the programme, as indicated by the overall negative responses of the males and females. For example, in School 1, 88 negative responses were received from the 126 males and 142 negative responses from the 169 female respondents.

Table 4.38: Student Response to Reactions to the LDEP

School 1 (Respondents: 169 females; 126 males)

Positive Reactions	Male	Female
Liked the novelty	0.8%	1.2%
Interesting or fun activities	40.5%	39.6%
Useful or informative	50.8%	58.6%
Liked Group work	13.4%	11.8%
Improved literary analysis	12.7%	7.7%
Liked integration	15.9%	9.5%
<i>No. of responses</i>	<i>169</i>	<i>217</i>
Negative Reactions		
Boring or uninteresting	30.2%	30.2%
Boring at times	22.2%	27.8%
Difficult or confusing	4.8%	6.5%
Disliked the integration	11.0%	13.0%
Disliked Group work	0%	0.6%
Concepts taught too basic	1.6%	5.9%
<i>No. of responses</i>	<i>88</i>	<i>142</i>

School 2 (Respondents: 106 males)

Positive Reactions	Male
Liked the novelty	8.5%
Interesting or fun activities	28.3%
Useful or informative	56.6%
Liked Group work	5.7%
Improved literary analysis	17.0%
Liked integration	20.8%
<i>No. of responses</i>	<i>145</i>
Negative Reactions	
Boring or uninteresting	30.2%
Boring at times	22.6%
Difficult or confusing	4.8%
Disliked the integration	2.8%
Disliked Group work	0%
Concepts taught too basic	11.3%
<i>No. of responses</i>	<i>76</i>

Sch 4 (Respondents: 36 females; 35 males)

Positive Reactions	Male	Female
Liked the novelty	0%	0%
Interesting or fun activities	31.4%	30.6%
Useful or informative	42.9%	61.1%
Liked Group work	20.0%	36.1%
Improved literary analysis	5.7%	8.3%
Liked integration	5.7%	5.6%
<i>No. of responses</i>	<i>37</i>	<i>51</i>
Negative Reactions		
Boring or uninteresting	57.1%	44.4%
Boring at times	8.6%	30.6%
Difficult or confusing	14.3%	25.0%
Disliked the integration	11.4%	2.8%
Disliked Group work	2.9%	8.3%
Concepts taught too basic	8.6%	8.3%
<i>No. of responses</i>	<i>36</i>	<i>43</i>

School 3 (Respondents: 87 females)

Positive Reactions	Female
Liked the novelty	6.9%
Interesting or fun activities	37.9%
Useful or informative	57.5%
Liked Group work	4.6%
Improved literary analysis	6.9%
Liked integration	25.3%
<i>No. of responses</i>	<i>121</i>
Negative Reactions	
Boring or uninteresting	46.0%
Boring at times	8.0%
Difficult or confusing	4.6%
Disliked the integration	24.1%
Disliked Group work	0%
Concepts taught too basic	1.1%
<i>No. of responses</i>	<i>73</i>

An examination of Table 4.38 reveals that in all schools the positive responses outnumber the negative responses. In all schools, except for the boys in School 4, more than half the students found the programme useful or informative (Table 4.38). Between 30% and 40% of the respondents in each school mentioned the programme being interesting or enjoying the activities in the programme. Though only about 30% of the respondents in School 1 and 2 found the programme boring or uninteresting, about 51% and 46% of the respondents in Schools 4 and 3 respectively reported finding the programme boring (Table 4.38). The reasons for this will be further explored in the next chapter on Qualitative Findings (pp.195-197 & pp.215-216).

In the next section, the overall student response from the four schools to the LDEP is presented. The averages of the percentages of each positive and negative reaction of the male and female students (c.f. Table 4.38, p.165) from these schools as a whole were calculated. The findings are reported in Table 4.39.

Table 4.39: Overall Student Response to Reactions to the LDEP

Respondents (292 females; 267 males)

Positive Reactions	Male	Female	Total
Liked the novelty	3.1%	2.7%	2.9%
Interesting or fun activities	33.4%	36.0%	34.7%
Useful or informative	50.1%	59.1%	54.6%
Liked Group work	13.0%	17.5%	15.3%
Improved literary analysis	11.8%	7.6%	9.7%
Liked integration	14.1%	13.5%	13.8%
<i>No. of responses</i>	351	389	740
Negative Reactions	Male	Female	Total
Boring or uninteresting	39.2%	40.2%	39.7%
Boring at times	17.8%	22.1%	20.0%
Difficult or confusing	8.0%	12.0%	10.0%
Disliked the integration	8.4%	13.3%	10.9%
Disliked Group work	1.0%	3.0%	2.0%
Concepts taught too basic	7.2%	5.1%	6.2%
<i>No. of responses</i>	200	258	458

Table 4.39 reveals that close to 35% of the students reported finding the programme interesting and about 55% stated that the programme was

informative. Though 15.3% of the students mentioned liking the group work, only a small 2% reported disliking the group work (Table 4.39). While about 14% of the students reported liking the integrated English and Literature programme, about 11% mentioned disliking the integration. Around 40% of the students referred to the programme as boring and 10% of them reported that the lessons were difficult (Table 4.39). The number of positive responses (740) was higher than the number of respondents (559), thus indicating that on the average there were approximately 1.3 positive responses from each respondent (Table 4.39). The number of responses from the females was higher than the responses from the males but this could be due to there being more females who completed the surveys. When the percentages of the responses in relation to the male and female populations were calculated, it was found that positive responses for the males and females were similar with 131% positive male responses in comparison to 133% positive female responses. Negative responses, on the other hand, were fewer than the number of respondents (Table 4.39) with there being an average of about 0.82 negative responses from the respondents. Negative responses were calculated at about 75% for the males and 88% for the females. In all there were about 150 and 130 more positive than negative responses from the males and females respectively (Table 4.39). Overall 62% of the responses were positive while 38% were negative responses. Many of the students surveyed perceived the LDEP as either being beneficial or enjoyable in some ways.

The fewer negative responses from the males than the females might indicate that more boys were receptive to the LDEP, hence leading to general improvements in the writing skills and reading comprehension skills of the boys irrespective of the PSLE English grades (c.f. pp.125-126 & 150).

In summary, the LDEP appeared to have helped many of the students to improve in their writing (c.f. p.112) and to be conscious that they were learning skills that they would find useful in improving their writing.

Conclusion

In the three schools where there were experimental and control groups, while the experimental group was taught under the LDEP, the control group was taught the usual curricula of the respective schools. The usual curricula included the teaching of similar topics such as narrative elements and narrative techniques. In two of the schools, text-based grammar that made use of extracts to teach grammar concepts was taught. In the third school, which had a literature-based English curriculum, grammar concepts were taught as and when written assignments revealed a need for the reinforcement of certain concepts.

The findings indicate that the LDEP had a bearing on the improvement of student writing skills, irrespective of the variables. The 482 students in the experimental group achieved positive mean improvement scores in writing skills, irrespective of gender, language fluency, class or school cultures and teachers. On the other hand, the 290 students in the control group did not attain a positive mean improvement score in the writing skills assessment. In the writing skills assessment, the experimental group improved over the control group in the following areas:

1. When individual scores of the students were assessed, more than 60% of the students in the experimental group posted positive improvement scores. However, only about a third of the students in the control group achieved positive improvement scores (c.f. p.109).
2. The mean improvement score of the students in the experimental group was positive with an ES of 0.44. On the other hand, the control group achieved a mean improvement score that was slightly below zero, thereby indicating that there was no change in the performance of the group in the post-test of the writing skills assessment (c.f. p.112).

3. In all the five measured areas of outcome by which the writing skills of the students were assessed, the experimental group displayed greater improvement than the control group (c.f. pp.114-116).

Among the sub-groups within the experimental group, generally there was improvement in the performance of the students.

1. Students with different PSLE English grades achieved positive mean improvement scores, attaining improvement Effect Sizes of between 0.39 and 0.57 (c.f. Table 4.8, p.122).
2. Generally, boys and girls improved with their mean improvement scores attaining Effect Sizes of 0.50 and 0.39 respectively (c.f. Table 4.10, p.124).
3. Boys with different PSLE English grades improved with Effect Sizes of the improvement scores ranging from 0.43 to 1.40. Except for one sub-group of girls with PSLE English grade C, the girls with PSLE English grades A* to B achieved Effect Sizes of the mean improvement scores of between 0.37 and 0.61 (c.f. Table 4.11, p.125 & Table 4.12, pp.126).
4. All the schools attained positive improvement score Effect Sizes although School 3 achieved the lowest ES of 0.18. The other three schools attained comparable Effects Sizes of between 0.48 and 0.56 (c.f. Table 4.14, p.130).
5. Nine of the 14 experimental classes achieved improvement score ES of above 0.33. Of the remaining five classes, two of them achieved Effect Sizes of 0.29 and 0.30, which are close to the ES of 0.33 taken as an adequate measure of effectiveness in this research. Only three classes achieved lower, though positive, Effect Sizes (c.f. Table 4.16, p.134).

6. In School 4, where the teachers taught both experimental and control classes, the control classes achieved negative mean difference scores between the post-test and the pre-test in the writing skills assessment. The experimental classes, on the other hand, achieved moderate to large Effect Sizes of 0.41 and 0.68 (c.f. Table 4.17, p.135).

In the writing skills assessment, in almost all the sub-groups and as part of a larger group of experimental classes, the students from the experimental classes improved whereas the control group did not improve.

The reading comprehension formed a minor component of the LDEP while the major component comprised of the writing skills. Due to a lack of time, a greater emphasis was placed on writing skills and these skills were explicitly taught. On the other hand, reading comprehension skills were not explicitly taught, rather it was expected that students would improve in these skills from the literary analysis of the literature text and the teaching of the writing skills.

The LDEP appears to have also enhanced the reading comprehension skills of 40 Normal (Academic) students in School 1 and the 176 students with Grades A, B and C in Schools 3 and 4. In comparison to the control group of 189 students in Schools 3 and 4:

1. Overall, the experimental group of students achieved a larger ES than the control group of students (c.f. Table 4.25, p.146). Among students with different PSLE English grades, the experimental group also improved more than the control group (c.f. Table 4.26, p.148).
2. When making comparisons by gender, boys in the experimental group achieved a larger ES than the boys in the control group. Similarly, the girls in the experimental group also attained a larger ES than the control group of girls (c.f. Table 4.27, p.149).

3. Boys with Grades A and B in the experimental group achieved much larger Effect Sizes of improvement scores than the boys in the control group (c.f. Table 4.28, p.150). Among the girls, the girls with Grade A in the experimental group achieved a larger ES than the girls with Grade A in the control group. The Effect Sizes of the improvement scores of the girls in the control and experimental groups with Grade B were similar (c.f. Table 4.29, p.151).
4. The experimental group of students in Schools 3 and 4 attained similar Effect Sizes of improvement scores of 0.63 and 0.65 respectively. The Effect Sizes of the control group of students in Schools 3 and 4 were much lower at 0.39 and 0.46 respectively (c.f. Table 4.30, p.154 & Table 4.31, p.155).
5. Among the experimental classes, all except one class achieved moderate to large Effect Sizes of between 0.51 and 1.14. One class, Class 18 in School 3 achieved a much lower ES of 0.23 (c.f. Table 4.32, p.155). However, the ES was still statistically significant at the 5% level.
6. In the Normal (Academic) class, the students achieved a large overall ES of 1.10 and moderate to large Effect Sizes in the sub-groups based on gender and PSLE English grades (c.f. Table 4.33, p.156).

Generally, the experimental group of students achieved a larger improvement score ES than the control group of students in the reading comprehension assessment. The LDEP appears to have been more effective in enhancing comprehension skills through linking literary analysis with the teaching of writing than the conventional mode of completing distinct comprehension exercises.

The results of the survey of the students of the experimental classes support the findings of the pre-tests and post-tests that the LDEP had an impact on the

students. From the survey findings attained from the responses of the 604 students, many of the students perceived that learning had taken place during the programme. Over three-fourths of the students reported that they grasped three or more grammar concepts taught and many of the students also attempted one or more of the writing techniques taught (c.f. pp.159-160). There were also more positive than negative responses to the LDEP from the students surveyed (c.f. p.167).

The quantitative evidence to support the positive effects of the LDEP is clear. The LDEP had led to substantial improvement scores in the writing skills of the students. The LDEP had also led to substantial improvement in the reading comprehension skills assessed even within the limited time frame of the curriculum intervention. Furthermore, the majority of student comments from the exit survey were positive. The quantitative evidence has been argued in terms of both statistical significance and effect size. Both these measures justify the claim that the LDEP led to measurable, substantial increases in skills despite a variety of uncontrolled variables of initial performance (PSLE English grades), gender, school or class and academic grouping.

In the next chapter, the feedback from teachers and students about the LDEP expressed through interviews and log book entries are presented and discussed in detail. In addition, the observation and field notes are also used to inform the interview findings found in the next chapter. It is also hoped that through the voices of the people most closely connected to the programme, namely the students and the teachers, it would be possible to list the advantages and disadvantages of the programme. Finally, suggestions on how the LDEP could be further improved are also offered.