THE UNIVERSITY OF HULL

PEER EVALUATION: CAN IT IMPROVE THE WRITING ABILITY OF GRADE 12 STUDENTS IN THE UNITED STATES VIRGIN ISLANDS?

being a dissertation submitted for the degree

of Doctor of Philosophy

in the University of Hull

by

TREVOR BELFIELD PARRIS

(B.A. General with honours, University of the West Indies)

(M.Ed. University of Toronto)

October, 1989

ACKNOWLEDGEMENTS

The researcher thanks all those who assisted in this research project, especially the following individuals. First, sincere thanks and appreciation are extended to Dr. Robert Protherough for his expert guidance and concern throughout the period of this research project. Feedback was always immediate, and Dr. Protherough helped to close the distance between Hull and the Virgin Islands. Appreciation also goes to Mr Colin Brock, whose role as facilitator helped to make the project possible. Thanks also go to Mr Howard Thomas, and Dr. W.J. Wilkinson for their comments on the statistical analysis. Personnel in the Department of Education, and the high schools were particularly helpful. It was particularly gratifying to work with those teachers of English who participated in the study. Finally, thanks to Miss J. Martin for assisting in the typing of the script.

CONTENTS

		Page
Acknowledgements		i
Table of Contents		ii
List of Tables		viii
List of Figures		x
CHAPTER ONE:	STATEMENT OF THE PROBLEM	1
	Justification	2
CHAPTER TWO:	THE VIRGIN ISLANDS BACKGROUND	6
	Development	7
	Pre-1960 period	7
	Post-1917 period	13
	The Board of Education	14
	The Teachers' Federation	16
	Department of Education	17
	Schools	17
CHAPTER THREE:	REVIEW OF THE LITERATURE	19
	The Rhetorical Context	19
	Distance	20
	The Writer Model	22
	The Product Model	23
	School Writing	24
	Writer-based Prose	27
	Audience	29
	Definition	29
	When to Consider Audience	34
	Discourse Communities	36
	Summary	37
	Group Work	39
	Definitions	39
	The Importance of Group Work	41
	Arrangements and Procedures	46

Empirical Research 50
Talking to Learn 50
Revision Studies 52
Evaluation Studies 58
Summary 62
Language Development 63
Personal Growth 65
The Contribution of Moffett 68
The Contribution of Britton 68
The Wilkinson Scales 70
Cognitive Model 71
Affective Model 72
Moral Model 73
Stylistic Model 74
Summary 74

CHAPTER FOUR: THE PILOT STUDY 76
Abstract 76
Subjects 77
Procedure 78
Lesson Plans: Experimental group 79
Activities: Control group 81
Controls 82
Randomisation 82
History 82
Testing 82
Morality 84
Other Effects 84
Data Analysis 85
Results 85
Grammar 89
Validating the Rating Scale 90
Discussion 91
Attitudinal Data 92

Writing Letters 94
Receiving Letters 95
Rating 95
Working in Groups 97
Using Booklets 98
Writing Assignments 98
Writing Improvement 98
Summary 99
Conclusions and Implications 99
Changes: Main Study versus Pilot Study 101
Larger Issues 102
Other Changes 103
Assignments 103
Materials 105
Scales 106
Measures: pre-tests 106
Grading 107

CHAPTER FIVE: RESEARCH DESIGN AND PROCEDURES 108
The Sample 108
Experimental Group 109
Control Group 109
Personnel 110
Teacher A 110
Teacher B 110
Teacher C 110
Teacher D 111
The Variables 111
Internal Validity 112
History 112
Maturation 114
Testing 114
Instrumentation 115
Other Variables 116
External Validity 116
Population 116
Description of Instruments 120
STEP 120

	Composition	121
	Writing Apprehension	122
	Multiple Measures	122
	Instructional Objectives	124
	Experimental Group	124
	Language Activities	126
	Control Group	126
CHAPTER SIX:	CONSTRUCTION OF THE SCALE	129
	Scale Data	132
	Terminology	132
	The Diederich Scale	135
	Other Analytic Scales	141
	The Case of England	145
	Analytic Scales	146
	Essay Scales	146
	Impression Marking	148
	The Assessment of Performance Unit (APU)	151
	The Scale	154
	Format	154
	Scale	156
	Content	158
	Organisation	158
	Structure	159
	Wording	159
CHAPTER SEVEN:	LESSONS TO TEACH THE COMPONENTS OF THE SCALE	160
	Teaching the Scale	161
	Using the Scale	162
	Objective	164
	Level of Intensity	165
	Instructional Roles	166
	Use of Space	167
	Materials	167
	Teachers' Lesson Plan	170
	Teachers Comments	173
	Ratings	174
	Editing	175

Organisation 178
Materials 179
Ratings 181
Structure 185
Ratings 187
Wording 189
Ratings 190

CHAPTER EIGHT: ANALYSIS OF THE DATA 193
Measures 195
Pre-test 195
Post-test 195
Pre-test Data 196
Effect of the Programme 200
Total Score 200
Content 204
Organisation 206
Structure 208
Wording 210
Summary 212
Children as Raters 213
Inter-reliability 213
Estimate of Reliability 214
The Whole Scale 214
Content Component 215
Structure 217
Wording 218
Summary 219
Indirect Assessment 219
English Expression 219
Mechanics 221
Free Response to the Programme 223
General Comments 224
Attitude 224
Writing Improvement 226
Group Work 227
Benefits 227
Disadvantages 228
Rating own papers 230
Comparisons 231

Summary 232

CHAPTER NINE: SUMMARY AND CONCLUSIONS 234

Results and Limitations 235

Effect of the programme 235

Sub-scales 237

Pupils as Raters 238

Student Attitudes to Writing 238

Implications 239

Teaching 239

Research 242

APPENDIX A PILOT STUDY DATA 246

APPENDIX B MAIN STUDY DATA 251

APPENDIX C MEASURES AND SCALES 258

APPENDIX D ILLUSTRATIVE MATERIALS 280

BIBLIOGRAPHY 294

LIST OF TABLES

		Page
TABLE 1:	Means and Standard Deviations of Pre-test Scores	85
TABLE 2:	Effect of Programme on Quality of Post-test Composition	86
TABLE 3:	Effect of Pre-test Grammar Scores on Post-test Composition Scores	87
TABLE 4:	Effect of the Pre-test Mechanics Scores on Post-test Composition Scores	88
TABLE 5:	Effect of the Programme on Post-test Grammar Scores	89
TABLE 6:	Inter-Rater Reliability Co-Efficients for 3 Raters	91
TABLE 7:	Summary of Ratings for Student Questionnaire	93
TABLE 8:	Ratings of "Bedlam" by Class	174
TABLE 9:	Ratings of "Centre Cut" by Group	181
TABLE 10:	Ratings of "Roots Wedding" by Group	183
TABLE 11:	Group Ratings of Composition for 3 Qualities	188
TABLE 12:	Individual Ratings of Writing Sample	191
TABLE 13:	Pre-test Statistics for the Experimental and Control Groups	197
TABLE 14:	Pre-test Statistics for Composition Scores and Sub-Scores	198
TABLE 15:	Effect of the Programme on the Overall Quality of Composition based on Post-test Adjusted Means[a]	202
TABLE 16:	Analysis of Covariance of Post-test Total Composition Scores for Two Groups	203
TABLE 17:	Effect of the Programme on the Overall Quality of Content based on Post-tested Adjusted Means[a]	204
TABLE 18:	Analysis of Covariance of Post-test Composition (Content) Scores for Two Groups	205
TABLE 19:	Effect of the Programme on the Overall Quality of Organisation based on Post-test Adjusted Means[a]	207
TABLE 20:	Analysis of Covariance of Post-test Composition (Organisation) Scores for Two Groups	208
TABLE 21:	Effect of the Programme on the Overall Quality of Structure based on Post-test Adjusted Means[a]	209

Page

TABLE 22: Effect of the Programme on the Overall Quality of Wording based on Post-test Adjusted Means[a] 211

TABLE 23: Analysis of Covariance of Post-test Composition (Wording) Scores for Two Groups 212

TABLE 24: Inter-Rater Reliability Coefficient for the Whole Scale 214

TABLE 25: Inter-Rater Reliability Coefficient for the Content Component of the Scale 215

TABLE 26: Inter-Rater Reliability Coefficient for Organisation Component of the Scale 216

TABLE 27: Inter-Rater Reliability for the Structure Component of the Scale 217

TABLE 28: Inter-Rater Reliability Data for the Wording Component of the Scale 218

TABLE 29: Effect of the Programme on the Overall Quality of English Expression based on Post-tested Adjusted Means[a] 220

TABLE 30: Effect of the Programme on the Overall Quality of Mechanics of Writing based on Post-test Adjusted Means[a] 222

TABLE 31: Source Table for Table 2 246

TABLE 32: Source Table for Table 3 246

TABLE 33: Source Table for Table 4 247

TABLE 34: Source Table for Table 5 247

TABLE 35: Source Table for Table 6 248

TABLE 36: Source Table for Table 6 248

TABLE 37: Source Table for Table 6 249

TABLE 38: Source Table for Table 6 249

TABLE 39: Source Table for Table 6 250

TABLE 40: Raw Score Data for Main Study 251

TABLE 41: Pre-test and Post-test Means and Standard Deviations 253

TABLE 42: Correlation Matrix based on Population Data 254

TABLE 43: Raw Scores for Student Ratings: Total Score 255

TABLE 44: Raw Scores for Student Ratings: Content 255

TABLE 45: Raw Scores for Student Ratings: Organisation 256

TABLE 46: Raw Scores for Student Ratings: Structure 256

TABLE 47: Raw Scores for Student Ratings: Wording 257

LIST OF FIGURES

		Page
FIGURE 1:	Economic indicators	9
FIGURE 2:	List of objectives: Content	124
FIGURE 3:	List of objectives: Organisation	125
FIGURE 4:	List of objectives: Structure	125
FIGURE 5:	List of objectives: Wording	126
FIGURE 6:	Language activities: Content and Organisation	127
FIGURE 7:	Language activities: Structure and wording	128
FIGURE 8:	Summary sheet: Content	164
FIGURE 9:	Anchor paper: High (5)	168
FIGURE 10:	Anchor paper: Hugh (4)	169
FIGURE 11:	Teachers' Lesson Plan	170
FIGURE 12:	An edited script	175
FIGURE 13:	Student Sample Rated for Two Qualities	179
FIGURE 14:	Student Sample Rated for Two Qualities	180
FIGURE 15:	Student Composition Rated for Three Qualities	187
FIGURE 16:	Student Sample Rated for All Qualities	190
FIGURE 17:	Questionnaire	258
FIGURE 18:	Writing Apprehension Measure	260
FIGURE 19:	Rating Scale: Content (high)	262
FIGURE 20:	Rating Scale: Content (middle)	263
FIGURE 21:	Rating Scale: Content (low)	264
FIGURE 22:	Rating Scale: Organisation (high)	265
FIGURE 23:	Rating Scale: Organisation (middle)	266
FIGURE 24:	Rating Scale: Organisation (low)	267
FIGURE 25:	Rating Scale: Structure (high	268
FIGURE 26:	Rating Scale: Structure (middle)	269
FIGURE 27:	Rating Scale: Structure (low)	270
FIGURE 28:	Rating Scale: Wording (high)	271
FIGURE 29:	Rating Scale: Wording (middle)	272
FIGURE 30:	Rating Scale: Wording (low)	273
FIGURE 31:	Composition topic: "Carnival"	274
FIGURE 32:	Composition topic: "Christmas"	275

		Page
FIGURE 33:	Rating sheets: Content	276
FIGURE 34:	Rating sheets: Content and Organisation	277
FIGURE 35:	Rating sheets: Content, Organisation and Structure	278
FIGURE 36:	Rating sheets: Content, Organisation, Structure and Wording	279
FIGURE 37:	Summary sheet: Organisation	280
FIGURE 38:	Lesson Plan: Organisation	281
FIGURE 39:	Summary sheet: Structure	282
FIGURE 40:	Student Sample for Rating	283
FIGURE 41:	Teacher's Supplementary Materials	284
FIGURE 42:	Summary sheet: Wording	288
FIGURE 43:	Student Writing Sample	289
FIGURE 44:	Homework Exercise	290
FIGURE 45:	Lesson Plan: Wording	292

CHAPTER ONE

STATEMENT OF THE PROBLEM

This study compared the writing ability of students who were taught to rate their own compositions with that of students whose compositions were rated by their teachers. The pupils were two groups of grade twelve students in two schools in the United States Virgin Islands. The Experimental Group followed a special instructional programme designed by the researcher; the Control group followed the normal programme of the schools. It was hypothesized that:

(a) pupils who worked in groups to rate their own compositions and those of their peers would score significantly higher on measures of writing ability than pupils whose compositions were rated by their teachers.

The research was also to determine:

(b) the inter-rater reliability of the pupils as raters.

(c) the inter-rater reliability of teachers as raters, and

(d) whether student attitude to writing would improve as a result of the programme.

Justification

Collaborative learning is becoming increasingly important as a teaching/learning strategy. This technique, loosely called "group work", is distinguished from the traditional use of space by the nature of the task. In the traditional classroom, the class was divided into small groups to work on specific tasks or sub-sets of the same task, but they did not necessarily engage in collaborative learning.

Weiner (1976) points out that "students put into groups are only students grouped and are not collaborators, unless a task that demands consensual learning unifies group activity" (p.637). Other researchers, for example, Barnes (1982), also stress the nature of the task as the hallmark of collaborative activity. Barnes suggests that group work helps pupils to "take more part in the formation of knowledge" (p.170).

There is increasing evidence that collaborative learning provides an effective facilitating environment for writing. One consideration is the dominance of the teacher in the classroom in general and writing in particular. The work of Britton and his associates have underlined the limiting effects of this dominance. Rosen (1973), commenting on writing evaluation, observes that the teacher is not "simply a one-man audience but also the sole arbiter, appraiser, grader and judge of the performance" (in Barnes, 1982, p.92). Elbow (1981) makes the point that students do not write to teachers, but write for them and that there is usually something fictional about this

transaction between reader and writer in most school writing. Purves (1984) outlines a broad spectrum of roles the writing teacher may adopt when reading student papers so as to reduce her impact over evaluation.

Collaborative learning can also compensate for some of the limitations of the traditional modes of writing. Flower (1979, 1981) shows how such models can often result in writer-based prose with its egocentric focus, constricting narrative principle of organising data, and loaded vocabulary. Teachers of writing are finding traditional paradigms inadequate, for these models test only a limited rhetorical context (Mitchell and Taylor, 1979). Further, manpower considerations now make collaborative learning a necessity, especially for developing countries. Dusel (1958), in his California study, found that 90% of his sample of 430 writing teachers used written correction which took from 8.8 to 21.5 hours a week, mainly out of class time. This overload resulted in fewer class assignments, together with hasty and superficial correction. His recommendation of 4 classes of 25 students each, and 2 periods a week for correction was later adopted by the National Council of Teachers of English (NCTE). However, even that solution when implemented is inadequate. Collaborative learning can help to ease the teacher load, and potentially has become a powerful tool for teaching writing as process and evaluating it as product.

Researchers have investigated the effectiveness of peer groups for responding to writing. Some researchers, for example Russell

(1985), Katstra, Tollefson and Gilbert (1987), and Pierson (1967) have conducted revision studies. Researchers too have found that students working in groups could be taught to grade their own papers and the papers of their peers reliably, and their writing ability could improve as a result of such evaluation (Sager, 1975; Ford, 1975; Lagana, 1972).

One of the issues that has emerged with respect to the use of peer groups in assessment is whose responsibility it is to assign grades. In the Pierson study, grades came within the purview of the teacher only. In the Lagana study, a grade for the course was determined by teacher and student working together - grade assignment was a joint effort. But the Sager investigation showed that pupils could be as accurate as adults in rating essays for total and sub-scores. Inter-rater reliability coefficients for students in the Sager study were .99 for total score, and a range of .96 to .98 for subscores; the inter-rater reliability coefficients for adults were .97 for total score and a range of .81 to .96 for subscores. Her findings suggested that students could, on their own, take on more responsibility for grading than teachers might be willing to give. This and other issues raised above are explored in chapter three, the review of the literature.

The present study fits into the set of evaluation studies in which peer groups assigned grades to their own or other student papers. In purpose and intent, it replicates the Sager study but at a higher grade level - grade 12, not grade 6; and with a different

population - The Virgin Islands, not Massachusetts. Classroom teachers and not the researcher taught the classes, and the grades students assigned to papers counted toward the final grade for the course. The study is similar to the revision studies mentioned above in that peer editing was built into the programme, and pupils could revise their papers to improve their grade.

The researcher expected a significant difference in the performance of the two groups, in line with the research of Lagana, Sager, and Ford reported above. He expected pupils engaged in collaborative learning to show a positive attitude towards the method; he expected editing and revision exercises to result in higher mean scores on measures of grammar and mechanics for the Experimental, though not for the Control group.

CHAPTER TWO

THE VIRGIN ISLANDS BACKGROUND

Geologically, the Virgin Islands are a submerged rock artificially divided into two groups - the colony of the British Virgin Islands (BVI), and the territory of the United States of America (USVI). The second group, formerly the Danish West Indies, were sold to the United States in 1917; St.Thomas, St.Croix, and St.John are the chief islands in the group. St.Thomas, in turn, barely 28 square miles, is the hub of the wheel of political and socio-economic activity. The islands' strategic location, only forty miles east of Puerto Rico, and the most northerly of the Lesser Antilles made them historically a halfway house for European and American trade, the heart of West Indian life, and the key to military dominance in international conflict.

The Danish period (1672-1917) is not a main area of concern here, and constraints of space further limit treatment. However, Lewis (1972) lists three "generic factors" shaping the islands at that time: Protestant colonisation by the Danes, the capitalist ethos dominant from the start, plus the economic and social openness of the area which made it a haven for peoples of all nations, and the centre of all activities from trade to piracy. Despite flashes of prosperity, and the early development of a Creole burgher class that dominated social life,in 1917 according to President Hoover, the new territory was "the effective poorhouse of the United States" (Dookhan, 1974, p.270).

The first section of this chapter sketches the economic, political and social development of the islands since 1917. After that comes an outline of the development of education which is followed by a look at the current organisational structure of the system, the Board of Education, and the American Federation of Teachers. Finally, the focus shifts to the two public schools involved in the study.

Development

Pre-1960

The legacy of the Danish period was inadequate social services, an impoverished economy based on a depleted sugar industry, dwindling commercial activity, and a government controlled by vested interests. When the United States bought the islands, they were used as a military base run by the navy. The naval administration effected some improvements in health, water, sewage and other social services, but as Dookan notes, the islands were "incapable of adopting and implementing an adequate revenue system" (p.270). Consequently, a steady flow of federal funds was needed for the supply of even the most basic services. More and more residents were emigrating to seek their fortunes in places like Puerto Rico, Panama and Cuba.

In 1927, natives and residents living in the territory at that time became citizens of the United States; in 1932 the law was extended to include all people born in the Virgin Islands regardless of place of residence. With the question of status settled, and with the coming of civilian rule in 1931, the islands were on the move. The U.S. Department of the Interior oversaw the territory, and a period of rehabilitation followed.

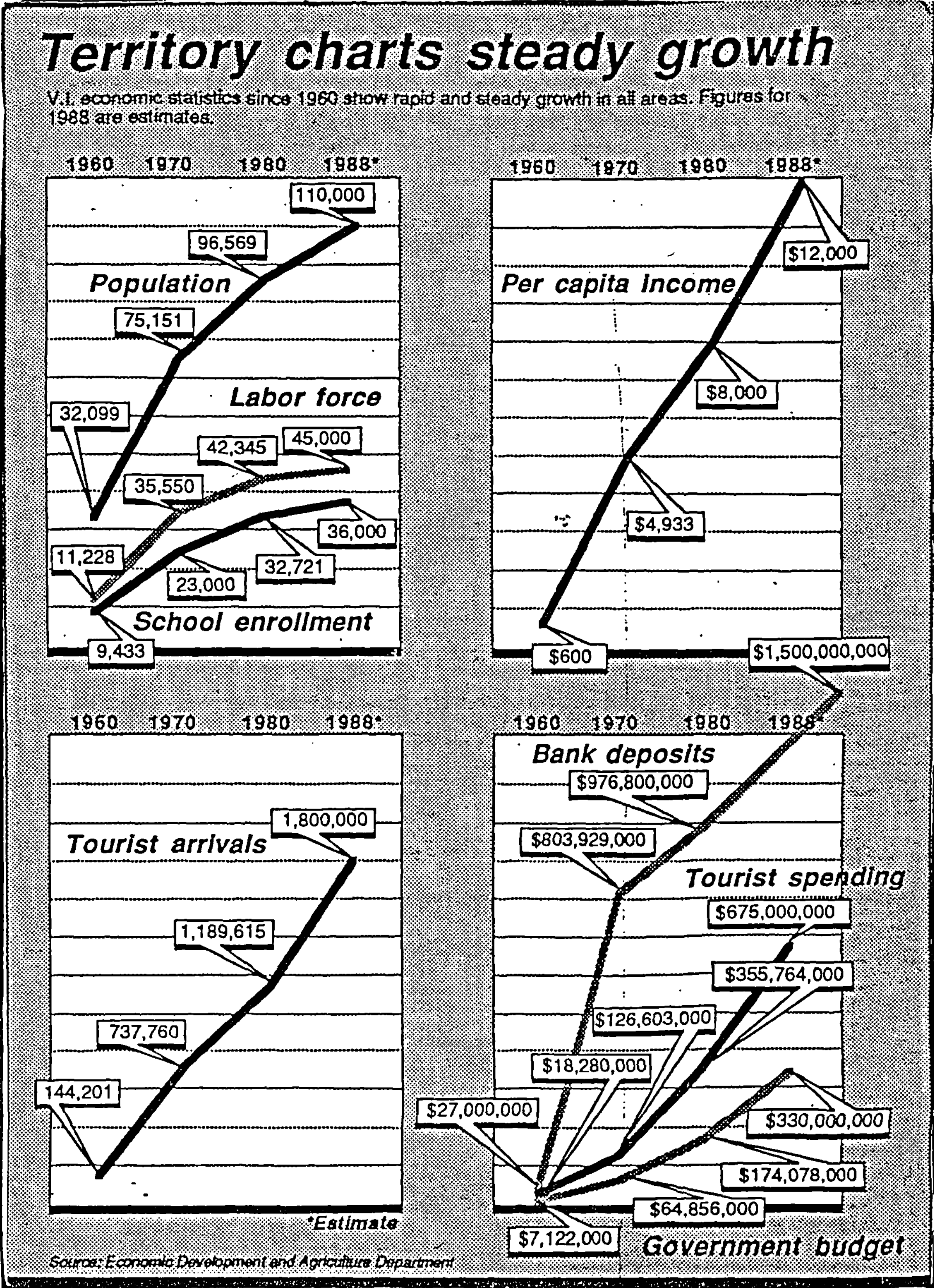
Plantations were bought and subdivided into six acre lots for homesteading, tourism and shipping became significant, and federal programmes, especially in education, made a difference. The Virgin Islands Company set up in 1934 became the engine to fuel this activity. Meanwhile, the population agitated for more political reforms and got them. Women got the vote in 1935, and in 1936 the passage of the Organic Act created a local legislature. The size of the electorate increased substantially; before this period it was a mere 5% of the population.

More recent times have witnessed increased self-government, a vibrant economy based on tourism, and a steady growth in industry, especially with the coming of the Hess oil refinery and the Harvey alumina plant to St.Croix. Further, the Tourist Development Board was founded in 1952, and a revised Organic Act in 1954 increased political and economic benefits. The franchise was further extended, political parties developed, and in 1968 the people elected their own governor for the first time. Figure 1 presents a graphic picture of economic indicators for this period (Rockstein, 1988).

The table graphically presents a profile of development in ten year spans by focussing on certain economic indicators. The data in the first block compare the growth in population, the labour force and school enrollment between 1960 and 1970.

The increase in population was sudden and dramatic - an increase

FIGURE 1: Economic Indicators 1960-1988



of 134% in ten years. Put another way, the 1970 figure more than doubled that of 1960. The labour force more than tripled, and the school roll more than doubled. The islands are still recovering from the shock of the tidal wave of immigrants. In 1960, the labour force was 34% of the population; in 1970 it was 47%. In 1960, the school enrollment was 29% of the population; in 1970 it was 31%, but in real terms this represented 13,567 more children in Virgin Islands schools.

Turnbull (1982) describes the immediate impact on the public schools: seven new schools had to be built in short order, and five more were being planned; double sessions were introduced. Turnbull gives the following statistical data. Public school enrollment was 2,455 at the time of the purchase in 1917; in 1925 it was 3,161; in 1930 it actually dropped to 3,061. In 1950, the population was 4,653; in 1960 it was 7,132; and in 1969 it was 12,655. In 52 years the school population had increased by a mere 10,200 - under 200 a year; then suddenly in the decade of the sixties, there was a crisis in the system.

The reason for this phenomenon was Public Law 91-225 passed by the U.S. Congress in 1969. The children and spouses of bonded workers were admitted to the territory. However, children were not automatically admitted to the schools. Two alien children had to sue the governor and commissioner of education, and win in court before legally resident children were admitted to the public schools.

Turning to the other blocks of data, during the decade under review per capita income in 1970 was over eight times that of 1960. Tourists arrivals were five times the 1960 figure. People earned more money and saved more - bank deposits in 1970 were 30 times higher than in 1960. The budget in 10 years went from just over seven million to well over fifty-seven million.

Growth in the decade from 1970 to 1980 was much slower as the population levelled off. Indeed, the comparative decline was just as drastic. The population increased another 28%, but that was less than half the increase of the previous decade. The labour force went up 19%, negligible compared to the decade before. The school population increased less than 10,000.

The pace of growth after 1980 has been steady, but more normal. The population has risen another 14%; the labour force another 6%; and the school enrollment another 10%. This has given government more breathing space, but the pressure on human services are still severe.

Utility services are no longer adequate: there are frequent power failures, telephone lines are overloaded. Add to this daily traffic congestion on bad roads, and the need to halt development and improve infrastructure is evident. But the waves of immigrants from the U.S. mainland, the steady flow from the former British colonies, the influx from Puerto Rico and Latin America have given the government no respite.

Still signs of prosperity are everywhere: in the 1970s per capita earnings about doubled, and increased another 50% in the 80's. Tourists continue to come and government revenue continues to rise. In the 1970s, the budget more than doubled; in the 80s, it nearly doubled again.

Government, aided by federal programmes, are now reorganising to solve the myriad social problems touched on. Virgin Islanders in November will go to the polls to decide what political relationship they want with the United States. They have seven options, including independence, statehood or status quo.

Education

Danish Period

The Danes were interested in trade, not in settlement, so the introduction of formal education came relatively late. By the turn of the 18th century, children were being sent to Denmark or the British West Indies to school; those that remained were taught at home by "informators" (Turnbull, 1981, p.20). The first school for whites and freedman only opened in 1787, but parochial mission schools for the slaves had been established in 1736 by the Moravians, and 1773 by the Lutherans.

It was not until 1839 - well into the 19th century - that the first comprehensive law was passed regulating education in the islands. This Country School Ordinance instituted a system of

compulsory universal education for all, irrespective of status. Three main types of schools emerged - public or semi-public, parochial or denominational, and private.

The goals were to christianise the slaves through a moral and religious-centred curriculum. The governor and the churches ran the system with the aid of a commissioner, school boards and an inspector of schools. Education was at the elementary level only; secondary education was not established before the American period. Despite the presence of a St.Thomas College (1876 to 1883), and a law to establish secondary education in 1913, there was only one private secondary school in the islands in 1917.

Reading was the most important subject by law in a curriculum that stressed the 3R's: reading, 'religion', and 'rithmetic'. Writing was considered a subversive activity, and parents had to pay for extras like music and needlework. Lewis (1972) describes graphically the state of education at the end of the Danish period. The education system was:

> "characterised by incompetent teaching by church 'teachers' and untrained boy monitors, petty graft in administration, and a grand total of nineteen tumbledown 'schools', a regime, in brief,....leaving about everything in the way of an adequate system to be desired." (p.38)

Post-1917 period

Turnbull (1982) records the important developments since 1917. He notes that the United States set out to Americanise their tropical

outpost through education based on essentialism geared to safeguarding democracy. Two school districts eventually emerged - St.Thomas/St.John, and St.Croix (1917-1955) - to match the two municipalities, but under one director of schools. From 1955-1980, there was one combined school district under a commissioner of education; in 1969 each district had a superintendent responsible to the commissioner.

Turnbull points out that the typical American plan, (6-3-3), gave the system its basic structure: 6 years of elementary education, 3 years of junior high, and 3 years of senior high school. This eventually developed into the current 6-2-4 plan. However, some time elapsed before the system was in place. For example, St.Thomas only got a permanent grade 12 in the 1930-1931 school year; St.Croix followed in 1935-1936.

From 1955-1980, the secondary curriculum had three tiers - general, college prep. and vocational. With project introspection improvements were made. The entire programme was revised: remedial, vocational, and minority education - mainly for students of Spanish background-was stressed, and the curriculum was made more relevant to local needs.

<u>The Board of Education</u>

The governor and the board of education share responsibility for education in the territory. Turnbull (1988) traces the history of the board. The board had its origins in the school commissions

of the early period. The 1921 school law established a Board of Educational Review in each of the two municipalities: St.Thomas/ St.John, and St.Croix. The boards' charge included recommending more schools, adopting curricula and recommending appropriations.

From 1968, the boards were elected and not appointed. The new law called for four members from St.Thomas, four from St.Croix, and one from St.John. The boards' powers now included the following - co-operating with the U.S. office of education, providing for the proper spending of the federal funds, and approving education personnel, both faculty and staff. In 1987, the Reorganisation Act established two district boards of education, with five members each for St.Thomas and St.Croix. Members serve four year terms, and joint meetings are held at least quarterly.

From the start, there has been a power struggle between the governor and the board. The latter wants similar powers to school boards on the mainland United States - to form policy, control schools, and hire and fire <u>all</u> personnel. The struggle was exemplified in August 1988 over the length of the school year. Akin (1988) called it "the steal-the-bacon war of nerves between the administration and the board over the public school system." (p.1).

The Department of Education develops a calendar for the school year. The calendar must be approved by the board and the governor

before it can be implemented. The law stipulates that the year should run from the first Tuesday after the first Monday in September until the last Friday in June. Only the board can change school dates.

The issue in question was whether the working but non-teaching days should be included in the stipulated 180 day school year. The board wanted to exclude those days, but the teachers' union and the governor disapproved. In the end, the governor got his way by declaring the disputed days "special holidays". June 22, 1988 was the last day of school for that school year.

The Teachers' Federation

The St.Thomas/St.John Teachers' Association affiliated with the American Federation of Teachers (AFT) in 1968. The local union found it had no influence when a Mrs. Charlotte Dogherty refused to substitute for a colleague and was fired. The case went to court, she won, and was reinstated. However, the union was advised that it had no effective power (A.F.T. pamphlet, May, 1983).

The St.Thomas/St.John Federation of Teachers has become one of the key power blocs in the territory negotiating separate contracts with government and using its political influence to improve the salaries and conditions of service for its members across the district. There is a similar union in the district of St.Croix.

Department of Education

The ladder of responsibility within the Department includes the commissioner, superintendents, assistant superintendents, directors, and co-ordinators. The commissioner of education is appointed by the governor and serves at his pleasure. Under the commissioner are two insular superintendents - one for the St.Thomas/ St.John district, the other for that of St.Croix - and a cadre of assistant superintendents. School principals are directly responsible to the assistant superintendents.

The directors of the various programmes are responsible to the insular superintendents. There are directors of curriculum, plant and maintenance, special services, volunteer services and similar services. Co-ordinators report to the directors. Co-ordinators oversee a variety of programmes - subject areas, drug free programmes, vocational education, and so on. Supervisors come within the purview of the co-ordinators.

Schools

In December 1988, there were 17 public schools in St.Thomas, 2 in St.John and 15 in St.Croix. In St.Thomas, there were two junior high schools, two senior high schools and thirteen elementary schools. The school population in St.Thomas was 10,625 pupils, with 1,854 pupils at Charlotte Amalie High School, and 1,013 students at Eudora Kean High School. There were 477 pupils in St.John. There was one high school in St.Croix, 3 junior high schools and 11 elementary schools. A grand total of 23,320 students were registered in schools in the territory.

For the same period, the non-public school enrollment stood at 6,556 pupils. Non-public schools were parochial, denominational or private. There were 17 in St.Thomas, one in St.John, and 24 in St.Croix. The total school population for the whole U.S.V.I. was 29,876 pupils.

The empirical study reported here was conducted in the only two high schools in the district of St.Thomas and St.John: Charlotte Amalie High School (CAHS) and Ivana Eudora Kean High School (IEKHS). The population of the two high schools was mixed and varied. It included pupils from across socio-economic boundaries; however, the majority of students came from large government housing projects. Students from the island of St.John were transported to Eudora Kean from St.John by ferry, free of charge. CAHS was an accredited school, but IEKHS had not totally met required accreditation standards for the Middle States Association of the United States.

CHAPTER THREE

REVIEW OF THE LITERATURE

The first section of this three part review examines the concept of audience. In the introduction, traditional approaches to writing are discussed. More specifically, the work of Moffett, and Britton and associates on a writer's sense of audience is examined. In the next sub-section, the need for new paradigms is established, and the limitations of the traditional models discussed. The work of Rosen, Elbow, Purves, and Flower helps to throw light on this area. After that, and in the next sub-section, the discussion shifts to different, and sometimes conflicting, definitions of audience. Various models of audience are discussed: the traditional views of commentators like Pfister and Petrick, and Hairston on the one hand, to the opposing views of Ong, and Long on the other, and then again to the synthesis of Ede and Lunsford which integrates those two perspectives. After that certain sub-issues are treated, and a short summary is given.

The Rhetorical Context

In a pioneer work, Moffett (1968, 1983) discusses writing in the context of rhetoric - although not a full rhetorical context. He writes of a "trinity of discourse", or a triangle of relationships, with three points to the triangle: the writer (I), the audience (you), and the product (it). Moffett shows that these three elements interrelate in various ways.

His discourse treats of the connection between writer and audience (I/you), and that between the writer and her information (I/it); however, there is little or no elaboration on the relationship between the reader and the text (you/it). Scholars like Park (1982) and Walzer (1985) have emphasised that rhetorical considerations must be enlarged to include the relationships between and among audience, genre and conventions of the discipline. Some of these relationships are discussed elsewhere in this literature review.

Distance

In the rhetorical relationship between writer and reader, Moffett distinguishes four levels of audience with an increasing distance between them. As physical separation increases, the possibility of immediate feedback decreases. At one level, there is no distance; in this intra-personal relationship, reader and writer are one person as in journal writing or note-taking, or other kinds of writing for self. At other levels, the writer and reader are separate persons. The closest relationship now becomes an inter-personal one as in conversation. At level three, writer and reader are further apart and can only communicate at a distance as in correspondence. Finally, at level four, the writer writes for publication - for a generally unknown audience, or if known, known only in a very general way. With this increasing distance, the likelihood of feedback becomes more and more remote; the feedback loop becomes more and more tenuous until it disappears, or becomes feedback through a public medium.

The writer also has a special relationship with his subject - she processes information at different levels of abstraction. The four rung of Moffett's hierarchical ladder are recording, narrating, generalising and theorising: a movement from the concrete to the abstract. Moffett sheds new light on the traditional genre categories which recent scholars of writing are finding increasingly inadequate as an accurate classificatory scheme. This issue is explored in some detail in chapter six.

But there is also the relationship between the audience and the text or product. The reader processes information in various ways. The text itself may be viewed as a kind of middleman, a bridge between writer and reader. The writer attempts to infuse it with her intent and purpose; the reader creates a meaning from the script somewhat like a musician interpreting a musical score (Kroll, 1984). The text itself may be autonomous holding a meaning divorced from that intended by writer, or gleaned by reader since "noise" in the text may cause the reader to miscue. There is a danger too, when the reader views the text outside of context. Witte and Faigley (1981) distinguish between cohesion and coherence. Cohesion is the web of relationships within the text, but this is a necessary - not a sufficient condition - for coherence; coherence is "a pragmatic unity, a unity of a text and the world of the reader." Witte and Faigley remind us that "all discourse is context bound - to the demands of the subject matter, occasion, medium, and audience of the text." (p.101-102).

The Writer Model

Traditionally, the teaching of writing has been informed by a classroom pedagogy derived from either the writer model - Moffett's first person (I), or the product model - Moffett's third person (it). There has been a patent imbalance; either the one or the other paradigm has been given too much weight, and audience has often not been treated at all. Mitchell and Taylor (1979) have argued that both the writer model and the product model are inadequate paradigms unless they are mediated by considerations of audience, and urged teachers to adopt an integrated approach.

The hallmarks of the writer model, according to Mitchell and Taylor, are sincerity and self-expression. The writer must be honest as he records and shapes experience, or faithful to fact as he reports and shares perceptions. The writer puts his personal stamp on the piece in creative even if idiosyncratic ways. Since reading is a creative act, the reader's reaction is also personal and idiosyncratic.

Mitchell and Taylor show that this model is problematic. The teacher reacts to the piece, but there is no systematic way to explain the source of that reaction to the student. Readers inside the culture often share experiences with the reader; both are connected to the work by common bonds of feeling. However, readers from outside the culture, or across cultures, may have trouble responding since they are often disconnected from the experience.

Another aspect of the model which Mitchell and Taylor find inadequate is the high regard for facts as information in their own right, divorced from context. Such an approach separates the medium from the message and distorts the communication process. One way of solving the problem is to think of sincerity as a rhetorical effect in which the writer projects sincerity, cutting off himself from his persona.

The Product Model

Mitchell and Taylor also reject the product model. The product model is feature oriented; the guiding principles are "clean" copy and an error free text. The assumptions are, argue Mitchell and Taylor, that teachers can and will agree on what these features are and apply them consistently when evaluating writing.

They quote a wide range of research as a basis for rejecting this model. They stress that there are no absolute standards as different disciplines will have their own standards, or have a right to set such standards without interference from English departments. Second, the concept of error is constantly changing; attitudes to error are becoming less fixed.

For example, the work of Kline and Mmering (1977) has thrown new light on the concept of fragment, and the work of Shaughnessy (1979) has shown that error must be placed in the context of larger

discourse elements if teachers are to make any impact on it in their teaching of writing. Shaughnessy points out that most of the errors basic writers make derive from inexperience in using what for them is a new technology. Minsky (1975) stresses the notion of discourse frames which are necessary heuristic devices for writing as problem-solving. The distinction between text and context has already been touched on above.

The inadequacies of the writer and product models underlined the need for a new focus. However, there were other compelling reasons for the search for a new paradigm. One consideration was the traditional conditions of school writing.

School Writing

In school, students often write for the teacher only. From one point of view, they are expected to. Teachers are the coaches; students are the learners of the game. However, pupils often fail to transfer even basic skills to writing tasks outside English classes. *English*, they insist, is the domain over which only the teacher of English rules. But writing for the English teacher only, also poses a variety of problems, and sets limits on effective communication. Scholars distinguish between school writing and out-of-school or "real world" writing. And there are telling differences!

Rosen (1973) comments on the power of the teacher on the classroom in general, and his dominance over the field of writing in particular. The traditional teacher initiates writing. He tells

his pupils what, when, how, why, where and if to write. The teacher also dominates response. Teachers themselves hardly read student essays for pleasure, or to learn about a topic since they are experts who know already, and are merely scanning texts with a hypercritical eye. This often results in unenthusiastic and dull student efforts designed to meet minimal acceptable standards.

Elbow (1981) explores this same theme. He opines that the teacher as sole judge is all right as a testing device, "but it's peculiar as a communicative or audience relationship". (p.219). There is a fictional quality about classroom writing transactions as both teacher and student play roles in the land of make believe. However, it must be pointed out that nothing is inherently wrong with that if the writing simulates the real thing. Writing for teachers has definite advantages. Teachers have to read student essays; they are parsimonious with their criticism and quick with their praise. But students need to play a game against a real opponent at times. Moffett urges teachers to "create more realistic communication dramas". (p.12). Students should write for the class group using raw material from their world of experience, to discuss themes workshop fashion.

Britton, Burgess, Martin, McLeod and Rosen (1977) illustrate the dominance of the teacher as audience in school writing. In a developmental study, the Britton team developed a scheme for classifying school writing by audience and function. The team then collected over 2,000 pieces of writing across the curriculum from pupils aged 11-18 for analysis.

Uni ers'ty
Library
Hull

In an analysis of the data, the London research team found that the bulk of the writing (95%) was done for the teacher as audience, and writing in all other categories represented a mere 5%. In turn, the teacher as examiner categories dominated the teacher categories (49%), with the teacher/learner dialogue sub-category (39%) a distant runner-up.

Teachers then must be very conscious of the roles they play in the writing situation, and then make pupils aware of those roles. Teachers must fill out the rhetorical context for writing "exercises" to indicate purpose, audience and function. The point has been well made by Purves (1984).

Teachers, according to Purves, play a variety of roles in responding to school writing: the teacher may read and respond, read and judge, read and analyse, or read to improve. The first role is very much like the teacher in the writer-mode as described by Mitchell and Taylor. The pupil expresses himself and the teacher gives a reaction. It is non-judgemental and non-threatening. In the second role, the teacher is a proof-reader guiding students to produce "clean" copy and a presentable text. Here the product model is dominant and the teacher focuses on features of writing at the word and sentence level.

Other roles include that of the copy editor - the teacher focuses on larger elements of content, style and organisation; or reviewer - the teacher judges the quality of the text for potential readers; or as a diagnostician - the teacher determines and treats

students on an individual basis. "In class, a teacher may pursue any of these roles or at times ... all of them in reading a student paper." (p.262)

Writer-based Prose

Still another reason for a new paradigm is to counter egocentric writing. The work of Piaget and Vygotsky in the language development of children has helped writing theorists to understand what Linda Flower calls writer-based prose. Flower (1979) uses the features of inner and egocentric speech to explain why some writers are disconnected from their readers, and why some pupils have problems with writing. Writer-based prose and reader-based prose differ on the dimensions of function, structure and language. For Flower, the one is writing for self, has a narrative structure reflecting process, and language that is private with an unexpressed context; the other is communicative, issue-centred reflecting purpose and has a shared context and language. She argues that students must learn to transform writer-based prose to reader-based prose.

In writer-based prose, the writer is not writing for an audience outside self. Reasons for this vary. First, the task may be difficult, and the writer therefore spends the time thinking, exploring, problem-solving. Cognitive psychology teaches that since processing space is limited, for such a writer there is no room in short-term memory for much else. One strategy the writer uses to explore is narration since it is an efficient way to rehearse information, though not to analyse it. In such writing, words may be loaded with

specific meaning for the writer only, or the lack of context may cause difficulties for other readers.

Writer-based prose can be a useful first step in the writing process. Flower (1981) explains that at the pre-writing stage of composing, writer-based prose can have a real use; it "comes naturally to us; it can be an efficient strategy for exploring a topic and outwitting our nemisis, short-term memory." (p.164)

Elbow (1987) also discusses the point in presenting an argument for ignoring audience. He points out that, in some instances, writer-based prose can be an end in itself as in journal writing. He observes also that a writer's voice can be so powerful that it forces the reader to decenter. As a problem-solving technique, both Flower and Elbow see writer-based prose as a useful strategy. Elbow advises students struggling to solve a particular writing problem to write to no audience, to self, or to an inviting audience. But, once problems are solved, they should revise with an audience in mind if the writing is to be shared.

To summarise, theorists have found traditional models of writing inadequate, and consequently made the case for a new focus. Writing as self-expression is inadequate because at worst it has an ego-centric focus, or at best, a limited context with consequent problems for evaluation. Writing as product concentrating on an autonomous and error-free text and focussing on specific features presents certain problems as well. The full rhetorical context of writing,

especially the dimension of audience, becomes a necessity for the effective teaching and learning of writing.

Audience

Definition

Composition theorists disagree on what audience is. Indeed, they disagree not only on the definition of audience, but also on other issues such as how to teach, and when to teach it. Some of these considerations are discussed in this section.

Some composition theorists think that a writer's audience is very much like a speaker's audience. They view the audience as real people - a collective located in time and place. A writer analyses her audience to gauge its characteristics and to adjust to its needs.

(Pfister and Petrick, 1980; Flower, 1979 and 1981; Hairston, 1982). These scholars have suggested certain heuristics for analysing audience. The assumption is that a writer can know and analyse her audience *before* writing to make for effective communication.

Hairston describes a rhetorical square - 4 points in a quadrangle of communication: purpose, persona, audience and content. Relative to school writing, she cautions that exercises are only *exercises* and that student writers "need to move beyond the context of their composition class into *real writing* situations, either actual or simulated." (p.73-74). Hairston's heuristic for analysing audience is as follows.

With a known audience, one close to the speaker, writers may use a commonsense psychology derived from experience. For more remote audiences, writers can catalogue traits. Writers can collect demographic data on the values and concerns of the readers, their socio-educational and economic background, and their attitudes toward religion, sex and politics. After analysis of this data, writers are more likely to influence readers by adapting to their needs. The Hairston heuristic is based on wh-questions: <u>Who</u> are my audience? <u>What</u> is important to them? <u>What</u> is their economic and social class? <u>Why</u> are they reading my paper? <u>What</u> do they know about the subject? <u>What's</u> their attitude to the subject?

Ede (1979) holds similar views. In her classes, she demands that her students submit an audience statement with each assignment, as well as a possible means of publication, after making a purpose-oriented audience analysis. They must ask themselves questions like - Why am I writing? Who is my audience? What is the occasion? What constraints are operating?

Flower (1981) agrees. She urges the writer to know the needs of the reader, to find common ground with him and close the gap between knowledge, attitudes and needs. She encourages the writer to ask questions to discover what the reader knows, or needs to know; to determine the writer's image of the subject, and if the writer can adapt to its needs. For Flower, knowing the needs of the reader is critical, for readers do not passively receive information but actively turn message into meaning.

Pfister and Petrick (1980) urge student writers to construct as nearly as possible a replica of real readers. They have organised their writing courses around a 4-pronged heuristic. First, they focus on audience/self: its status, experience, values and biases. Then they focus on audience/subject: its knowledge of, and attitude to the subject, and the basis and strength of that attitude. After that they explore audience/writer relationships: knowledge, attitude, shared experience. And finally they address audience/form concerns: the best method, best time, best diction, best syntax for the writer to adopt so as to achieve his purpose. They teach the concept of audience by giving their students practice in composing for particular rhetorical situations. Pupils work in pairs to write and criticise character sketches, analyse and write for an imaginary audience, and create their own audience for a process paper.

The three models above are based on the assumption that a writer can know his reader in the same way that a speaker can know his audience. This is the traditional stance deriving from Aristotle's (Ede, 1979 and 1984; Kroll, 1984). But other theorists feel that there are fundamental differences between speech and writing.

For the classical rhetoric, the audience was a physical reality, situated in a particular context and providing feedback that the speaker could immediately respond to. However, the audience of the writer is a general concept. Moffett's sequence of intrapersonal, interpersonal (near), interpersonal (far), publication shows that

the wider the distance and space between writer and reader, the less writing is akin to speaking. Scholars like Gere (1987) suggest that speaking and writing are complementary, not polar activities.

Another limitation of the rhetorical perspective is the adversative nature of rhetoric. A writer is perceived as having an eristic purpose, but all communication is not adversative. Ong (1975) and Long (1980) share this view. Long considers the rhetorical perspective agonistic, too assertive and "noxious stereo-typing". (p.223) Ong sees the writer's audience as a fiction created by the writer. A writer's audience for him is different from a speaker's audience. What happens in the writing act is that both the writer and the reader are playing roles. The writer creates a role for the reader and inserts cues in the text to help the reader understand that role. The reader, in turn, adopts the role - "fictionalises itself" and plays the role imposed upon him. Hence, written discourse is a construction of the mind.

Ede and Lunsford also find limitations in this position. They see it as overstating the writer's power and independence, just as it over-emphasises the differences between speaking and writing.

Generally, there has been a move away from these extreme positions. The model proposed by Ede and Lunsford takes into account many of the issues involved. A description of the model follows.

Mitchell and Taylor argue for an integrating perspective. They show how the writer and product model can be reconceived in relation to the audience. However, the power they give to the audience has been criticised by some, including Ede and Lunsford.

There is often no hard and fast distinction between writer and reader. Ede and Lunsford see audience as a complex, fluid and multiple concept. A writer is also a reader of her own text who may adopt a range of roles for herself, as well as her audience. The audience (as self or outside the self) processes the text; however, it is not a passive receiver of information. The audience may adopt the role assigned by the writer, and so agree to be acted upon; on the other hand, the audience may not and use its own knowledge, attitudes and interpretations to make meaning.

The audience too, may be invoked or addressed - created or real. Ede and Lunsford suggest a taxonomy of distance which reminds us of Moffett's. Audiences invoked run the gamut through writing for self, friend, colleague, critic, mass audience, future, past, and anomalous audience.

A writer writes for himself, or reads his own text with some end in view, for example, to proof-read or revise for a reader. Or the writer may write for a friend. This reader may be a part of the intended audience - say, a particular academic community - who will give him honest feedback in regard to the effectiveness of the communication for those readers, somewhat like Elbow's enabling

audience. A colleague is like a friend, but comment may be more guarded and neutral in intent. The function of the critic is different - somewhat like that described by Purves above. Then the audience becomes more like a collective. With a mass audience, the writer must rely on general characteristics, but must also think of subgroups with their special interests. Finally, a writer will rely on past experience with different types of present or future readers so as to adapt the message to their needs. In short, an audience is not just actual; it includes "all those whose images, ideas or actions influence a writer during the process of composition." (p.167) Audience addressed also covers that range, but omits past and anomalous audiences.

But if the real audience is outside the text, then only through the text can it come to life. In the same way, it is only through the text that the reader can see the writer at work, and experience his play of mind. In summary, "writers create readers and readers create writers." (p.36) There is a wide and shifting range of roles, and writers must relate audience to the full spectrum of discourse.

When to Consider Audience

One subissue debated is when the writer should consider audience. For Pfister and Petrick, a writer should consider audience before he writes. Flower (1979) suggests that students with problems in writing use writer-based prose as a first step, and think of audience after the first draft. However, she points out that a writer can

think of audience <u>anywhere</u> in the writing process - considerations of audience are "nested, or embedded, within other phases". (p.130) It is "a major, functional stage in the composing process and a powerful strategy well-fitted to a part of the job of writing." (p.34)

Elbow (1987) suggests ignoring audience at first to avoid writer's block or to explore a topic, and <u>then</u> considering audience. Writing demands a complex of skills and the writer can suffer from cognitive overload. Again, an audience may be threatening to a writer causing her not to write well, or at all. Writing becomes an unnatural process. In such cases, Elbow advises, one can write for "a safe non-audience" (p.188) or for the self; or for a safe non-threatening audience.

He warns though that that is only applicable to getting-it-right writing where one can ignore audience at the beginning and choose "<u>when</u> to enter its magnetic field." (p.199). In other writing acts, for example, transactional writing, one must attend to audience from the start. Elbow warns that to always figure out your meaning before you start can be bad advice. Better advice is to consider audience "sometimes before you finish, figure out meaning and think about audience." (p.198)

Roth (1987) studied the composing process of three students in an advanced placement class. He asked them to write an essay for

publication on campus - for a general audience. The students shared all notes and drafts with him, and interviews were taped.

Roth found that one writer revised her audience as she composed, and that his advanced writers addressed audience selectively. One student said that at times he addressed his "best self" (Elbow), or an ideal reader or wrote for himself at times. Roth concluded that viewing audience only as a real world reader <u>external to and predating</u> a text may be misleading. Audiences are fixed and definite as well as indefinite and multiple.

<u>Discourse Communities</u>

The next issue to be explored is the relation between writing and a particular universe of discourse. Flower (1981) warns that "if you depart too greatly from your genre ... you are likely to confuse or disorient your reader." (p.144)

Berkenkotter (1981) did a protocol analysis of 10 professional writers - 5 professors of rhetoric and 5 from other disciplines. She was interested in the intellectual process of experts as they write. Writers played the role of speakers giving a speech on career day to a group of high school students. Some wrote in the narrative mode, others wrote informative compositions, and still others wrote in the persuasive mode.

Berkenkotter found that experts seem to have scripts which are audience related stored in their long-term memory. Although

her research was not conclusive as to whether a writer's discipline determines his representation of audience, Berkenkotter notes that a writer is constantly revising her own work with a particular audience in mind.

Walzer (1985) did a case study of the concept of audience. He analysed 3 articles written for 3 journals on the same research project. Walzer found that the rationale and "rhetorical ground, form and significance of data were determined by the discourse or interpretative communities" for whom the writing was intended. He urged teachers to "introduce students to the concept of an interpretive or rhetorical community defined by the conventions of its discipline." (p.157)

Summary

A teacher's image of audience will determine how he or she teaches it. There will be a variety of methodological techniques to match the different perspectives. Kroll (1984) isolates three such perspectives: the rhetorical, the informational and the social, but each - as Kroll points out - has its limitations. In the final analysis, the text is the link between writer and reader.

Kroll warns that one danger of the informational perspective is that writing can then be viewed as writer encoding and reader decoding. Scholars like Hirsch stress that texts should be designed for readers so that there are few obstacles to comprehension.

It is true that this minimises the role of the reader as a creative participant; still it is only via the text that a writer can address or invoke his reader, and a writer must somehow get the reader to realise his purpose by adopting, style, content and language to genre and discourse.

Further, if audience is a multiple concept, a writer must address different sections at different times through the text, and students should be taught to analyse different audiences for the same communication. In a way, this is like writing for a general audience which can be viewed as the top level of the Britton and Moffett taxonomies of audience. So the wheel in this respect has come full circle. Indeed, Britton, Park and Elbow urge that normally pupils should be taught, in the final analysis, to write for a general reader.

Despite the various points of difference, theorists and practitioners seem to agree that good writing must be informed by considerations of the reader(s). One way teachers can help pupils to do this is by devising realistic tasks in which roles of the pupil and teacher are made explicit; as Moffett notes; students must write often as first person or second person. Again, pupils too often are writing for the teacher only. An alternative is to let students write to real people- their peers in class, their peers in another school, to friends and family, to the press; or write for different purposes and functions; or write for wider publication in some form.

The teaching of audience is of necessity ultimately connected to group work. It is this which is treated in the next section.

Group Work

In this section, the use of peer groups as a pedagogical technique is examined. An advance organiser for the section is presented here. First key terms and concepts are defined, and then the value and importance of group work as a technique explored. Next, the advantages and disadvantages of group work are treated, and conditions for its effective use explained. After that, there is an overview of strategies for setting up and organising groups, student and teacher roles, and the nature of reporting. The above and other related issues are discussed in the light of a number of research studies. Some of these studies focus on revision in peer groups, others on the efficiency of the group method, and still others on formative evaluation techniques.

Definitions

In chapter one, group work was distinguished from children working in groups. The first is collaborative learning; the second involves arranging the traditional classroom in clusters for efficiency, and is not necessarily group work as the term is used in this research report. The heart of collaborative learning is the "group's effort to reach consensus by their own authority." (Wiener, p.55). Moberg (1984) reminds us that group work is "as old as bones", and that the ancient Greeks as well as Medieval scholars used it, but the distinction made by Wiener is not stressed.

Consensual learning then depends on the nature of the task.

This issue is examined in some detail below, but Barnes (1982) points out that when the task assigned students is open, the climate is set for pupils to operate in a hypothetico-deductive fashion. Commenting on the way four eleven-year-olds discussed a poem, he concludes that in open discussion pupils "work out their interpretation in collaboration: one puts forward a view, another takes it up and modifies it, another finds evidence, and another sums it up." (p.25) Group work, as used in this study, means collaborative learning.

Pierson (1967) defines peer groups as "students who are similar in development and educational status and who are assigned to the same English class section." (p.2) The definition raises some vexing problems, for example, the question of who are peers is not always clearcut. Should there be heterogeneous clusters where pupils are different in development in a gross way, or should there be an arrangement where homogeneous groups within the classroom are formed according to student ability? On what other bases could students be assigned to groups for maximum efficiency? And are students across grades taught by different teachers or by the same teacher peers? Some of these issues are treated by Harris (1986). Despite these concerns, the definition of peer groups in this study follows Pierson's definition.

The Importance of Group Work

In the first section of the literature review, the limitations of the teacher as a sole audience for writing was outlined. Peer groups present a potentially powerful alternative for students to extend the scope and range of their work by writing for different purposes to different audiences. Barnes notes that the small group offers a way of distancing a teacher's control." (p.190) He suggests that when the teacher's dominant instructional mode is teaching the class as one group, students do not make full use of language for different purposes. Some of these uses are talking to themselves in their own language, exploring ideas, planning and recording experience: in short, talking to learn.

Educators have used group processes for a variety of purposes from editing and rating to conferencing. For example, Bean (1979) has used it in upper college literature where his students code and then grade their own work using a 6-point rating scale. He reports few disagreements, and gratifying results since there are few cases where the disagreement exceeds one score point. Fox (1981) used group work to improve student grammar competence.

Composition experts think of writing as a process. Emig (1975) points out that these so-called stages are not locked into an algorithmic sequence; they "occur and reoccur (sic) throughout the process." (p.67) One composition text "encourages each student to develop an individual writing process - a process which is not

linear but recursive" (Blum, Brinkman, Hoffman and Peck, 1984, p.vii). It is conceivable that students use language differently at different stages in this process: at one stage for exploration, at another for recoding, and at still another for public presentation. If this is true, peer group discussion and interaction can help students to work through the different stages of the writing process.

Reference was made in chapter one to the Dusel California study. Stanford (1979) refers to the NCTE endorsement. He argues that if every week

> "you assign a composition to all classes, you will need approximately seventeen hours to respond to all of them. By having students write a composition only every second week, you could complete your grading by spending one hour a day on weekdays and two hours on weekends. Not a very unreasonable expectation." (p.xii).

However, Stanford argues that a five class teacher load with a maximum class size of 30 students was the norm. He concludes that simple mathematics shows that the job of teaching English, considering present-day realities, may not be humanly possible. Stanford was writing of the United States; in most developing countries, the situation is even more acute. Where school districts have reduced workloads, peer group activity can make for an increase - not a decrease - in student themes per week. In developing countries, group work must be a sine qua non in the average classroom.

The research studies to be examined spell out the specific benefits of collaborative learning, but the larger context of talking to learn, effective writing strategies, and practical necessity provide a useful point of departure for a closer study.

Moffett (1968) points out that group work "promotes the social art of conversing, the intellectual art of qualifying, and the linguistic art of elaborating." (p.92) Davis, Scriven and Thomas (1981) point out that group work can provide useful formative and even summative information on writing skills. They underscore the value of collaborative learning for its anecdotal richness, flexibility, and helping teachers to understand more subtle programme effects; they conclude that group activity is "sound pedagogical practice". (p.94)

Beaven (1967) sums up the advantage of peer groups: the use of peer models seem more efficacious than professional models, editing and revising is more palatable, and the reduced paper load frees the teacher to conference more with students, and give more individual attention to those who need it. She discusses group work in the larger context of the development of adolescents.

Beaven stresses that adolescents writing for other adolescents - their significant others - can develop their own voices, correct distortions in their perception of the world, and strengthen interpersonal skills. Achievement depends on the fulfillment of the adolescent's need for affiliation and power, and his development of a sense of self.

In sum, the educational value of group work, the opportunity it provides the adolescent for personal growth, and the development of inter-personal skills make collaborative learning a useful tool. However, teachers need to be aware of the disadvantages of the method - or rather, its possible drawbacks - and the preconditions that must be satisfied to make it an effective and efficient strategy.

The first drawback is that the process takes time. Beaven notes that the teacher can spend half her time on process and half on task. The typical syllabus tends to be overloaded in terms of content coverage, and when teachers feel they must cover everything, frustration sets in. Besides the development of interpersonal skills take time which also comes at a premium. The teacher too, may need a short course or workshop before she can feel comfortable with the process. For the teacher's responsibility can be taxing. She has to make the necessary physical arrangements, supply materials, monitor behaviour, consult with various groups as required, and become a member of a group when necessary. Thus, logistical considerations, together with a variety of roles often make collaborative learning a nightmare for the teacher.

Again, teachers are often diffident. Pupils are often perceived as usurping teacher roles. Both teachers and students must make a psychological adjustment. Some issues include the following: Can students be taught to assign a final grade to a piece of work?

Can teachers, administrators, parents and other agencies have confidence in the process? Under what conditions can pupils share in the evaluation process? Beaven stresses that a precondition for effective group work is the development of a climate of trust. This is so for both teacher and student.

And other composition specialists too, stress the need for a climate of trust as a necessary condition for collaborative learning. Harris underlines the need for a "caring, thorough, and honest peer evaluator" (p.8), Davis et al, mentions the need to set the proper climate before group activity starts.

Katstra (1987) reports a significant difference in attitude between students writing in groups and students who did not. She notes however, that her teachers used a special unit to develop an enhancing climate before the treatment started. The unit included special reading assignments, a library search, a field trip, the viewing of a film and much discussion. Even so, the peer method is no cure for all ills. Harris cautions that "teachers need to be aware that peer evaluation of writing is not an experience in which every student needs to be involved. A few good students prefer to work alone." (pp.5-6)

Despite the caveats raised above, group work has the potential to be a powerful teaching/learning instrument. Professionals who use the method, and a number of researchers, have developed efficient ways of setting up and organising groups. It is to these that we now turn.

Arrangements and Procedures

A number of teachers who use collaborative techniques have described the procedures they use. Their comments may be collapsed under the broad headings of arrangements and procedural strategies. Relative to setting up groups, they discuss the issues of size of groups, their composition, permanence, physical arrangements and functioning. In regard to the group process, they discuss rater response guides, teacher and student roles, kinds of tasks, time allowed, and feedback arrangements. These are examined in this section of the review.

Hawkins (1976) comments that 5 is the ideal size of group, and many commentators agree. Moffett advises that the maximum should be 6 since, as a rule, large groups are poor for students' learning discussion techniques. Barnes (1982) used three groups of three girls each, and one group comprising two boys. Moberg (1984) recommends an ideal size of five, but notes that clusters of four or six can work as well. However, groups of three or seven hardly work. In the first case, one student can be isolated; in the second instance, the group often divides into subgroups. Grimm (1986) cautions that groups of four or five do not "automatically insure that everyone will receive useful response." (p.91) One issue not yet quite resolved is how permanent groups should be - how long the same students should work together. Moberg advises that pupils should switch for the first few days. Hawkins recommends permanent arrangements.

For group work, the furniture has often to be rearranged. In the traditional classroom, chairs and desks were arranged to give the teacher maximum control; in most classrooms today, each pupil has a separate chair and desk. Groups of five are often arranged in a horseshoe format so that pupils face one another, or can turn to face the teacher when necessary.

Beaven (1976) details a four-step procedure. At first, students work in pairs on any subject, and on tasks that last no longer than 15 or 20 minutes. The ground rules are that pupils work with someone they do not know, or have not worked with before. After that, pupils work in groups of four on tasks lasting for the same period; group composition changes with each task. However, the teacher assigns roles and one group may model procedures. At the next stage, the teacher assigns students to groups for longer projects. She focuses on group dynamics and interpersonal skills. Finally, students choose their own groups for sustained projects and support. Pupils may return to an earlier stage if and when necessary. Beaven indicates that "the desired growth in writing seems to occur when students work with the same group for an extended period and where there is less structure." (p.148) Groups may be set up in various ways.

Sometimes, arrangements for group functioning are rather formal. Students select a chair, and a recording secretary who keeps a log to be returned to the teacher after each meeting. At college level,

students exchange telephone numbers. Efforts are made to foster a sense of belonging, for example, a catchy group name is selected; or to develop leadership qualities; chairs and secretaries rotate. In sum, the logistics of group work call for careful planning to avoid chaos and confusion in the classroom.

The key to collaborative learning however, is the writers' workshop. What pupils discuss is important, but how they talk is crucial. Hawkins and Barnes urge teachers to assign open-ended tasks that call for a variety of possible solutions, so that pupils have to co-operate to piece together a solution. Students, too, must have a framework for responding, if not, Wiener warns, they "will just pat each other on the back, attack each other counter-productively, or fall silent." (p.57) Teachers discuss the assignment, including grading criteria, and circulate rating scales or other response guides.

Students normally divide into groups to examine drafts, or discuss their papers. Often xeroxed copies are circulated when pupils are responding to the same script. Specialists recommend spoken as well as written feedback. Hawkins recommends a sophisticated sequence. First, pupils must read and not write. Then they talk to the author about the good features of the paper, and the parts they like. After that, they talk to one another. The author next has the opportunity to ask questions on specific areas on which she wants advice. The group then write specific comments at both the macro

and micro levels, and all responses are passed to the student who revises the script. The writer turns in both drafts as well as group comments to the instructor who reads, responds and criticises.

Other practitioners follow a similar sequence - oral and written responses - though often the comments are written at home. The important consideration is the expressive talk "which helps the writer find a stronger focus and develop his or her points more specifically" (Grimm, 1966, p.92). Students must be taught to ask the right questions, a difficult task since their usual role is to answer questions posed by the teacher the answer to which he already knows.

As indicated earlier, the teacher has to function at different levels. He must circulate to answer questions, guide response or participate in discussions. At another level, he must be a manager keeping order, facilitating group processes, or solving logistical problems. He must also manage the mechanics of social relations. He must direct plenary sessions and help groups resolve conflicting reports.

Wiener comments that an important role of the teacher is that of synthesiser, helping his class to join another discourse community: "By synthesising the results of the individual groups, and comparing that synthesis with the consensus of the larger community of knowledgeable peers - the teacher's community - the teacher helps complete the movement into this larger community." (p.59)

Empirical Research

Talking to Learn

The procedures described so far have been informed by research in the field. The next section surveys this research according to function. The research of Barnes (1982), and Russell (1985) focus on talking to learn. Other studies examine the use of peer groups to enhance student revision or editorial skills. But many researchers have carried out experiments to determine the importance of groups in writing assessment (Pearson, 1967; Lagana, 1972; Sager, 1973; Ford, 1973; Harris, 1986; Katstra, 1987).

Barnes (1982) was interested in the effectiveness of the verbal strategies pupils use in problem-solving in groups. Other research considerations included how best to set up groups for learning, and teacher involvement in group processes. Barnes conducted a study at an unstreamed suburban comprehensive school in England. He used a small sample of 11 second-year students (pupils aged 12 to 13). Teachers divided the pupils into four clusters to respond to tasks in science, English and history. There were three groups of three students and one group of two. Discussions were recorded and protocols studied.

Barnes found that pupils used either a closed or open approach to problem-solving. When students used a closed approach, they limited themselves to the parameters of the task. They focussed only on what was asked for, and asked no new questions. There seems to be a lack of ability to extrapolate or go beyond the data. Pupils worked for consensus - or rather census - and rarely disagreed with one another.

Conversely, students who used an open approach used language to elaborate thought. They operated in the hypothetical mode to extend, modify, and summarise information given by other students. Barnes suggests that if pupils are to understand texts when they are working alone, or in groups without a teacher's guidance, students must have an "open and hypothetical style of learning". (p.52) Barnes also found that success depended on children working in the interactive mode, using the language of active engagement.

This is the position held by Grimm who states that "expressive talk . . . helps the writer find focus and develop his or her points more specifically" (p.92). He remarks that "when one student picks up another's comment and extends it further, she is heightening the credibility of peer response, deepening the group's understanding of what is often undeveloped, and also insuring that one student doesn't dominate the discussion" (p.94). Barnes agrees but warns against expecting final draft language from students without allowing enough time for student knowing how to do it.

Barnes also found that there were essential differences between the small peer group and the class as one large group. Specifically, he noted that small groups provided pupils with the opportunity to use language for exploration and discovery. This use of language was characterised by incomplete and inexplicit statements, hesitation, shifts in direction, and what Flower calls the use of code words. In contrast, when students were reporting to a larger group, language served a different function. It was more ordered and public.

Consequently, Barnes recommends a two-part sequence for teaching students how to talk to learn.

During the focus stage, the teacher presents the topic, discusses it and models behaviour if necessary. The students attempt to convert school language into "action knowledge" through assimilation and accommodation. This is followed by the exploratory stage during which the pupils manipulate materials and talk about the task. In phase two, students need to reorganise and refocus to prepare for reporting back before presenting their findings to the larger group where further discussion can follow.

Finally, Barnes suggests four considerations for setting up groups. Students should experience a feeling of competence. This happens when there is an open approach, and an explicit task; when pupils feel their contributions are valued, and there is a sense of relevance of the assignment to what they are doing. Students should find a common ground for discussion by sifting and recoding evidence. They must be taught to focus by asking the right questions, and there should be adequate pacing during which language as performance is balanced with language for learning.

Revision Studies

The Pierson, Russell, and Katstra investigations were <u>revision</u> as distinct from evaluation studies in the sense of pupils assigning

a rating to the writing of their peers. The studies of Lagana, Ford, Sager, and Harris were evaluation studies per se. In the Lagana, as in the Pierson study, pupils did not assign grades. However, each student met with the teacher to assess individual and group progress and to agree on a letter grade. The revision studies are discussed next. After that, the evaluation studies are examined.

Pierson (1967) investigated the effect of peer versus teacher correction on the writing ability of 153 grade 9 students. The sample was drawn from a middle-class, suburban, public junior high school in New York. In a carefully designed experiment, Pierson controlled for sex, intelligence and pre-experimental differences in writing ability. The six teachers in the study taught prewriting and brainstorming techniques before students composed. Students did all their writing in class, and then were grouped into clusters of four or five to correct drafts with the aid of guide sheets. Writers were encouraged to ask questions and question corrections. After themes were corrected, students revised drafts and stored them in their portfolios. Periodic tests were given in correcting and revising. The writing of the control group was corrected by the teachers after school.

Pierson found no significant difference between groups after analysis of post-test data. There was no difference in writing ability as a result of the programme. Intelligence, sex and pre-test differences in writing ability had no significant effect on

post-test performance. Pierson also found that "revision was occurring, but not always. Some writers did not revise adequately; although they made some corrections, they omitted others. There were wrong corrections, particularly in spelling. Some criticisms were misinterpreted" (p.59). However, group work was a more efficient teaching/learning strategy. The peer method fostered "esprit de corps," independence of thought, and students had generally a positive attitude towards the method.

One methodological issue arising from consideration of the Pierson study is who is responsible for the assignment of grades; a design issue is whether the same instructor should teach both the control and the experimental groups. Relative to the second issue, Pierson observed that the "teachers agreed to devote as much enthusiasm and effort to teaching the control group as to teaching the other; nothing in the actions or attitudes of the participants was found to contradict this agreement" (p.61). The researcher was also an administrator in the school where the experiment was conducted, and had the opportunity to monitor proceedings closely. Still, questions of possible rater bias remain.

Russell (1985) conducted a study to discover how peer conferencing related to the revision of student writing. Conferencing was defined as a type of critiquing whereby a peer of a teacher questions students about the content of their writing" (p.4). Calkins (1981) had shown that students as early as grade 3 could do four types of revising after conferencing. For some, the questions of their peers seemed

to have no effect. These wrote several drafts without considering earlier ones. Other students revised for mechanics only; others moved back and forth between drafts to consider the best possibilities. Still others - the brighter students - internalised their audience, stood back from the material to look at it objectively, and revised selectively.

Following Calkins, Russell used the case study method with a carefully selected sample of four students from a rural school with a population that showed great contrast in socio-economic levels. The sample was categorised as low, middle, or high in writing ability on the basis of scores on STEP tests taken in grades 4 and 5, reports from other teachers who had taught the students, and the researcher's own judgement of the ability of the group. Russell taught the skill of conferencing during one semester, and then conducted the study. Observation data were collected - but not by the researcher - and protocols analysed.

Russell found that students used three levels of questions: questions on specific details, <u>what</u> questions; general questions calling for reflection on the part of the writer, <u>how</u> and <u>why</u> questions; and questions involving syntax. Russell also found that poor writers could effectively conference with their peers, but depended on the questions of others to revise their own writing. Sally, a low ability student, asked 70 questions which included 44 type two questions. However, the two low ability students asked no type three questions. What is significant though is that they made

25 type three revisions among them after conferencing. Average and good writers had the ability to conference with themselves, and revise on their own. Russell comments that "improvement in the writing of these students is slow and more difficult to see . . . than improvement in poor or average writers" (p.8). She opines that one reason seems to be that because such students have hardly have the need to proof-read for mechanical errors, they are less likely to revise at the macro level.

Katstra (1987) conducted a study to discover whether student evaluation of first drafts would lead to increased fluency as defined by word count, and whether peer evaluation would improve student attitude towards writing. The sample included the entire cohort of 177 grade 9 students with very high I.Q.'s, from a white, suburban, upper middle-class school. Three teachers were randomly assigned to the 7 intact classrooms.

Both the experimental and control groups worked through a unit specially designed to create a sense of trust. They shared personal writing, visited severely deaf students, were assigned special reading, watched a film, and did a library search - all based on the topic "Understanding people who are different." After that, both groups wrote first drafts of a paper as a pre-test. Two attitude measures were given at the beginning and at the end of the study. The experimental group then received training in peer evaluation techniques, and then met in groups to discuss their drafts. The control group rewrote their papers with help from the teacher - but only if they asked specific questions.

Katstra found a significant increase in positive attitude in the experimental group over the control group. Girls in both groups showed a more positive attitude than boys. Students said that peer evaluation reduced fear, that they enjoyed reading one another's scripts and that their writing had improved. Katstra warns that the study was not concerned with writing improvement in a global sense, although fluency is one index of competence in that area. It was also found that pupils in the experimental group had a decreased word count on the final draft of the essay post-test, but their writing was tighter and more controlled. There was a three-way interaction between and among teachers, sex, and group. Two females and a male taught the seven classes.

It is instructive to compare how these researchers handled certain methodological concerns. Both placed a premium on pre-training and sensitisation. In the Katstra study, more time was spent on the unit to create a climate of trust than on the actual experimental treatment! Russell spent the first semester of the school year teaching her class to conference. One difference is that Russell taught her own class; she did however, use an independent observer in her study. Although concerned mainly with revision, the Pierson study related "clean" copy and writing ability. The Katstra study treated writing only tangentially. The Russell investigation treated revision and writing in a larger context. The researchers wanted to discover where students fell on a continuum of skills on a growth scheme of revision types. Nevertheless, none of these

studies treated the use of students working in groups to assign grades - formative as well as summative - to themes. The Lagana study was different.

Evaluation Studies

In a 15-week study, Lagana (1972) investigated the effect of individual and peer evaluation on the writing ability of 60 tenth grade students. The researcher used multiple measures to collect and organise a variety of objective and subjective post-test data. Students were divided into groups of 5 on a permanent basis. Group work came late - during phase 3 of a 4 phase programme. He, like Pierson, used diagnostic data to balance strengths and weaknesses within groups, and pupils responded to papers but did not assign grades. However, groups had more control of the composition process, selecting objectives and deciding on writing tasks. Both the teacher and the student assessed individual and group progress; the awarding of a final course grade was a joint decision. A manual was provided for the course to acquaint both teacher and student with the logistics of each phase, roles and objectives. Finally, self evaluation preceded peer evaluation.

Lagana found a significant difference between the mean essay post-test score of the experimental group (n=30) and the control group (n=30). The experimental group was superior on the dimensions of organisation, critical thinking and appropriateness; the control group, who followed the school method, was superior in conventions.

Subjective and affective data indicated a positive response to the learning materials, group work, and the relative freedom from teacher control.

Ford (1973) wanted to discover the effect of peer editing and grading of themes on the grammar usage and writing skills of students. The sample comprised 50 freshmen students at a large state university in the Mid-western United States. The experimental group (n=25) edited and graded 7 themes during the 18 week programme. The compositions of the control group were assessed by their instructors. Students were administered a grammar and an essay test as pre-test and post-test measures. Ford found that the experimental group made significantly higher gains in grammar and usage, and writing ability. However, Ford had controlled for a number of variables including sex and age, but found no significant difference between pre-test and post-test scores.

Sager (1975) studied the effect of peer evaluation on the writing ability of 83 grade 6 students in Boston, Massachusetts. Like Pierson, she controlled for pre-experimental differences in intelligence, sex and writing ability. In the Sager study, groups used a rating scale to respond to their own work and that of their peers. In phase one, students first studied the characteristics of the scale through the use of special materials prepared by the researcher. Pupils then rated their own papers. In phase two, students responded to scripts in groups. For peer evaluation, the sequence was as follows. First, pupils rated for vocabulary, next for structure, and then for

both. During the final stage, students rated scripts for all four qualities. At first, they worked for consensus, but later rated independently.

Sager found a significant difference between the writing of the experimental group and the control group. The former showed more growth in overall ability, content, organisation, and structure. Interaction of sex did not affect performance, but girls did better than boys in overall writing ability. Finally, the experimental group showed great proficiency in using the scale with a reliability ranging from .96 to .99. The reliability of the scale when used by teachers ranged from .87 to .97.

In a sense, the study of Harris (1986) can serve to tie together the different strands of the discussion of collaborative learning. Harris conducted a study to ascertain the most effective strategies for collaborative learning. His first research question was the order of self and peer evaluation as pedagogical techniques: Should students work alone before working in groups? Which strategy was the more effective for revising composition? His second research question was centered on the mechanics of group work. Harris also investigated the best ways to set up groups, devise evaluation forms, and make arrangements for group work.

The sample was drawn from Advanced Placement students reading grade 12 English in a North Caroline public high school. The control

group of 43 students used peer evaluation techniques for a school year; the experimental group of 33 students used self evaluation for one semester and peer evaluation for the second half of the year. The control group responded to 427 papers, while the control group evaluated 129 scripts. Rough drafts, evaluation forms, and final essays were analysed and students responded to a questionnaire. Papers were ranked as showing strong, moderate, or little or no revision.

Harris found that peer evaluation made for stronger revision than self evaluation, and that self and peer evaluation were almost equal in moderate revision. The remarkable thing was that 279 of the 697 papers, 40% were not affected by either self or peer evaluation, as they showed no evidence of revision. He opined that pupils opted not to revise, or could not act on suggestions because the feedback forms, or the feedback itself might not have been helpful. Lamberg (1980) defines feedback as "any information on a performance which affects that performance" (p.63). Thus, following Lamberg, these students essentially received no feedback.

Data were collected from a questionnaire concerned with issues in peer evaluation. On the question of whether evaluators should be anonymous, students generally preferred to be anonymous when they were the evaluators, but wanted to know who their evaluators were when they were being evaluated. Forty of the 71 students who responded thought their writing had improved. Thirty-nine of the 70 thought that personal bias had affected the grades they had received.

Finally, students preferred their work to be graded by a friend, or a peer in another class taught by the same teacher, but not by a peer from a class taught by another teacher, or students in a lower grade.

Harris also found that the best evaluation forms had the following qualities. They were comprehensive, covering all areas of the composition. There was a section for the writer to point the collaborator to areas where the writer wanted specific feedback. Pupils preferred at least three raters with the names of the evaluators on the forms. Students, they felt, should respond as editors, and not give subjective comments they could not support with evidence from the text. Finally, the evaluator should have space to make specific suggestions.

Summary

To sum up, practising teachers at all levels - elementary, secondary and college - have found collaborative learning useful for a wide range of purposes. Empirical research has been generally positive: significant results have been evidenced in studies at all levels. Some studies, of course, show no significant effect of group work on student learning, but this seems to be partly a function of design of the study. It is also a function of the length of the project, since growth in writing takes time. However, it must be noted that the Pierson study ran for a year, but showed no significant difference in writing ability between groups.

Later, studies on revision, as illustrated above, have suggested that collaborative learning made for better revision and an improved writing product. The best strategies and procedures seem to be in place to make the method a breakthrough in easing the teacher load, and making writing for other audiences a reality for the student.

Questions remain: Are we as teachers willing to share our traditional monopoly of power over evaluation with students? If yes, what kind of evaluation - summative or formative? And if students can have an input in grade assignment, how much weight can such an assessment carry? Other questions are whether teachers are willing to put in the time required to learn effective collaborative strategies, and whether governments and other agencies are willing to take on the challenge. In the final section, the discussion is put in the context of student growth and development.

Language Development

Gere (1981) discusses the influence of two psychologists, Piaget and Vygotsky, on developmental theory. The Piagetian theory of maturation has been mentioned elsewhere: the child moves from egocentric speech for self to decentered speech for others.

Gere characterises this view of language as asocial since its proponents see language starting from within the individual and moving outwards. This Cartesian epistemology has resulted in a

particular school of thought in which writing is seen as a solo performance. It is reflected in the metaphors in current use - brainstorming, writer's block, and the image of a locked box.

It has been translated into polar and dichotomous thinking: writer-based versus reader-based prose (Flower), reflexive versus extensive thinking (Emig), expressive versus transactional writing (Britton, et al). Gere opines that this view of language development is narrow, and contrasts it with the Vygotskian model.

Vigotsky views individual language as internalised social language. The origins of language lies outside, not inside the individual; it is social not asocial. Gere comments:

> "Instead of being a transition from asocial to social language, egocentric or inner speech is a _continuation_ (italics added) of socially and environmentally oriented language development." (p.81)

Gere points out how recent research is bringing these seemingly opposing views together: language development is individual and social, and there is a continual dialectic between the two.

Group work becomes increasingly important if individual language is social language internalised. Through small groups "individuals engage in concrete social interaction...explainable in terms of small group dynamics and communicative practices" (p.60). Group work is not a means to an end - Piaget's movement from dialogue to monologue - but an end in itself - the centre of the writing act. Bruffee (1984) shares this view of thought as internalised conversation, and writing as that conversation re-externalised. Thus, Gere sees speech and writing as functionally complementary.

It is against this philosophical background that the contributions of Dixon and Moffett, and the developmental studies of the Britton and Wilkinson teams are examined. First, the report of Dixon on the Dartmouth seminar is treated. Then the contributions of Moffett are discussed. After that, the research sponsored by the Schools Council at the London Institute of Education is explored, and the Crediton research examined.

Personal Growth

The historic Dartmouth seminar involving scholars from both sides of the Atlantic resulted in the reconceptualising of English language teaching. Dixon (1965) explains the main lines of the personal growth approach to language which effectively modified - if not replaced - the skills and heritage models in use at that time. The emphasis was on new ways of knowing in a person-centred curriculum which connected the logical with the psychological.

The new emphasis was on whole language - "language in use from day to day" (p.8). Put in operational terms, language was defined by use: the corpus of activities to which language was put. The new emphasis was on personal experience, and the student's own language, however idiosyncratic. The view was that pupils use language to make sense of the world: to recall experience; clarify it and shape it; assimilate and accommodate it; and share it in various ways. They do so through talk and drama, and through writing and reading.

With such a curriculum approach, the teacher's role is also

redefined. She becomes ideally a trusted adult, conscious of her role as facilitator, engaging her pupils in dialogue, encouraging them to teach one another by monitoring the kind of talk they do as well as how they do it. She is conscious of the function of drama.

Drama should open up "to the inarticulate and illiterate that engagement with experience on which literature rests" (p.38). Because drama is primary, talk and gesture and movement will harmonise as pupils take on dramatic roles in encounters with new situations. In such a setting, development would be from the literal to the symbolic; from the simple to the complex; from pupils improvising language to students interpreting their own scripts.

From 12 to 18 children would be changing roles, seeing situations from many angles or perspectives, writing their own plays, exploring their own work - all in preparation for leaving their own world to enter the world of adult literature. Relative to literature, development is from the literal to the symbolic; from the symbolic representing actions of fiction steeped in myth and fairy tales to new ways of interpreting reality; from reading aloud to silent reading.

Following Piaget, the view of the seminar seemed to be that writing was a movement, a progression from collaborative enterprise to the solo performance of monologue. Thus, Dixon: "to write then is to move from the social and shared work to an opportunity for private and individual work" (p.44). Moffett writes:

> "To ask a student to write is to ask him to make all the adjustment between dialogue and monologue that I have been describing... The most critical adjustment one makes is to relinquish collaborative discourse, with its reciprocal prompting and cognitive co-operation, and to do it alone." (p.87)

Dixon suggests that talk and drama should be incorporated into writing and reading. Discussion in the form of exploratory talk should precede written assignments.

Dixon's book was a seminal work. The new concepts were built on, modified and adapted by theorists and researchers from Moffett to Wilkinson. One criticism was that the seminar in its emphasis on self over-emphasised the power of the individual. Language for others was neglected - language "to inform, convince, persuade, report, invite, order, request, instruct ....It's a large body of language to neglect" (Dixon, 1975, p.123). Both the individual and social aspects of language were stressed by NATE, The National Association of Teachers of English (1973).

Teaching and learning were conceived of as having both a vertical context and a horizontal context. The model stressed interaction with others in the widening contexts of family, social, ethnic or national group, and culture on the one hand; and growth upwards through widening conceptions of distance and levels of abstraction on the other. The work of Moffett is in the tradition of Dartmouth.

The Contribution of Moffett

Moffett (1968 and 1983), in his theory of discourse for language teaching K-12, expanded the main lines described by Dixon. His developmental sequence has been discussed elsewhere. He was among the first to redefine the traditional categories. For those he substituted recording, reporting, generalising, and theorising. Different levels of the taxonomy calls for new organising principles: the sequence is from chronologic through analogic to tautologic.

Like Dixon, Moffett calls drama "the matrix of all language activity" (p.61). He also values the connection between speech and drama. The development of speech is from soliloquy to dialogue to monologue. Egocentric speech leads to decentered social communication. Dialogue potentially becomes dialectic as pupils make meaning together. Monologue or ordered extensive speech is the bridge between speech and writing.

Moffett suggests similar teaching methods to Dixon: students improvising, performing, writing and analysing their own scripts; recording and analysing those of mature playwrights. Development of the narrative is via the following sequence: interior and dramatic monologue, to letters, diaries, autobiography, and memoirs to biography.

The Contribution of Britton

The London research was in the same tradition. As the basis of empirical research on writing across the curriculum, Britton took Moffett's categories further. His spectrum of discourse was

organised around audience and function.

For the Britton team, the expressive mode is the matrix of discourse. Development is in two directions - transactional and poetic. For Britton, the expressive is the matrix of discourse. This is essentially drama as discussed by Dixon and Moffett. From expressive language which is close to the self and for the self, growth is in two directions: the transactional (language to get things done), and the poetic (language as art). Users take on the role of participant or spectator. The latter concept was discussed at Dartmouth: "the sense of the role of spectator came to define literature" (Dixon, p.58). Literature was not confined to books, but included film and television, the pupil's personal writing and spoken narrative.

The transactional categories are record, report, generalised narrative/descriptive information, low analogic, through speculative to the tautologic. The audience categories are the writer to self, teacher, wider known audience, unknown audience. The teacher as audience is sub-divided: writer to trusted adult, pupil to teacher - student/teacher dialogue, pupil to teacher, particular relationship, and pupil to teacher as examiner. The wider audience has three levels: expert to layman, child to peer, group worker to working group.

Moffett's work then tied theory to practice. It has been criticised as being too cognitive, but if a curriculum springs from

dramatic activity and dialogue, then affective and moral issues may become part and parcel of such activity and dialogue. Nevertheless, Moffett's work stressed cognitive growth almost to the exclusion of other dimensions. The work of the London school treated the cognitive as well. It was left to the Crediton researchers to build a multi-dimensional concept of growth incorporating whole language with the whole person.

The Wilkinson Scales

The Wilkinson scales allow for a comprehensive description of writing from word level to discourse level. The team felt the Britton model has "grave limitations" and "is crude in that it gives a single procrustean description to each piece of writing ... in cognitive terms only" (p.223). One strength of the Wilkinson scales is that there are four levels of development - cognitive, affective, moral, and stylistic. The focus is on the child "as a communicating being"; emphasising both "communicating" and "being" (p.223).

Another strength is that they are concerned with diagnosis and formative assessment and not essentially to assign grades. A response to composition then is on many dimensions, not one; and at many levels - word, sentence and text. Following Piaget, the team distinguish between concrete operations (7-11) and formal operations (11-16). For example, describing and interpreting are concrete; generalising and speculating are formal and abstract.

The researchers in the Crediton Project analysed samples of writing from 150 children aged 7+, 10+ and 13+. Pupils responded to the same tasks under normal classroom conditions. The influence of the London research was evident. The tasks were classified by function and audience. The four types of audience were pupil to trusted adult, peer, expert to layman, and wider audience.

<u>The Cognitive Model</u>

The cognitive taxonomy has four dimensions: describing, interpreting, generalising and speculating. Each dimension has its own growth scheme or levels of development. For instance, the scheme for describing is labelling, naming, partial information, recording and reporting. The last two categories are drawn from Moffett; the difference is that the first two push the discussion back to the word and sentence levels. But the two stage level of abstracting are essentially the same.

The interpreting category is new. Wilkinson, like Britton, considered Moffett's leap from reporting to generalising too wide. Britton's team inserted "generalised narrative", and divided generalising into "low-level analogic" and "analogic". The Wilkinson team used interpreting as a bridge or transition between the concrete and the abstract.

After this, pupils should be able to classify, and use a different principle of order. Again, starting at the word level - abstracting -

the student summarises evaluates, reflects on, and classifies information. Finally, the student uses hypotheses. At first, three levels of competence are possible. Hypotheses may be irrelevant, relevant but inadequate, and inadequate. Later, the pupil extrapolates and explores and finally theorises - Piaget's hypothetico-deductive use of language.

The Bloom-like classicatory scheme allows the teacher to discover where a student is, and make certain pedagogical interventions to take the student through the various levels. The model is a teaching rather than an assessment instrument in the traditional sense.

The Affective Model

This is the first move beyond the cognitive. The measuring of attitude is problematic. Perhaps the most well known scheme is that of Krathwohl (1974) - receiving, responding, valuing , organising, and characterising. Wilkinson uses three broad dimensions of growth. These are the self, others and reality; but there is no attempt at a hierarchy or classification scheme in the sense of Krathwohl's above.

The movement is from the centre outwards: from self awareness to awareness of others. Another facet of growth is responding to the environment, the physical world, and finally "the inter-engagement of reality and imagination" (p.85).

The broad indicators of maturity are as follows. First, the writer deals with his own emotion: he expresses it, evaluates it, finds reasons for it. The writer is concerned with images of self. Another strand of growth is towards people outside the self: the writer is aware of others, gives them separate identities through quotations and the use of dialogue, analyses their comments, and establishes their character in context. Still another aspect of growth is in the concept of audience; there is a movement from context bound language to using context free prose. Yet another growth strand is the development of an awareness of the environment. From making a non-response to the environment by ignoring it, the writer starts to describe his physical surroundings - to respond to it, and use it to achieve an effect in his work. The final strand of growth is from the world of the imagination to that of reality.

The Moral Model

The analysis of moral thinking is problematic: Should it stop at describing the kind of judgements students make? It is true that pupils will catch ideas of morality from the world, but should morality be taught? Certainly, teachers indirectly influence moral behavior through the selection of reading texts for literature. If a student's "morality" is unacceptable, how can or must the teacher intervene? And by what criteria are we to judge acceptability? However, perhaps the model is not ultimately concerned with these issues, and even if it is descriptive only and not perscriptive, the model does help teachers to understand the whole student.

Growth is perceived as moving from the self outward: from self-gratification or anomy; through the effect of our actions on others resulting in rewards or punishment, heteronomy; to socionomy in different forms, considerations of status and the rule of laws; to autonomy, judgement using the criteria of human and individual rights.

The Stylistic Model

In what sense does style develop? Wilkinson defines style as "the result of a series of choices made to diverge or not to diverge from the norm represented by the sentence" (p.41). The writer makes choices to the end of effective communication. The range of choices will include syntax, verbal competence, structure, cohesion, reader awareness, appropriateness of discourse, and effectiveness. Space does not allow an extensive description here. But syntax will move from the simple to the complex; verbal competence from the literal to the metaphorical, and from the general to the particular; organisation from mere juxtaposition of elements to control; cohesion to a web of ties to establish relationships within the sentence; appropriateness will be adjusted to the field of discourse; and effectiveness to the realisation of the writer's purpose.

Summary

The work of Wilkinson and his colleagues has given the teacher additional ways of responding to student writing. Evaluation now moves beyond the cognitive to the affective and the moral. How the

teacher uses the non-cognitive data seems to be problematic. Are teachers of English competent to go beyond general statements to make judgements about such matters? Would teachers refer "problem" pupils to the school or guidance counsellor? What treatment follows teacher diagnosis of compositions on those dimensions?

Because of the nature of dialogue, collaborative learning or group work certainly can help teachers to know the whole student in a way that writing perhaps cannot. Indeed, the Crediton project seems to make the necessity for group work more important. Add to that notions that the group work provides a way for pupils to teach themselves, provides alternative audiences for student writers, is at the heart of writing as a social activity, and is a necessary stage for those who see writing as an individual activity, and the importance of this methodology is evident.

CHAPTER FOUR

THE PILOT STUDY

Abstract

A pilot study was conducted during the first semester of the 1987-1988 school year. The purpose of the preliminary investigation was to test the design and procedures, and assess responses to the programme. The hypothesis to be tested was that peer evaluation would improve the writing quality of a sample of grade twelve students in two public high schools. Pupils in four classes were taught by four different teachers with one Experimental and one Control group at each school. The Experimental group used a programme designed by the researcher to teach students how to use a scale to rate their own papers and those of their peers. The Control group followed the regular English programme of the school, and their compositions were graded by the teacher. The subjects wrote one of two essays, and STEP tests in grammar and mechanics of writing as pre-tests. Students wrote the other essay and alternate forms of the STEP tests as post-tests. The Experimental group also responded to an attitude questionnaire, at the end of the programme. There was a significant difference in writing quality between the two groups at the .05 level of significance. Covariates of grammar and mechanics were significant at the .07, and .10 levels of significance respectively. The rating scale, constructed by the researcher, was validated by three experienced teachers. Inter-rater reliability co-efficients were .89 for total score, and a range of .79 to .84

for sub-scores. Finally, an analysis of subjective data suggested a positive attitude toward writing.

Subjects

The sample comprised four high school grade twelve classes taught by four female teachers in two high schools. These were the only two public high schools in the school district of St.Thomas/ St.John in the United States Virgin Islands. Students at both schools covered the range of socio-economic levels. One school, Charlotte Amalie High School (CAHS), was a city school with a long tradition; the other school, Ivana Eudora Kean High School (IEKHS), was out of town and generally perceived as less advantaged. Further, this school operated under certain physical constraints which had affected its accreditation. The enrollment at the two schools at the time of the experiment in October 1987 was 3,123 pupils. The sample of 69 students (12%) was drawn from a grade twelve public high school population of 557.

Two regular grade twelve classes at IEKHS took part in the study. These pupils were selected at random and comprised a sample of 44 (26%) from a grade twelve school population of 170. Two junior composition classes at CAHS participated in the study. Junior composition was an elective, so the students were volunteers. The small sample of 25 students (6%) was drawn from a grade 12 population of 387 students. The Experimental group at IEKHS comprised 21 students and the Control group 23 students. At CAHS, the Experimental group

was composed of 10 students and the Control group 15 students.

Procedure

Prior to the study, preliminary data were collected. The chairs of the English departments made student essays on file available to the researcher. This material was analysed for writing quality on the dimensions of content, organisation, structure and wording. In addition to this, a number of grade twelve classes wrote two compositions on topics suggested by the researcher.

Grade twelve English curricula were obtained from both schools, and the knowledge, objectives and skills to be taught were incorporated into the programme of study designed for the investigation. Finally, the researcher held meetings with English faculty at both schools to explain the purpose, characteristics and importance of the study, and exchange information.

The Experimental group followed a programme designed by the researcher. Pupils worked in groups of various sizes to edit, respond to, and rate their own papers, compositions of their peers, or samples in student booklets. Lessons and exercises were designed to teach them the components of the scale. Other lessons involved student use of the scale to respond to compositions.

The Control group teachers taught the same objectives, but followed the grade twelve English curriculum of the school. Students

also worked in groups, but their compositions were graded by their teachers. For the Experimental group, one essay, a working paper, went through various drafts and was evaluated by the group. The essay at the end of the unit, a test essay, was graded by the teacher using the same rating scale. This was the basic plan for all units.

Lesson Plans: Experimental Group

During the first week, a writing sample and objective tests in grammar and mechanics were collected from both groups. After this, preliminary work was done to develop a sense of audience. Students wrote and read letters of introduction to their peers within the classes. Then, the Experimental group at each school wrote letters of introduction to one another. They worked in pairs to edit these letters, but once the flow of communication started, pupils only shared letters if they wished. The inter-school exchange of letters continued throughout the programme; in all children exchanged four letters.

The programme really started from the second week. The unit on content was composed of four phases. During the first phase, students were taught the criteria for papers rated high, middle and low. At the end of phase one, students handed in an out-of-class composition which their teachers held in storage; these papers were evaluated during phase four. During the second phase, students learned to distinguish between papers rated 1 and 2, and papers rated 4 and 5. During the third phase, students rated sample papers

in groups. Students made oral and written comments on each paper in workshop fashion. Writers could then rewrite papers to improve their scores. A rating for each paper was determined by the group, and the teacher recorded the grade in his markbook. These grades became part of the final grade for the course. Chapter seven presents detailed examples of lesson plans, composition of the student handbooks, and teacher manuals.

The second unit was on organisation. The four phase sequence described above was common to all units. Students internalised the criteria, responded to sample compositions, and rated their own papers. They rated papers for two qualities: first for organisation and then for content. For the unit on structure, pupils were encouraged to edit essays written earlier in the course for grammatical and mechanical errors. Students rated papers for structure only. The final unit was on wording. During the rating phase, students rated scripts for wording and then for structure. Finally, they rated papers for all four qualities.

The experiment ended with a series of post-tests conducted during the final week of the programme. Students from both groups wrote alternate forms of the objective tests in grammar, mechanics, and a final composition. Students in the Experimental group also responded to a take home questionnaire.

Activities: Control Group

During this period, the Control group wrote the same pre-tests as the Experimental group. Then teachers reviewed the writing process, and students did writing awareness exercises.

Students did a unit on content. They wrote a description of a mystery object and shared descriptions. They worked through the various phases of the writing process, writing and rewriting drafts. They built, displayed and responded to collages and completed exercises on imagery to sharpen their observation skills. For the unit on organisation, students completed exercises, and analysed writing samples using a proof-reading guide. They also wrote and shared compositions after examining models for atmosphere. They did free writing in their journals and studied transitional devices.

Unit three was on structure. Students continued freewriting exercises. Then they wrote a skit in dialect, rewrote it in standard English, and then compared the two products. Work continued on the use of transitional devices and exercises to improve student composition skills. Students then did sentence-combining exercises and a test on vocabulary. The final unit was on wording. Students wrote descriptions of persons, and participated in palm-reading exercises to sharpen their observation skills. They wrote letters and descriptions of their classmates. They did exercises in grammar and videotyped collages.

During the last week, students wrote alternate forms of the objectives tests in grammar, mechanics, and a writing sample as post-tests.

Controls

The researcher used a pre-test/post-test control group design with random assignment as detailed in Campbell and Stanley (1966), and Borg and Gall (1983).

Randomisation

At both schools, pupils were assigned to regular classes by computer at the beginning of the school year. The two classes at IEKHS were regular classes and thus randomisation was assured. However, the classes at CAHS were junior composition classes, and junior composition was an elective. So, selection was a factor that could probably have influenced the internal validity of the pilot study.

History

Attempts were made to schedule classes at the same time to reduce the effect of external events on the results of the experiment. At IEKHS, both classes met at the same time, so that events that affected one group would, presumably, have affected the other group. However, because of scheduling constraints, the junior composition classes at CAHS met at different times of the day.

Testing

Students responded to standardised tests administered under formal conditions by the testing officer of the local university, assisted by the classroom teachers. The following efforts were made

to minimise the influence of testing. First, the class teachers were present during each administration of the tests and served as proctors. Second, there were attempts to make conditions as normal as possible in order to reduce test anxiety.

Third, post-testing was incorporated into the normal examination period of the schools. In fact, the post-tests for the programme were the school examinations for the four classes, and the compositions and exercises done during the study were the components of the grade for the course. Pre-test scores also counted in the determination of the final grade. Fourth, the technique of "shuffling" was used to reduce rater bias. Two composition instruments were designed by the researcher. Students wrote both essays - one as a pre-test and the other as a post-test. However, half the students wrote one essay as a pre-test while the other half wrote the other. For the post-test, the situation was reversed; students wrote the essay they had not written before.

Fifth, raters were experienced teachers connected with the university but who had experience in teaching at the high school level. The raters did not use the rating scale designed for the programme but used a four-point holistic scale, and the criteria they usually applied to grading their own students' papers. This helped to control for a possible incestuous relationship as another mode of scoring helped to validate the results.

Mortality

For various reasons, only partial data on some students were available. First, at the beginning of the school year there was a movement of students from one class to another in order to resolve scheduling conflicts. Because of this, some of the students entered the programme late. These wrote the post-tests only. Second, post-testing was done just before schools closed for the Christmas break, and some students were absent. Consequently for statistical purposes, data were only analysed for students who wrote the entire battery of pre- and post-tests.

Other Effects

Both programme and control group students and teachers were aware that they were taking part in the testing of a new programme. The importance and the relevance of the study were explained to both groups. This was an attempt to control for a possible Hawthorne effect. Since the programme classes used specially prepared materials and wrote letters to one another, this special attention could, by itself, have had some positive effect on student performance. The programme materials were to be treated as confidential and were not to be shared with other teachers or students. The Control group teachers displayed much enthusiasm. One teacher kept the researcher informed about class progress and was eager to discuss techniques and methodologies. Here, the John Henry effect seemed to be operating.

Data Analysis

The researcher used analysis of covariance (ANCOVA) procedures. The dependent variable was writing ability as measured by post-test composition scores. The independent variable was teaching method. The controls were three covariates: - a pre-test composition, the STEP test of English Expression and the STEP Mechanics of writing test. Separate analyses were conducted for each covariate.

Results

TABLE 1

MEANS AND STANDARD DEVIATIONS OF PRE-TEST SCORES

	Experimental Group N (21)		Control Group N (21)	
	Mean	S.D.	Mean	S.D.
Composition[a]	4.19	1.83	4.24	2.12
Grammar[b]	18.10	5.80	16.95	4.79
Mechanics[c]	22.33	5.94	19.00	7.53

[a]Maximum Total Score = 10
[b]Maximum Total Score = 40
[c]Maximum Total Score = 45

Table 1 presents the means and standard deviations of the pre-test scores for composition, grammar and mechanics. The Control group mean was higher in composition, but the Experimental group excelled

in grammar and mechanics. All groups were equated statistically, at the end of the programme to account for pre-experimental differences on the variables of composition, grammar and mechanics.

Table 2 presents the adjusted mean scores showing the effect of the programme on the writing quality of both groups.

TABLE 2

EFFECT OF PROGRAMME ON QUALITY OF POST-TEST COMPOSITION
(Unadjusted[a] Means in Parentheses)

Group		Difference	F	P
Experimental (N = 21)	Control (N = 21)			
6.00[b]	4.42	1.58	4.06	.0481
(6.00)	(4.43)	(1.57)		

[a]Adjusted for pre-experimental differences in writing ability.
[b]Maximum Score = 10
Note: Source table is in Appendix A.

The Ancova for the effect of the programme using the pre-test composition covariate showed a significant difference between Experimental and Control groups as predicted, $F\ (1,39) = 4.06$, $p < .05$. The pre-test composition mean score showed a difference of .05 favouring the Control group M = 4.24) over the Experimental group (M = 4.19). At the end of the programme, the Experimental group had shown more growth, with an unadjusted mean difference score of 1.81 compared to

a mere .19 for the Control group. The post-test Control group mean was adjusted downward since that group had been slightly better in writing initially. The gap between the two groups of 1.57 points on the dependent variable was increased in favour of the Experimental group to 1.58 which was statistically significant.

Table 3 presents the adjusted mean scores and the effect of previous knowledge of grammar on the writing ability of the two groups.

TABLE 3

EFFECT OF PRE-TEST GRAMMAR SCORES ON POST-TEST COMPOSITION SCORES

(Unadjusted[a] Means in Parentheses)

Group		Difference	F	P
Experimental (N = 21)	Control (N = 21)			
5.89	4.54	1.35	3.40	.0694
(6.00)	(4.43)	(1.57)		

The Ancova analysis for the grammar covariate showed a significant difference between the Experimental and Control groups, $F\ (1,39) = 3.40$, $p < .07$. The Experimental group was initially better in knowledge of grammar (M = 18.10) compared to the Control group (*M* = 18.10) compared to the Control group (*M* = 16.95), a difference of 1.15 score points. The dependent mean was consequently adjusted downwards in

favour of the Control group with an adjusted mean difference score of 1.35. This difference was not as significant as that for the pre-test composition covariate. The influence of the pre-test on the post-test depends on the size of the initial difference between the two groups, and the strength of the relationship.

Table 4 presents the adjusted means scores and the effect of previous knowledge of grammar on post-test composition scores.

TABLE 4

EFFECT OF THE PRE-TEST MECHANICS SCORES ON POST-TEST COMPOSITION SCORES

(Unadjusted Means in Parentheses)

Group		Difference	F	P
Experimental (N = 21)	Control (N = 21)			
5.88 (6.00)	4.55 (4.43)	1.33 (1.57)	2.71	.10

The Ancova analysis for the mechanics covariate showed a significant difference between the Experimental and Control groups but not as strong as the grammar or composition covariates. Initially, the difference score of the covariate mean was 3.33 in favour of the Experimental group. Consequently, the dependent mean for the Experimental group was adjusted downward to a mean of 5.88; whereas that of the Control group was adjusted upwards to a mean of 4.55. The adjusted mean difference of 1.33 was significant at the .10 level.

Grammar

Table 5 presents the adjusted mean scores for the two groups, and the effect of the programme on grammar knowledge and use in those groups.

TABLE 5

EFFECT OF THE PROGRAMME ON POST-TEST GRAMMAR SCORES

(Unadjusted Means in Parentheses)

Group		Difference	F	P
Experimental (N = 22)	Control (N = 25)			
18.25 (18.14)	16.62 (16.72)	1.63 (1.42)	1.60	.2098

There was no significant difference in the post-test grammar score means of the two groups. An examination of the means, however, suggested that the Experimental group mean between pre-test ($\underline{M}$ = 18.25) had increased by 1.25 score points after the means had been adjusted for pre-experimental differences. But the Control group students, who initially scored higher, had a lower post-test adjusted mean score. The pre-test mean for this group ($\underline{M}$ = 17.36) was reduced by .74 score points ($\underline{M}$ = 16.62). Thus, the adjusted post-test difference score mean of 1.63 score points in favour of the Experimental group was very much higher than the pre-test mean difference score of .36 in favour of the Control group.

The data were further analysed to gauge whether the post-test grammar and mechanics scores had increased by any statistically significant degree between pre-test and post-test. The scores of all students completing both the pre-test and post-tests were considered. So, the size of the sample increased from 21 to 26 for the Experimental group, and from 21 to 25 for the Control group, a difference in size for the total sample of 9 students. There was no significant difference revealed by the analysis.

Validating the Rating Scale

The researcher checked the reliability of the rating scale when used by teachers. Three teachers of English were asked to participate in a validation exercise. Two of the teachers had at least ten years experience at the high school level; the other tutored English at the college level.

The group met in a conference setting where the raters were briefed and trained. They then rated 12 scripts using a 5-point holistic scale. The papers were a random sample drawn from pre- and post-test essays collected during the project. Each script was marked three times - once by each rater. Cronbach's alpha was used to calculate the inter-reliability of the scores.

Table 6 presents the inter-correlation co-efficients for three teachers using the holistic scale. For a full description of

procedures used, see chapter eight on the analysis of the main study data.

TABLE 6

INTER-RATER RELIABILITY CO-EFFICIENTS FOR 3 RATERS

Category	Co-efficients	P
Total Score	.89	.01
Content	.74	.05
Organisation	.84	.01
Structure	.79	.05

The inter-rater reliability for adults using the scale was significant for total score (R = .89, p < .01), for content (R = 74, p.05), for organisation (R =,=.84,p < .01), structure (R = .79, < .05), and wording (R = .84,p, < .01).

Discussion

Reader agreement has been an issue in the evaluation of the direct assessment of writing for many years on both sides of the Atlantic. Important research reports - Godshalk and Swineford (1966); Diederich (1974); and Britton (1966) have done much to solve the reliability problem, or at least to make it less severe. But what level of rater agreement is satisfactory or acceptable for essays?

The value of the reliability co-efficient tends to be a function of the number of tasks and the number of raters. Breland (1983) reviewed 17 studies done between 1935 and 1983 for reader reliability estimates. The method of scoring was holistic, analytic and atomistic. Reliabilities ranged from .80 to .94. There was a median of .70 for the three types of scoring taken together. These coefficients were for one task scored by two raters.

Breland (1983) also found that, in six studies, reporting reliability co-efficients for analytic sub-scales that the reliability co-efficient for one task by two raters ranged from .52 to .83. Diederich (1974) found a reliability of .80 quite acceptable for one task scored by two raters.

The reliability of .89 for the total essay reported in this research report then seems to be in line with what is obtained on this type of assessment. All reliability co-efficients ranged from .74 for content to .89 for total score. Table 7 showes that all the reliabilities were significant either at the .05 level or the .01 level of significance.

<u>Attitudinal Data</u>

Table 7 presents a summary of ratings for a take home student questionnaire. The return ratio for the questionnaire was 70%. An analysis of the data follows.

TABLE 7

SUMMARY OF RATINGS FOR STUDENT QUESTIONNAIRE

Category	Ratings 1	2	3	4	5
Corresponding					
Writing	-	-	3	4	7
Receiving	1	1	2	2	9
Rating Papers	3	1	2	3	5
Working in Groups	1	-	2	4	7
Using Booklets	3	-	2	4	7
Writing Essays	2	2	1	5	4

Note: Number of questions distributed = 20.

Number returned = 14.

At the end of the study, the Experimental group responded to a take home questionnaire. The questionnaire was divided into three sections. In section one, pupils responded to items on a Likert-type scale with a low of 1 and a high of 5. They were asked how they liked writing and receiving letters, rating papers, working in groups, using booklets and writing assignments. In section two, students gave reasons for their response to each item in the spaces provided. In

section three, students responded to whether they thought their writing had improved and in what ways. Finally, the pupils commented on anything else they liked.

Writing Letters

The respondents enjoyed writing letters; there were no low ratings on this item. The eleven pupils responding, representing 79% of the 14 pupils, gave this category a high rating. Some of these students responded positively but did not give a reason. They used words like "enjoyed" and phrases like the "best part". Although these were positive reactions, they were not helpful beyond this point. Others gave general reasons. Student A's comment was typical of this group: It "helped me to expand my thoughts and things."

But others were more specific. These had fun expressing themselves to "new people" to get to know their "personality. However, Student C - a girl who rated this item 3 - had problems finding common ground with her correspondent and was extremely sensitive about writing to a boy: "It was hard writing to a boy and our interests were totally different." Student E, with self-effacing candour, found it "sort of fun", although she herself was "not a very interesting person." Student J found it to be a challenge, not knowing the person with whom she was corresponding. Both rated the item 4.

Receiving Letters

Nine students, 64% of the respondents, gave receiving letters the highest rating. Student C gave this item the lowest rating and commented, "I didn't really care to receive a letter from a guy." In general, the choice of words suggested excitement: "great!"; "special thrill!"; "the fun part!"; "delightful!" Student A's response was typical: "There was always a special thrill in receiving the letters because of the effort I put in mine, and their letters gave some ideas of what they thought of my letters." Feedback appeared to be extremely important to students. Thus, Student F remarked: "Getting a response from my letter makes me know someone reading my letter." Learning about the goals of other students, reading what other people had to say, and evaluating their responses were other reasons given for the high rating. For most, the experience was emotionally satisfying. Witness Student K: "It was very good to know someone (was) writing me a letter and *saying nice things in it*" (italics added).

Rating

The reaction to rating papers was more uneven. Four students gave it a low rating - three of these the lowest. On the other hand, eight students (57%) gave it a high rating - 5 the highest. Two students were lukewarm. This mixed reaction was to be expected.

First, peer response is difficult. Davis (1981) cautions that its "success and utility ... will depend on the effort that student

and teachers expend on it and the amount of guidance the student receives in the process" (p.94). He also warns that pupils need detailed guidance before they begin. Although the programme attempted to address those concerns, problems remained. Students were unaccustomed to the method. Student A noted: "I've never really comprehended the different techniques."

Apart from the novelty of the exercise, Student F admitted, "It's difficult rating your friend's paper." H found rating difficult without giving a reason; for L the difficulty was having to read the paper "at least two times or more to understand and then rate it." Student J was quite frank in expressing a concern no doubt shared by many: "I hate (italics added) rating papers because I did not like to give a person a grade, because I am not equipped to do so." The notion that the teacher was the sole dispenser of grades, and the student could not share in the process, rang familiar. A note of frustration was sounded by Student M who found "rating rather boring and unpleasantly unsatisfying."

To summarise, some students found it helpful seeing things the way their classmates saw them, learning from the mistakes of others, judging how creative they were, comparing responses, and discovering errors and correcting them. Despite the difficulties and concerns expressed, some students welcomed peer evaluation as a helpful learning experience.

Working in Groups

The related category of group work brought greater consensus. Student F was the only student to give this item the lowest rating, but the reason for this was essentially in favour of group work: "We get to share each others comments; then all of us decide on the rating."

The advantages of the procedure were well illustrated by student comments. First, group work matched the learning style of some students. These pupils enjoyed sharing their thoughts and had "a lot of fun". Second, other students found that group work provided opportunities for them to become better acquainted." Third, still others commented on the practical utility of the method: - sharing ideas, voicing opinions, hearing different views. Student E made the point well: "Working in groups was O.K. because you had three different opinions to go on to evaluate the writer's grade, instead of doing it alone and getting all of the blame for rating someone low." Such forces of classroom dynamics and the social pressures within the walls of the school were inhibiting factors and tended to limit student responses - limitations though which could be overcome by the nurturing of a climate of trust. Peer evaluation is sometimes perceived as tolerable at best and as extremely unpleasant at worst. Many pupils are comfortable when the responsibility is shared.

Using Booklets

Three students (D, G, and N) gave this item the lowest rating. Student D disliked the fact that student samples were unedited and presented as they came so that the class could edit in groups. G found the booklets confusing, and N "hated using the book because she found them "quite distasteful and a waste of time."

Other students, in contrast, found the booklets easy to read and carry around, thought the information useful, liked the fact that they were reading the work of their peers and felt the booklets helped them a lot.

Writing Assignments

Responses were spread out covering the range from 1 to 5. Nine pupils rated this item above average. However, Students G and J gave it the lowest rating. The reasons given for the ratings varied. Student G found the assignments "unnecessary". J found them "terrible" because she found writing extremely difficult and was unable to translate her ideas into words. L did not have much fun since she was always trying to get things right for the teacher. In contrast to this group, other pupils welcomed the chance "to open up" or "use their imagination", or put their thoughts in perspective.

Writing Improvement

Only 2 of the 14 students felt their writing had not improved. Many students felt their writing had improved at the micro-level.

For example, Student M was concerned that he still had troubled writing adverb clauses. But some others felt their writing had improved at the macro-level. These referred to larger elements such as paragraphing, organising and planning, focusing and elaborating. Generally, all 12 pupils who responded to this question offered positive encouragement and recommended the programme be introduced in other parts of the school

Summary

It seemed that the general response to the pilot study was positive. The difficulties expressed by students appeared to be inherent in the activity of rating papers. The concerns of some students about certain aspects of the booklets, and the need for some adjustment to the writing tasks assigned were considered at the revision stage of the study.

Conclusions and Implications

The results of the pilot study data suggest that peer evaluation may lead to improvement in the writing ability of Virgin Islands students. Thus, they tend to support the findings of Sager, Lagana and Ford who used samples from the mainland United States.

It appears that the unit on structure had some influence - though perhaps a tenuous influence - on writing ability. There

was not, it is to be noted, the traditional particle approach to the teaching of grammar, but an emphasis on sentence combining and organic correction. Peer editing for mechanics had the least effect on writing ability. What does all this mean?

It suggests that when students know what teachers are looking for and learn to apply those standards themselves, they may transfer this knowledge and use in their own writing. It suggests that certain approaches to the teaching of grammar can make a difference. However, the programme did not show that peer editing for mechanical errors would necessarily lead to an increase in writing ability or mechanics scores. Indeed, the approach of the Control group instructors seemed at least just as effective. It must be pointed out though that emphasis on mechanics was down-played, and students conferenced and responded to papers relative to the larger elements of writing.

The attitude of students in the Experimental group to collaborative learning was generally positive, although the method was not consonant with the learning styles of some students, and not without inherent difficulties for those comfortable with the method. The exchange of letters between schools seemed promising for developing in students a sense of audience. It appears that experienced teachers could use the scale reliably and consistently, and discussions with teachers of the programme suggested that students may be taught to rank order scripts on a scale of quality.

Changes: Main versus Pilot Study

In this section, specific differences between the pilot study and the main study are explained. After the first study was completed, all parties concerned took a hard look at the experiment to gauge its impact and make modifications where necessary. The investigator met with principals, English chairs, and teachers together to have a bird's eye view of the programme. These sessions were essentially negotiations to determine under what conditions the main study would be conducted. Finally, a series of meetings was set up with the teachers by themselves to discuss the programme in detail.

Schools in the USVI, in conjunction with the Department of Education, issue school-leaving diplomas. Schools are responsible for the constructing, administering, grading and reporting of tests, although they follow course outlines and curriculum guides approved by the Department. This is in contrast with what is obtained in Britain where different examination boards, usually connected to one of the universities, are responsible for examining and certifying students both at "O" level (now GCSE) and "A" level. Covering the grade twelve syllabus then is critical for teachers and students alike.

In addition to examination pressures, seniors especially during their final year, participate in such activities as introductory

nights, language arts showcases, and so on that compete for class time. Then there are the not so extraordinary activities like strikes and demonstrations which may occur at the start of a given school year.

Larger Issues

The first meeting was with the parties at CAHS. Both teachers were very enthusiastic about student response to the programme, but were very conscious of the time it took and the constraints on grade twelve students mentioned above. First, the chair pointed out that the research paper had to be done. After discussion, it was agreed to make it partly an out-of-school activity, but where lessons could be modified to serve both purposes these opportunities should be taken.

Second, there was the question of teacher contracts. Teachers urged that their workload be reduced so as to avoid extra preparation. Contracts stipulated the number of preparations per class, and the number of classes that should be assigned to teachers. The teachers involved did get the released time requested. Indeed, the co-operation given by the administration at both schools was encouraging.

The meeting at IEKHS addressed similar issues. There, it was decided to postpone the start of the programme until the next school year. The teacher who finally consented to teach the programme wanted more time. Teaching the programme was voluntary, and the

teacher assigned refused to participate unless certain conditions were met. Similar constraints of time were mentioned, but the following specific decisions were made: there were to be no inter-school letters, the programme should address as many modes of writing as possible, and lessons plans should be reduced to the main only.

It had always been planned to allow students to write for different audiences and purposes. However, the lack of inter-school communication meant that pupils could not have the opportunity to write for a relatively unknown audience - their peers outside school. We turn now to other changes.

Other Changes

Assignments

School syllabi are based on the traditional modes, and were designed to take pupils through description and narration, and through exposition and argument to the research paper. The pilot study treated assignments in the descriptive/narrative modes. Modifications to the programme were along the lines of the classification schemes suggested by theorists like Moffett, Britton and Wilkinson.

Britton (1966) rejects traditional classification schemes because of their limitations, while admitting that no scheme is going to be completely satisfactory. A classification scheme is necessary since writing is not a global quality, and people write to different audiences for different purposes.

Britton points out that both the categories and the rule structures for producing them are derived from the products of professional writers; that the writing product not the writing process is emphasised, that the system suggests that the four major activities match four separate and distinct mental activities existing somewhere in pure form. Besides, the categories are not equal since description and narration can fit into persuasion and exposition, so the categories are not discrete.

The limitations of the classical rhetorical modes are most glaring in the mode of description. Description hardly exists as a separate mode playing a supporting role for other universes of discourse. The term is difficult to define as well: description may be suggestive, literary, technical; specific or general.

Thus, one way to approach the topics treated in the study was by re-classifying them. The topics in the pilot study were descriptive/narrative in the traditional sense. In the Moffett classification scheme, most would be classified as drama - what's happening; or narrative - what happened; or to put it another way, recording and reporting.

In the main study, topics included but went beyond those categories. Some questions fell into the category of what Wilkinson calls interpreting; sometimes pupils had to give reasons. From one viewpoint, the pre- and post-test topics were the opposite poles of a possible growth scheme for the course. The one question

pointed to the interpreting of a personal experience; the other pointed to the use of description with a persuasive intent. For a further illustration of the topics and discussion of this issue, the reader is referred to chapter six.

Materials

The teacher's manual was modified in the light of teacher comment. Most lessons were reduced to the main line and only a limited amount of teacher supplementary material included. This physical adjustment in bulk seemed to reduce psychologically the amount of preparation required. The purposes of the supplementary materials were met through discussion and conferences.

Further, some of the illustrative material used in the pilot study, mainly found data collected before the study, was replaced by writing samples collected during the trial run of the programme. But this was not common to the teacher's manual, since the same changes were also made to the students booklets.

New lessons were added and end of unit test items introduced. There was also a change in the order of the presentation of the units. The original order was content, organisation, wording, and structure; the new order placed structure after organisation and before wording. Some teachers felt that the mechanical aspects of composition should come earlier. Finally, space for teacher comment was provided at the end of each lesson. Student booklets were modified to reflect the changes outlined.

Scales

Rating scales were also adapted. The writing programme treated writing on four dimensions, each described on a 5-point scale. At first, each point on the scale was described: scale criteria presented and illustrated with anchor papers. However, for efficiency, this system was modified.

The researcher felt that, once students knew the differences between high, middle and low papers, they could learn to make a judgement as to how high or low the paper was. Once these differences were internalised and anchor papers for the full scale selected, scales for 2 or 4 papers were not really necessary. How the system worked is illustrated in chapter seven. As a consequence, the only scale descriptions given were for high, middle, and low papers; however, all five points on the scale were illustrated with anchor papers.

Measures: Pre-tests

The same objective test instruments were used in the main study as in the pilot study. The composition probes were the same as well. However, attitudinal measures were added to supplement the cognitive data and get a more rounded picture of how pupils viewed the writing act. The Miller/Daly measure of writing apprehension was given as both a pre-test and post-test.

The Experimental group also had post-study measures that were

specific to that group. Pupils graded 5 papers for each of the 5 qualities taught. In addition, they were encouraged to write a letter to the researcher reacting to the programme, especially to group work and peer rating. This was in addition to the modified free response questionnaire used in the pilot study.

Grading

Objective tests were machine scored. Pilot study essays were graded by two experienced teachers, according to their own standards. The only directive was that they use a 5-point scale. They graded the papers holistically. Conversely, main study composition scripts were rated by teachers trained to use the same analytic scale used during the study. The scale had been validated after the pilot study as described in an earlier section.

All hypotheses in both studies were tested using analysis of variance techniques. In the pilot study, interaction effects were analysed but were not significant. For the main study, such interactions were not explored.

CHAPTER FIVE

RESEARCH DESIGN AND PROCEDURES

The main study was conducted at Charlotte Amalie High School from February to June 1987-1988, during the second semester. The study was conducted at Ivana Eudora Kean High School during the first semester of the 1988-1989 school year. The statement and justification of the research problem were presented in chapter one. In addition, an elaboration of the rationale for the study was presented in chapter three in the review of the literature.

The Sample

The sample comprised ninety-eight grade twelve pupils from two public high schools in the school district of St.Thomas and St.John in the United States Virgin Islands. Classes were classified as Experimental or Control groups. Fifty pupils were in the Experimental Group, and 48 in the Control group.

At Ivanna Eudora Kean (IEKHS), 50 students were selected: 25 in each group. These pupils represented a sample drawn from a population of 177 grade 12 students. The sample represented 28% of all grade 12 pupils at the school (Department of Education, 1988). Pupils were assigned to classes at random with the help of the School System, a computer programme distributed by the Associated Business Corporation, in California, the United States.

At Charlotte Amalie High School (CAHS), 48 students were selected: 25 in the Experimental Group, and 23 in the Control group. These were a random sample drawn from a grade 12 population of 398 pupils, in similar fashion to that at the other school. The sample represented 12% of grade 12 pupils at the school.

There was a total of 70 girls and 28 boys participating in the experiment: 36 girls and 12 boys from CAHS, and 34 girls and 16 boys from IEKHS. Most students were aged 17-18 years, with a low of 16 and a high of 20.

Experimental Group

The Experimental Group, comprising two grade 12 classes, used special instructional materials designed by the researcher. Pupils learned to use the four components of an analytic scale to evaluate writing for content, organisation, structure and wording. Four handbooks covering the material were distributed to each student - one at the start of each unit. Teachers also received four manuals with the teaching/learning materials.

Control Group

The Control group was composed of two grade 12 classes. These pupils studied the same components of composition as the Experimental group, but each teacher followed the curriculum and procedures outlined in the school's English programme. However, teachers in the Control group were given the scale description for each unit.

At one school (CAHS), the Experimental and Control groups met for 50 minute periods, 5 days a week for 13 weeks. At the other school, (IEKHS) pupils met for 45 minute periods, 5 days a week for 11 weeks. At CAHS, both classes met at the same time during the first period: 8.30 to 9.20 a.m. At the other school, the Experimental group met during period 3, 9.30 to 10.15 a.m; the Control group met during period 7, 12.30 to 1.15 p.m.

Personnel

Teacher A

This teacher, a male, taught the Experimental group at CAHS. He graduated from the World University, Puerto Rico in 1970 with a B.A. in English, and from the University of Illinois in 1980 with an M.S. in journalism.. He had over 17 years' teaching experience at the high school level, with some experience at the college level teaching English and journalism.

Teacher B

This teacher, a female, taught the Experimental group at IEKHS. She received a B.A. in English with a minor in education from the University of the Virgin Islands in 1985. She had been teaching high school English from 1986 to the present.

Teacher C

This female had wide experience teaching at both the college andhigh school levels - 8 years college and 6 years high school.

She read for a B.A. in reading at Boston, Massachusetts and graduated summa cum laude in 1974. She received her M.A. in reading from the same school in 1981.

Teacher D

This female, a graduate of the University of the Virgin Islands, received her B.A. in English with a minor in education in 1975. She also graduated from the University of Conneticut in 1978 with an M.A. in English. She had been teaching high school English for 11 years.

The Variables

Efforts were made to control variables that could possibly contaminate the results of the experiment. Campbell and Stanley (1966) distinguish between internal validity and external validity. Internal validity refers to the extent to which extraneous variables are controlled. These are factors outside the study that can possibly affect the treatment given to a particular cohort of students, and therefore influence the result of the study. Internal validity is an important issue since other researchers might want to replicate the study. External validity is an important consideration as well, since there must be some attempt, however tentative, to generalise findings from the sample to the experimentally accessible population, and eventually to the target population.

Snow (1974) distinguishes between the "systematic" and

"representative" design of experiments. He sees the traditional controls as in some ways artificial, unnatural, and generally lacking in generalisability. Snow urges that research reflects the "real life" environment - the ecology of the classroom. The assumptions are that children, as well as classroom environments are complex, and their characteristics interact in different ways. Finally, Snow recommends, among other things, that the researcher go into the school, prepare pupils for the experiment, and pay special attention to the "social context" in which the study is being conducted.

The challenge in this study then, was to maintain the integrity of the experiment by making it valid and reliable, and at the same time making it natural and generalisable. The next section describes attempts to control some of the threats to validity discussed by Campbell and Stanley, but in the light of the comments made by Snow.

Internal Validity

History

First, it was possible for events happening outside the walls of the school to influence events occurring within the school in general and classrooms in particular. The use of a control group helped to reduce the effects of history on the study where there was a problem, since what affected one group would perforce have affected the pupils in the other group as well. Further, the design of this study called for both Control and Experimental groups to

meet at the same time to minimise the influence of external events. Efforts were also made to monitor what was happening in the classroom so that the researcher could at least be aware of intra-session history.

The Experimental and Control groups at CAHS did meet at the same time every day during the course of the study. The researcher in various ways monitored closely what was happening in class. There were weekly conferences with the teacher in the Experimental group. At these conferences, materials, procedures, problems and constraints were discussed. At the end of each lesson, the teacher also made written comments in the teacher's manual.

Events happened during the study that could have affected student and teacher performance. First, at CAHS early in the year, some students and faculty demonstrated at the legislature to effect improvements to the physical plant. Again carnival, an island festival, is scheduled for April every year, and this activity closed schools for a week. And all this was in addition to other scheduled closings listed in the school calendar - 5 in all.

At IEKHS, the programme started two weeks late. Faculty went on strike over unfinished school buildings, class schedules and other matters. Indeed, six classrooms were under construction throughout the entire period of the study. A change of principal, 7 scheduled holidays and teacher absences also contributed to

reducing the time allotted to the study. Thus, class and teacher performance could have been affected by events other than the programme. This would limit comparisons between schools, although not within schools.

Maturation

Second, maturation which has to do with physiological changes within the student - changes not dependent on external events - could affect the results of a study. Factors such as boredom and fatigue may affect student performance from pre-test to post-test. This phenomenon was unlikely to have been a factor at one school; however, there was a gap between class times for the Control and Experimental group classes at Ivana Eudora Kean High School. The school had a shift system in operation, and because of this, one class met during period 3, and the other during period 7 right after lunch. This was the last period of the day.

Testing

Third, psychometricians point out that children can do better on achievement tests a second time as a result of test-wiseness alone. This made it even more desirable to minimise the effects of testing. In this study, all four classes were administered two objective pre-test measures: a grammar test and a test on the mechanics of writing. Students also wrote one or the other of two essays, and responded to an attitude measure. At the end of the study, alternate forms of the objective tests were administered; students wrote the other essay as a post-test, and once more responded to the attitude questionnaire.

The objective tests were administered by the testing officer of The University of the Virgin Islands; the researcher administered the other tests. Thus, instruments were administered under controlled conditions, although teachers were there to proctor the examinations. Post-testing was done at the same time as final examinations, at the end of the semester or marking period. Grades earned during the programme for composition and other exercises counted as part of the cumulative grade for the course. Hence, testing was made as natural as possible.

Instrumentation

Instrumentation refers to changes within the measuring instrument which could cause a difference between pre-test and post-test scores. The researcher used printed objective tests and children wrote the same two compositions. Again, both pre-test and post-test essays were graded at the same time after the two sets of scripts had been shuffled and coded. Raters were trained to use the analytic scale and rated papers on their own only after they had rated and discussed a number of scripts and internalised the criteria. Each paper was rated twice. Where there was a disagreement - viz., where one rater passed the composition, and another failed it - the script was read by a third rater. The paper got the score given by the two raters who agreed. Thus, shuffling and coding, together with careful grading techniques minimised the influence of instrumentation. Objective tests were, of course, machine scored.

Other Variables

Regression was not a factor in the study that could possibly have affected the result since the sample was not selected from an extreme pool of students. This was in contrast to the one class in the pilot study who selected a composition course as an elective. The assumption in that case was that they had a more positive attitude towards writing, and at least some were above average in writing ability.

The students involved were selected at random for the general English course. Again, all students completed the programme, although a few of them did not take all tests. The scores of the students who did not take all tests were not considered in the analysis.

Stanley (1966), and Borg and Gall (1983) discuss a number of factors that may limit the external validity of experiments. In the next section, certain considerations of validity are examined.

External Validity

Population

Researchers make a distinction between the experimentally accessible population and the target population. The first can be reached with relative ease by the researcher; the second is more remote and distant. The study was done on a sample of grade 12 students in public high schools in one school district in the Virgin Islands. Ideally, the researcher would have liked to extrapolate

the findings from the sample of grade 12 pupils to the population of such students in the two public high schools, and from public high schools to all high schools (public and private) in the school district, the target group.

The first leap from sample to the population in public schools was at least plausible since the two public schools in the district cater to the needs of most students of that age. The leap from the public to private high schools was not attempted. The issue was whether the two types of schools were so much alike to make such generalisations valid. Borg and Gall remind us that "generalising research findings from the experimentally accessible population to a target population is risky" (p.639). In general, private schools tend to be more homogeneous, students pay for their schooling and many schools are church schools. Again, the majority of students at these institutions could be identified as middle-class or better. These factors alone make these schools different from the public schools which have more heterogeneous populations.

The particular characteristics of students at one school - personalogical variables - could make findings inapplicable from one school to another. The two schools under study possibly had more similarities than differences. There was no initial evidence that the programme could have been more effective at one school than at the other. However, at CAHS, the longer tradition of the school, a more congenial and adequate physical plant, and apparently

a more confident faculty contributed to the public perception that CAHS was the preferred school.

Indeed, there was a sense of this difference even among faculty of both schools; among students it was more explicit. One indication of this involved the exchange of letters between schools - one feature of the pilot study. The teachers involved at IEKHS insisted that the main study should not include such an exchange of letters. However, reasons for this decision were not made explicit. This was regrettable since that aspect of audience awareness had met with very positive responses during the pilot study. Bracht and Grass (1968) suggest certain procedures to make generalisability more valid from one situation to another. First, the experimental treatment should be so clearly explained that another researcher could easily replicate the study.

Second, efforts should be made to control for the Hawthorne effect which might result when some groups are singled out for special attention. Such attention and not the treatment could be the stimulus for change. It is true that students and teachers used specially prepared materials, but there was not much apparent contact between students of the programme and Control group classes to make this factor likely. Students moved from class to class between periods, were grouped differently for different subjects.

Third, novelty and disruptive effects should be reduced. The content of the course was developed from the curriculum of the

English Departments of both schools and approved by all parties involved. As noted before, the assignments and grades for both Experimental and Control group classes were embedded in regular examinations or classwork. They replaced the assignments and exams normally given. Peer evaluation of students, essays was a novelty at first, since it seemed there was no systematic group work as a teaching strategy used at these chools, but after a few weeks, any novelty present had worn off, and it was not evident as a contaminating factor.

Bracht and Glass also discuss interaction and other effects which could have an impact on the results of a study. It is important that both Experimental and Control group teachers follow the protocol and procedures outlined. This issue of procedures was treated elsewhere in this chapter.

Finally, there are the interaction effects. The treatment, together with some other factor present, could affect the result of a study. For example, the interaction of the treatment with history in this study has been touched upon earlier. The interaction of testing with the treatment is also a consideration for researchers. The effect of the pre-test can affect scores on the post-test. However, Campbell and Stanley make the point that testing is not a phenomenon in schools but a common occurrence. To gloss this comment, it must be noted that even formal testing by outside agencies happens at least twice a year at schools throughout the territory.

Description of Instruments

STEP

The Sequential Tests of Educational Progress (STEP) Series II were developed between 1966 and 1970 to measure knowledge, skills and abilities "every well-informed citizen" (p.12) of the United States should have. The battery of tests measures objectives in reading, English expression, mathematics, science and social studies at grades 4 through 12. Each multiple choice test has 4 levels of difficulty with forms 2A and 2B appropriate for high school students.

The mechanics of writing instrument evaluates competence in composition skills at the micro-level in four categories. One section measures the ability of pupils to capitalise proper nouns, titles and use capitals with quotations. Another measures the use of the apostrophe with possessives and contractions. Still another measures the use of the comma with the following: independent clauses, words in a series, appositives, direct address, parenthetical elements, places and dates, and introductory elements. Finally, a miscellaneous section includes items on the semi-colon, period, hyphen, question-mark and quotation marks. The mechanics of writing test was normed in the Spring of 1970. Data were collected from 2,233 grade 12 pupils from across the United States. The 25 minute test is a power test with 45 items which are machine scored.

The English expression test assesses competence in evaluating the correctness and effectiveness of sentences. The test has 2 parts: Part 1 has 40 items and part 2 has 25 items. The test measures

the following skills: one subsection tests student knowledge of the correct forms of the verb, pronoun, adjective and adverb. Another section measures diction and idiom - the appropriate use of like/as, between/among; conjunctives, prepositions and correlatives; redundant and awkward sentences, and levels of diction. The final subsection measures structure: double negatives, modifiers, word order, fragments, shifts of subject, coherence and parallelism. Students wrote part 1 of this test only - a 20 minute test with 40 items.

For the pre-tests, student wrote Form A of the tests. Fifty-eight grade 12 pupils of Charlotte Amalie High School wrote the tests on Monday, February 8, and Tuesday, February 9, 1988. Fifty grade 12 pupils of Ivana Eudora Kean High School wrote the tests on Monday, September 26 and Tuesday, September 27, 1988 respectively. They wrote Form B as post-tests.

Composition

Two composition items were constructed by the researchers and field tested during the pilot study. Copies of both items are found in Appendix C. Meredith and Williams (1984), in discussing the characteristics of writing prompts, advise that prompts should provide the topic and audience, and specify the format; the wording, length, appropriateness and methods of development are also considerations. These variables were considered in the construction of the two items.

Writing Apprehension

Students responded to the Daly/Miller instrument to measure writing apprehension. This self-report instrument is designed to identify students who are anxious about writing. It was developed in an empirical study during Spring 1974 on a sample of 164 undergraduate students. The final version has 26 items in a Likert-type scale. It includes both positively and negatively worded items. A copy of the measure is found in Appendix C.

Multiple Measures

Researchers suggest that these three types of measures - indirect assessment, direct assessment and attitude towards writing are important for the assessment of writing ability. Research findings show that highly anxious students avoid writing situations or react anxiously to them since their writing acts tend to result in failure rather than success. Anxious students are less motivated or have a low self-concept. Of particular relevance to this study is the finding by Daly (1974) that such pupils "seldom engage in small group interaction" (p.247).

The evaluation of writing is complex, and one solution lies in the use of multiple measures to assess a range of skills and sub-skills. For instance, writing samples are generally more valid measures of writing skill than indirect assessment, but they are less reliable. It is often difficult to find out what a composition really measures because of the lack of consistency of measurement.

Readers are not consistent in assigning grades since irrelevant factors such as bias, boredom, fatigue, length of essay, neatness, handwriting quality can influence the scores they give. So, one can never be sure what a particular composition score means. Things are not much better when two raters read the same essay since there is the additional problem of inter-rater reliability.

Score reliability presents a second problem since the performance of students vary from mode to mode, and even within the same type of task. When we add inconsistencies of student performance from one occasion to another, what an item measures may be very unclear. One way to approach the question is to use both direct and indirect to gauge writing ability. Breland (1983), in reviewing the literature on essay reliability, writes that "direct assessment in which a simple topic and a simple discourse mode are used, clearly are limited in content validity" (p.18). He writes that the scores of one or two writing samples used to assess writing ability must be interpreted with caution.

In sum, although an analytic scale was used in this study and the scale descriptions provided an operational definition of what writing is, indirect measures of subskills helped to fill out the picture. However, although there was a unit on sentence structure, there was no separate unit on mechanics. Emphasis on correctness at sentence and word levels came into play as students did editorial exercises as part and parcel of the instructional programme. The focus was on "organic correction" (Cooper, 1975, p.113). Pupils

responded to their own errors or those of their peers; they saw problems in context rather than in the isolation of exercises in a text book.

Instructional Objectives

Experimental Group

Figures 2 through 5 present the instructional objective for the Experimental group.

FIGURE 2

LIST OF OBJECTIVES FOR LESSON

Component		Objective
Pre-tests	1.	Writes STEP objective tests
	2.	Writes composition
	3.	Responds to attitude measure
Content	4.	Knows characteristics of a paper rated high - professional models.
	5.	Consolidates criteria for a paper rated high - student models.
	6.	Knows criteria for a paper rated low.
	7.	Knows criteria for a paper rated middle
	8.	Discriminates between a paper rated 5 and a paper rated 4.
	9.	Discriminates between a paper rated 2 and a paper rated 1.
	10.	Rates sample papers in groups with consensus
	11.	Rates own papers in groups without consensus
	12.	Writes test composition.

FIGURE 3

Component		Objective
Organisation	1.	Knows characteristics of a paper rated high - professional model.
	2.	Consolidates criteria for a paper rated high - student model.
	3.	Knows criteria for a paper rated low
	4.	Knows criteria for a paper rated middle
	5.	Discriminates between a paper rated 5 and a paper rated 4.
	6.	Discriminates between a paper rated 2 and a paper rated 1.
	7.	Rates paper in groups for organisation
	8.	Rates sample papers in groups for content and organisation with consensus.
	9.	Rates own papers for organisation and content.
	10.	Writes test item.

FIGURE 4

Component		Objective
Structure	1.	Knows characteristics of paper rated high - professional model.
	2.	Consolidates characteristics of a paper rated high - student model.
	3.	Knows criteria for a paper rated low
	4.	Knows criteria for a paper rated middle
	5.	Discriminates between a paper rated 5 and a paper rated 4.
	6.	Discriminates between a paper rated 2 and a paper rated 1.
	7.	Rates sample papers for structure
	8.	Rates own papers for structure
	9.	Writes test item.

FIGURE 5

Component		Objective
Wording	1.	Knows characteristics of a paper rated high - professional model.
	2.	Consolidates criteria for a paper rated high - student model.
	3.	Knows criteria for a paper rated low
	4.	Knows criteria for a paper rated middle
	5.	Discriminates between a paper rated 5 and a paper rated 4.
	6.	Discriminates between a paper rated 2 and a paper rated 1.
	7.	Rates sample papers in groups for structure and wording with consensus.
	8.	Rates own papers for all qualities.
Post-tests	1.	Writes S.T.E.P. objective tests
	2.	Writes composition
	3.	Responds to apprehension measure
	4.	Rates test essays for all qualities.

Language Activities

Control Group

Figures 6 through 10 present the language activities for the Control group.

FIGURE 6

Component	Activity
Pre-tests	1. STEP objective tests 2. Composition 3. Writing apprehension.
Content	1. Overview of the writing process 2. Pre-writing: freewriting, brainstorming. 3. Practice in focusing and elaborating 4. Proofreading in groups 5. Paragraph writing: topic sentence, body and summary. 6. Reading and summarising 7. Writing expository composition 8. Reasoning: deductive and inductive 9. Types of evidence 10. Writing argumentative paper 11. Writing descriptive paragraphs.
Organisation	1. Outlining: filling, and evaluating 2. Coherence exercises 3. Techniques of ordering material 4. Evaluating student drafts 5. Organising the whole paper 6. Organising note cards 7. Format for bibliography 8. Formal outlines

FIGURE 7

Component	Activity
Structure	1. Subordinating and co-ordinating conjunctions.
	2. Verb endings
	3. Editing scripts
	4. Sentence combining
	5. The use of the comma
	6. Revision of story
Wording	1. Increasing vocabulary through study of roots of words.
	2. Malapropisms
	3. The language of poetry
	4. Film on development of creole
	5. Differences between creole and standard English.
	6. Language in The Suffrage of Elvira
	7. Quiz on research terms
Post-tests	1. Writes STEP objective tests
	2. Writes composition
	3. Responds to apprehension measure.

CHAPTER 6

CONSTRUCTION OF THE SCALE

The main concern of this research project was peer evaluation of writing. Because the term "evaluation" is nebulous and slippery, it is used here with caution. Loosely, the term is often synonymous with "measurement" or "assessment". However, Gronlund (1981) makes a distinction between the two terms. In a sense, "measurement" stops at numbers; it suggests a precision and accuracy that are hardly possible in the judging of writing. "Evaluation" is more comprehensive and can be "quantitative or qualitative or both" (p.6).

It is perhaps not difficult to envisage a teacher in the traditional marking mode awarding a composition 9 marks out of 10; often it is more difficult to explain what the numbers really mean - as numbers. The counting of linguistic and other features used to be popular, but these quantitative measures when used are now being supplemented by more holistic considerations. Teachers and students using the scale in this research project were only 'counting' when they were rating a script for structure or other grammatical conventions in ways similar to those employed by The Assessment of Performance Unit (APU), or the Britton research on multiple marking where markers added a mark for mechanics.

It is now perhaps more common to envisage a teacher making

purely qualitative judgements - responding in a general way to a piece using either hidden or overt criteria. On the one hand, there is the teacher using objective or quasi-objective criteria to circumscribe her judgements of writing quality; on the other hand, there is the response of the teacher as a reader interacting with the text, attempting to go behind and beyond it to the writer. Judgement is always involved when the English teacher is evaluating writing competence in a way that it is not when the Mathematics teacher is measuring Mathematics ability. In this report then, "evaluation" is not "measurement" in the sense described above.

Satterley (1981) points out that historically "assessment" used to suggest a teacher "sitting beside children . . . a close relationship and a sharing of experience" (p.1). Now, he stresses, it is often viewed as punitive and biassed. "Assessment" as a helping function is closer to the term "evaluation" as employed here. In sum, in this research report, "evaluation" is defined as being both quantitative <u>and</u> qualitative - a modification of Gronlund's definition, or the original concept of "assessment" as distinguished by Satterley.

In similar fashion, the term "writing" needs an operational definition since it can mean different things at different times even to the same people. Protherough (1983) makes the point that there is "no single, universally accepted model of what writing in English should be like" (p.57). This issue was discussed in

chapter three. The descriptive scale used, and the type of tasks set students provided a working definition of what writing is in the context of this study.

Some specific questions and answers will help to fill out the context further. Who were the evaluators? In this case, the students in the Experimental group were the evaluators. What were they evaluating? They were rating their own papers, those of their classmates, and sample essays collected from former grade 12 pupils, the semester before the project started.

When did evaluation take place? Evaluation took place while the children were learning. It was formative, so that pupils and their teachers could monitor progress. It was also summative in intent so that teachers and pupils could determine competence at the end of a unit.

In the long run, if teachers are confident that they can trust the judgements that writing students make, they may allow pupils to share in the final assessment of papers at the end of the course. How will the evaluations be made? Students will use a rating scale developed by the researcher.

Scale Data

Terminology

The first consideration is again one of terminology. Britton (1966) writes that "a scale consists of a number of sample compositions, the qualities of which have been determined by competent judges and which have been arranged in order of merit" (p.5). To judge a composition, the rater matches it with a certain sample along the scale of quality. Sometimes, the characteristics of these samples are described in writing as in the essay scales of Boyd (1924), or those of the London Association of Teachers of English (LATE) 1965. Scales may be unidimensional - there is one holistic dimension - and readers respond to the total impact of the composition using the principle that the whole is greater than the sum of its parts. Sometimes, they are multi-dimensional - there are separate subscales - and raters assess papers on several dimensions, so that the whole is considered the sum of its parts.

Cooper (1975) uses the term "holistic" scoring to describe "any procedure which stops short of enumerating linguistic, rhetorical, or informational features of a piece of writing" (p.4). This definition is broad and will include "analytic" scoring, although it would exclude "atomistic" scoring where markers dissect a composition for 15-20 qualities assigning a value for each. Diederich (1974) uses the term "general impression" to describe his marking method which is generally described as

analytic. Once users, Diederich maintains, internalise the criteria and start to rate scripts, rating "is no longer a blur; it is the quick summing up of characteristics that determine whether a paper is high, middle, or low in general merit." (p.44)

Breland (1983) uses Cooper's definition, but only part of it. He observes that holistic evaluation is usually guided by a holistic guide which describes each feature and identifies high, middle and low qualities for each feature. However, Breland omits the rest of Cooper's more comprehensive definition quoted above.

Theorists also distinguish between "classical" holistic scoring and "focussed" holistic scoring. In "classical" holistic scoring, a number of readers use their own standards built up over time to judge the quality of a script, and then the scores are averaged to determine a total score. The goal is to minimise differences and reduce variability in raters judgements. This superficially appears to be very much like impressionistic marking as used by Britton (1966), but there are important differences. The aim of the Britton model is to include rater variability in the assigning of a total score, not to exclude it, so scores are summed, not averaged.

In "focussed" or "modified" holistic scoring, raters assess writing on certain dimensions, although they do not focus on each category separately but on the whole or the gestalt - on the parts working together. The term "analytic" as used in this research

report involves a look at the parts of a composition as parts. Teachers and students read the paper and then judge it for each feature in turn. However, there is the "quick summing up" in the manner of Diederich.

According to Britton, in the U.S.A., the history of scales goes back to 1903 when Rice published a writing scale for mechanics only. Holistic scoring as described above started with Hillegas (1912). Hillegas was interested in grading compositions for general merit only, and he had different raters grading compositions holistically, summing the two scores. The first use of analytic scales is credited to Willing (1918) whose essay scale had two values - style and form. By the turn of the third decade of this century then, the analytic/holistic debate had started.

In the next section, the development of scales in the U.S.A. is treated. There is a close look at perhaps the most well known analytic scale, the Diederich scale developed for ETS in the 1960s. After that, a selection of analytic scales in current use is examined - scales used in Illinois, Connecticut, South Carolina, Texas and Maryland. In the section following that, there is an examination of the use of scale data in England. Discussion centres on the work of Boyd (1924), Wiseman (1944), Britton (1966), LATE (1965), and APU (1975). In the final section, the format of the scale used in the experiment reported here is described. The scale itself is presented in Appendix C.

The Diederich Scale

Diederich (1974) reports on a writing assessment project for the Educational Testing Service (ETS) conducted in 1961. The aim of the study was to determine "what qualities in student writing intelligent, educated people notice and emphasise when they are *free to grade as they like*" (italics mine). In other words, the team was interested in what good writing is as perceived by different groups.

Fifty-three readers from across professional disciplines rated 300 freshman essays on a 9-point scale. They comprised college faculty - English teachers, social and natural scientists - editors and writers, college English faculty, lawyers and businessmen. They were to use their own standards to rank order scripts and write comments to indicate why they gave papers particular ratings. The design of the study therefore, precluded rater training. The raters responded to papers at home. A factor analysis of over 11,000 comments gave the researchers their answer.

The analysis revealed that 16 of the 53 raters put a premium on ideas, 13 on structure and mechanics, 9 on organisation and analysis, 9 on wording and phrasing, and 7 on flavour or the personal qualities of the work. The main result of the study was the formulation of the Diederich Scale. Teachers from across New York learned how to use the scale during a two-year period when they met to rate different kinds of writing in schools.

The Diederich analytic scale is well known. It has been used in a variety of different situations - classrooms, school districts and state projects (Cooper & Odell, 1980; Beaven, 1977). However, there have been mixed responses to this important research study.

Freedman (1982) discusses the scale in the context of reliability. The Educational Testing Services (ETS) had been conducting various studies to improve the reliability of readers assessing essays since the turn of the century. However, Freedman argues that the Diederich study did not solve the inter-reliability question. She refers to methodological problems in the study and states that "the theory that led to creating the scale was flawed... The scale never gained popularity because it was time consuming and limited to the expository prose of older students" (p.87).

Turning first to the reliability issue, it is true that the inter-rater agreement among Diederich's raters was a mere .31. This perhaps was expected, for raters came from different backgrounds, were given hardly any guidelines, had no training and rated papers at home. No doubt, they used yardsticks from their own disciplines (Mitchell and Taylor, 1979). Still, the disagreement was gross: all 300 papers got at least 5 different ratings; 69 got 7; 111 got 8, and 102 received all 9 ratings (Breland, 1983). This could be contrasted with the high reliability of raters in the Britton study; the raters in that study all had a background in English.

However, the low agreement illustrated Diederich's point that "standards for writing are neither well-defined nor widely accepted" (p.1). The rater agreement increased dramatically to about 80% after teachers learned how to use the scale during training sessions - a quite acceptable co-efficient for the evaluation of composition. Beaven (1977), comments on the use of the scale as a useful peer evaluation tool; she reports that it has been widely used in different situations with <u>high reliability</u> - in classrooms, in school districts and across states. Thus, there is some disagreement among scholars on the reliability issue.

That the use of the scale is "time consuming" also appears in need of modification. When the purpose of formative evaluation is considered, the analytic scale seems to be very helpful. Indeed, as discussed below, education departments in certain states are having scripts rated twice. First, all papers are rated holistically; then failing scripts are rated analytically to give more information to schools.

But even as a summative instrument, the use of analytic scales in general does not seem inordinately time consuming. Diederich reports that the average grading time after rater training was two minutes per essay. Each essay is rated twice, and there is a third reading to resolve discrepant ratings, so that the average marking time spent on a paper was ten minutes. Considering the <u>purposes</u> of analytic scoring and the techniques currently used to improve inter-rater agreement, the time spent using the analytic scale does not seem to be excessive.

Freedman also refers to "methodological problems". One underlying issue is whether writing ability is a global or general quality. Recent theory suggests that perhaps it is not. The assumption of the Diederich team was that, if different views of what writing is are incorporated into one scale, then the modes of writing favoured by particular disciplines - as represented by the different groups of raters - could not bias the total score since all views were represented in the scale description and categories. Indeed, the team added "graphics", handwriting, to complete the picture of what good writing is.

This may not be a satisfactory answer to the question, but it is an attractive answer. The scale is empirically based (Cooper and Odell, 1977); it was developed by practitioners who spent two years using it to grade the essays of their students; and since its introduction, a wide range of teachers has found it helpful.

The fact that it is mode specific has also troubled commentators. Cooper and Odell (1977) and Wilkinson et al (1980) note that the assumption behind the scale is that criteria relevant for judging one expository piece of writing are appropriate for judging other expository work. Further, the scale presumes that criteria used in one mode are relevant and can be applied to other modes of writing. Such scales however, are often not sensitive to the demands made by different topics, audiences and purposes.

These are important limitations. Discourse theory in the 1980s stresses writing for particular readers for particular purposes and suggests that effective writing, among other things, is sensitive to the conventions of the various discourse communities. Breland (1983) warns that factor analyses like Diederich's are limited to the particular mode of the discourse, and further comments that analytic scales do not cover the entire domain of skills, as does holistic scoring" (p.12). Mullis (1983), describing analytic scoring, says that "characteristics chosen are generally those important to any piece of writing in any situation, e.g. organisation, content, and mechanics. However, results are often more useful if the characteristics are derived from writing done for particular purposes and audiences" (p.18). It is debatable however, whether organisation is common to "any piece of writing".

Another difficulty seems to be that very often scales cannot be compared without the descriptions, for although there may be some features of writing generally applicable to most writing tasks, the labels used to describe them are not always very helpful. Terms like <u>organisation</u> and <u>style</u> may mean different things in different situations.

Still, if some characteristics are generally applicable, it follows that other features are not. Would we then want to incorporate the latter into a scale to be used by pupils? If so,

would not these mode or task specific scales have to change as the writing tasks change? Perhaps the ideal solution to this problem would be to make the scale mode/task free, and give general characteristics rather than specific ones. Then the circumstances of teaching and learning, together with the conditions and contexts of the various writing tasks, will give students a set for writing, and supply the specificity needed.

This has been suggested by Breland, and is currently used in the description of the APU analytic criteria as described by White. Evans (1977), in his work with Ontario teachers developing criteria for evaluating writing in grades 7 and 8, also distinguishes between general criteria and specific criteria. These criteria were developed by teachers from the various school boards and followed very closely the Diederich model.

In summary, the Diederich scale has certain advantages. It was the result of empirical research, and it was fashioned by teachers in the classroom. It has been used successfully in a variety of situations with good reliability. It has content validity and is successful for certain purposes of evaluation. However, the scale like other analytic scales also has certain limitations.

Focussing on the parts may cause us to lose sight of the whole: Presumably, in various types of holistic scoring the entire gamut of skills is taken into account. In more and more

situations, there has been a compromise; raters now consider the whole paper and then the parts. It seems that no mode of rating essays in use is without limitations, and for the purposes of this research, the use of an analytic scale seemed appropriate.

<u>Other Analytic Scales</u>

Despite the limitations of the analytic scale, it continues to be an important evaluation technique. The state of Connecticut has two large scale testing programmes in place, each containing a writing component. According to Baron (1984), the Connecticut Assessment of Education Progress (CAEP), established in 1971, tests proficiency at grades 4, 8, and 11. For the writing test, CAEP raters use an analytic scale to grade two 20 minute essays - narrative and persuasive. Examiners stress that competence at both the macro and micro levels is important; consequently, they assess compositions at the word level, the sentence level, and the text level. At the text level, papers are judged for focus, support, organisation, audience and style; at the sentence level papers are rated for syntax, and at the word level papers are rated for spelling, capitalisation and punctuation.

The new factor here is audience. This is in keeping with recent emphasis on audience in writing research. Diederich collapsed considerations of audience under ideas, organisation and wording, but he did not isolate it as a separate category.

Breland (1983), discussing the Steele subscales which included audience and purpose, notes that the inter-rater reliability for this category was only .48%. He comments that the dimension is difficult to measure. Wilkinson suggests some reasons for this difficulty. He notes that the Assessment Performance Unit in England does not include an audience category in the analytic scale for writing tasks. He comments that audience is hard to control in examinations; it is essentially a spoken concept, and its influence in writing is perhaps overestimated.

The state of Illinois also has a writing assessment programme run by the board of education. Chapman (1984) reports on the Illinois experience. The Illinois Inventory of Educational Progress (IIEP) has been using an analytic scale since 1983 to judge functional writing proficiency in grades 4, 8 and 11. The test is criterion-referenced and the analytic scoring is the scoring technique used so that detailed information could be transmitted to teachers on the strengths and weaknesses of student achievement in writing. A 25-minute persuasive writing prove is used across grades, and papers are assessed for 5 qualities on a 6-point writing scale. The categories of the scale are focus, support, organisation, mechanics, and overall quality.

The new category here is <u>overall quality</u>: how well the elements combine to address the demands of the assignment. The issue is how to keep the advantages of holistic scoring, and the

advantages of analytic scoring: the ability to give more information to interested agencies.

Chapman underscores the advantages of analytic scoring:

> "While each particular writing item... has its own unreliabilities and invalidities, taken together, they are quite powerful in describing the student's ability...Therefore, this approach gives information to both the instructional teaching and the evaluation/ measurement community." (p.25)

Connecticut and Illinois are only two of the states with state-wide writing assessment. States like Connecticut have two programmes - one to establish competency, the other to improve instruction. Three states with competency based programmes set up after legislation in the 1970s to determine minimum requirements for students are South Carolina, Texas and Maryland in 1977, 1978, and 1979 respectively. These programmes have common features: they are criterion referenced and focus on basic skills in certain curriculum areas. Texas has mastery exit tests in reading, writing and mathematics at grades 3, 5, and 9. Students may resit exams until they meet requirements. In Maryland, grade 7 and 9 students may do repeat tests.

The writing tests given must be viewed in the context of the larger assessment policy. Both Maryland and South Carolina use a combination of holistic and analytic scoring. All three states

use modified holistic rating scales where raters inform their holistic grade by considering certain specific features of the texts. This seems to be a compromise between general impression-marking and analytic scoring. However, South Caroline and Maryland score failing papers "analytically" to give feedback to schools. In Maryland, one rater assesses a sample of papers in 5 areas: content, organisation, audience, structure and conventions; in South Carolina failing papers are all scored analytically to give feedback to schools.

To sum up, there are two types of scales widely employed across the United States - holistic and analytic - but the terminology is not always clear, and both types of scoring have certain characteristics in common, for example, sample essays or anchor papers arranged along a scale of quality from low to high. The holistic is useful in certain types of assessment for certain purposes; the analytic scale is used for both summative and formative assessment and seems to have a wider range of use - especially when dimensions of audience, and overall quality are included. Modified holistic scoring is a compromise between the two. Both types of instruments have been developed to such sophistication that they yield inter-rater reliabilities of 80% or better. In the United States, the tendency seems to be the use of a combination of methods to give as much information as possible to interested agencies.

The Case of England

In Britain, large scale assessment of writing centres around a series of examinations. First, there was, until recently, the primary school examination at 11+, a placement test for English secondary schools. Then there are the various school-leaving exit examinations at 16+ or 18+ conducted through one or more of the examination boards. Finally, more recently, there is the writing component of the national testing programme run under the auspices of the Assessment of Performance Unit (A.P.U).

Both general impression marking and analytic scoring are evaluation approaches which have been used, or are currently being used, to assess the quality of writing at 11+ and 16+, and 11 and 15 for the APU. The Britton team distinguished between the official marking procedures used in secondary exit exams for writing evaluation, and the experimental marking used in their two research projects. Multiple marking is different from the traditional general impression-marking and has influenced marking procedures used in various assessments of writing.

In this section, two types of scale data are examined. First, the development of analytic scales is explored. After that, scale data for non-analytic grading are discussed. Next, the concept of multiple-marking popularised by the Britton team is treated. But multiple-marking calls for scales only in a special sense, since the points on the scale are not described and raters use their own standards as to what good writing is to assess papers.

Analytic Scales

Wilkinson (1980), summarising the history of scales, point out that those early scales revealed problems of definition and overlap of categories. There was no attention to audience and function, and the scales were used to judge all types of writing.

Hartog (1944) headed a committee to improve the validity and reliability of markers rating compositions at 11+ and 16+. The Hartog team introduced the categories of "sense", by which they meant function and audience, as an improvement in the judging of compositions at those levels. Their schedule listed general impression; expression (vocabulary and structure); sense (function and audience); spelling and punctuation. Thus, by 1944, general impression marking and analytic scoring seemed to have fertilised each other, and the validity question, one consideration of the Hartog Committee, seemed at least partly resolved.

Essay Scales

Protherough (1983) makes the point that there was a revolution in the philosophy of writing assessment between 1924 and 1965 which was reflected in a shift in scale criteria. He illustrates this philosophical shift during the forty-year period by contrasting two scales: that developed by Boyd in 1924, and that described by LATE in 1965.

Boyd (1924) used a panel of markers to judge a number of

essays written by 11+ pupils in English schools. Analysis of the illustrative papers along the scale of quality indicated the panel's views on what good writing is, and by inference the writing theory behind those views. The mimicking of adult models points to the artificial and sentimental being preferred to the sincere expression of personal feeling. Other qualities valued by Boyd and his examiners were an elevated style, a detached point of view, an adult voice, use of rhetorical devices, and "clean copy" - an absence of spelling and punctuation errors.

These were the qualities illustrated by an "excellent" essay written by a 11 year old girl. Protherough commented that the "objective sureness" which the scale suggested could have sent the wrong signal to teachers, with a possible backwash effect on examinations and teacher emphasis in the classroom.

In contrast, the model suggested by LATE seems to be influenced by the personal growth model that came out of the Dartmouth Conference with its emphasis on experience through an integrated English curriculum stressing language for life. One essay in the scale used to illustrate the qualities favoured was a piece recreating a pupil's own experience written in a vivid, sincere and spontaneous style.

Protherough notes that the "imaginative coherence" of the paper points to its significance for both reader and writer. The use of creative language - language used in personal and unique ways - is valued.

Finally, the dimension of mechanics has a low priority with the emphasis being on "adequate control". Protherough points out that one "A" piece included more than 30 mechanical errors! This perhaps was an extreme case, but the shift in emphasis and the premium placed on the student's own voice, language and experience underscore the philosophical shift. In summary, the LATE criteria stress experience and language: how far an experience is realised and explored, and how far the writer exhibits control over structure and punctuation.

Britton's markers in the 1964 experiment apparently shared a similar philosophy. They valued, in order of importance, the involvement of writers with the task, the internal organisation of the piece, and the "general shape" of the composition (sentence structure, paragraph structure, and "aesthetic form"), and mechanical accuracy. Britton notes that, although a separate mark was awarded for mechanic, "there was a general tendency to allot little weight to this criterion" (p.23).

Impression Marking

General Impression marking has emerged as the main alternative to analytic scoring, and to such essay scales as those of Boyd and LATE. Wiseman (1949), commenting on the use of the analytic scale, notes that in marking 11+ exams, the best essay did not always come out on top, and analytic grading demanded much time and labour with no incremental return over and above impression marking.

Wiseman adopted the concept of multiple marking, although he was not the first to use that method. Multiple marking has two features: first, different readers mark the same script; second, consistency between and among raters is not used as an index of reliability. Rather, intra-rater reliability is important: the consistency of a rater with himself as he assigns grades at different times to the same scripts; third, the goal is not to minimise differences between readers, but to include them. Raters normally re-mark a sample of previously scored essays and reliability co-efficients are calculated; fourth, there are no standard criteria which readers have to internalise before marking. Conversely, in typical large scale assessment exercises, there is careful training of team leaders and markers, standardising through anchor papers and check papers etc.

Britton traces the history of the method to Robertson who had a team using it to rate 11+ compositions. After an initial experiment proved promising, Britton conducted a large scale project using "O" level examination papers. One crucial feature of the Britton experiments was the technical efforts to achieve high validity: to ensure that compositions were actually measuring writing ability by using criterion related validity checks.

The issue was essentially a rater's consistency with himself on the one hand, and his consistency with other raters. A rater could be consistent in assigning the same grade to the same scripts

at different times; different raters could also agree on the ratings for the same scripts; however, the construct being measured might not be what the examiner set out to test in the particular examination. So the question of validity was not completely answered.

Content validity - a close scrutiny of the item and then field testing it could help to answer the question as to how far the item was measuring writing ability. Concurrent validity - correlation with a measure taken at the same time, for example, a multiple-choice test, could help. The Britton team used an external criterion of a sample of ten essays collected over a year from the same students writing "O" level examinations. Correlation coefficients were then calculated on the two sets of scores: the essays collected during the year, and the "O" level essays.

Britton was interested in, among other things, the self-consistency of his raters, and the agreement among the marking teams. He was also interested in the agreement among the official markers; and the extent to which the scores of both groups correlated with the external criterion.

The Britton examiners for the pooled assessment used a 10-point scale. No special criteria and no briefing were given. Raters were asked to indicate the standards they used only after the exercise was completed. The maximum mark awarded a script was 30. Examiners read about 12 scripts to get a standard and then marked papers.

Two months later, raters re-marked twenty per cent of the scripts as a check for self-consistency.

The official markers employed in 1964 by the G.C.E board, in contrast, were carefully sensitised. Each examiner had to study material in advance; there was a detailed marking scheme, sample scripts were photocopied and marked by raters, there was a general meeting for discussion and possible modification of the marking scheme. All this was in addition to other safeguards.

Results confirmed that multiple marking was more valid and reliable than individual marking with careful training. Reliability between the experimental teams ranged from .78 to .81. The self-consistency of markers was .91. In contrast, correlation among the official workers was only .52, despite the elaborate preparation. Official comparison with the criterion was .57; the experimental comparison ranged from .63 to .68.

> "The figures suggest that our multiple marking method has achieved higher reliability than normal marking methods without any loss of validity, indeed with some improvement in that respect also."
> (Britton, 1966, p.24)

The Assessment of Performance Unit (APU)

The Assessment of Performance Unit uses both general impression and analytic marking techniques. Satterley (1981) discusses the aims of the APU which are similar to those of the National Assessment

of Educational Progress (NAEP) in the United States. The goals are to provide for interested agencies a national picture of pupil performance and hence school effectiveness, to identify and appraise existing assessment instruments and create new ones, and give feedback to schools. There are six panels: 3 assess competence in Mathematics, Science and English. However, according to Satterley, controversy surrounds the measurement of personal, aesthetic and physical development.

As described in the working paper, Language Assessment (1978), the English Panel assesses a wide variety of writing tasks with special emphasis on audience and function to measure pupil performance at 11 and 15. The pamphlet gives a list of the types of tasks to be assessed. Wilkinson praises the writing tasks and objectives, but considers the analytic criteria unimaginative. There is overemphasis on conventions - orthographic and grammatical. He comments that these are the only ones carefully defined, and there is a mere "genuflection" to style, structure and content.

White (1986) discusses APU assessment of writing procedures. The above criticism loses some force in the light of her exposition. White points out that "the categories concerned with grammatical and orthographic conventions relate to standard features of written language" (p.6). Considerations of general merit, she points out, are more sensitive to task, audience and function.

There is impression scoring which is similar to the procedures developed by Wiseman and Britton. Papers are double marked with raters using a 7-point scale. The self-consistency of raters is considered in the determination of final scores. Markers are briefed, but instructions are not detailed. Scripts are read rapidly by raters using their own standards.

But White points out that analytic marking is also used for further analysis in order to give "insight into the exact nature of strengths and weaknesses of performance in relation to a given task." White describes the 5-point analytic scale used. A panel of 4-6 scorers double mark a 10% sample of scripts. The criteria are sensitive to audience and function, and the demands of particular tasks.

In sum, development in England mirrored that in the U.S. with one notable exception. In both cases, there was concern with validity and reliability issues, refinement of analytic marking scales, and considerations of audience and purpose. However, the movement in America appears to be away from "classical" holistic scoring to modified or "focussed" holistic scoring. In England, the move seemed to be from classical holistic scoring to impression marking. Finally, both impression and analytic scoring are used to assess the same essays. Or at least general impression marking seems to exist next to the official marking procedures of examiners rating 'O' level scripts.

The Scale

Format

The rating scale used in this research study comprised four components: content, organisation, wording and structure. Each quality was assessed on a 5-point scale; ratings on each dimension were low (1 and 2), middle (3), and high (4 and 5).

Most scales examined contained the sub-category, mechanics. Diederich's raters considered mechanics as very important and isolated punctuation, spelling and handwriting as separate categories. However, the mechanics category was not included as a separate dimension in the scale used in this study. There is evidence that, in many writing classes, there is overemphasis on grammar and mechanics. When these qualities are given too much weight in assessments, the problem is aggravated further. Therefore, although structure is a separate sub-scale, considerations of mechanics were organically built into the lesson sequences as described here.

Many of the student samples selected were included in the booklets as they came, unedited. Before working on a script, students had to edit it working alone, in groups, or as a class. Since pupils were rating their own work or the compositions of their peers, corrections were more meaningful and pupils learned from one another. Thus, for every unit, pupils potentially could be involved in at least four proof-reading exercises. Students also had conferences within their groups to respond to one another's writings and make suggestions for improvement, both at the macro and micro levels. Cooper (1975, p.113) notes that:

> "correcting usage and syntactical and rhetorical deficiencies organically by working with the students' own writing and not by pre-teaching rules [puts the emphasis] on diagnosis, on formative response and evaluation to enhance the complex and highly individualistic process."

The assumption was that proof-reading skills would transfer and become part and parcel of the students' own writing. Students in the experiment were told that, when the teacher was grading unit tests using the scale, he would include a mark for mechanics since they had spent so much time on it.

Again, there was not an audience category in this scale. However, considerations of audience were incorporated into the programme in the following ways. During the study, pupils wrote for two distinct audiences within the classroom. The primary audience was their peers; the secondary audience was the teacher. They talked about their work using the terminology of the field; often as many as four students responded to a paper and then the group leader led discussion to determine a final grade. The grade assigned stood and became part of the record.

During the pilot study, another distinct audience was in play. The experimental groups from both schools wrote to one another, so there were at first an unknown audience outside the school. Again, at the end of the pilot study one class collected selected essays for publication in a class booklet, so this expanded the audience.

All the writing prompts given included directions as to purpose and audience. Finally, the scale also did not include an overall quality as in certain scales described earlier.

The scale was odd-numbered. It is true that some raters have a central tendency and tend to assign many middle scores (Coffman, 1971; Sager, 1975). However, the use of a 5-point scale is influenced by Diederich (1973) and Evans (1979). Students learned to distinguish between papers rated low, middle and high. Characteristics of each were described and 'anchor' papers studied. Once students decided on the placement of a paper, they then made another decision as to how low or how high the essay was.

One advantage of the 5-point scale is that with large numbers scores tend to have a normal distribution and the numbers become standard scores. Again, when sub-scores are totalled, they fall into a certain range and grades of E, D, C, B and A can be assigned. Teachers using the Diederich scale gave double weight to ideas and organisation. In this study, categories carried equal weight so the lowest possible category score was 5 and the maximum 20. Since the teachers involved use percentage scores, they did the necessary arithmetic to transform the ratings.

Scale

Certain assumptions were considered when using the scale. The first assumption was that the analytic scale is the best tool for

formative evaluation, despite the limitations discussed above. The second assumption was that the conditions of each assignment and the classroom environment helped to fill out the rhetorical context for the typical writing task. The third assumption hinges on the second point - certain categories were made as task free as possible to make them more generally applicable. The mechanics categories were standard for most types of writing, though this must not be overemphasised.

White (1966) points out that, for the APU, assessment content and the organisation of that content are determined by audience and function and suggests that "the range of conventions employed in writing to ensure that texts are unambiguously comprehensible and legible" (p.6) are not. In other words, general merit categories are task bound; mechanics categories are task free. The second point certainly is generally true, but if reading is a creative activity demanding an engagement of the reader with the text as the reader makes meaning, then following conventions alone (though aiding information processing), may not lead to "unambiguously comprehensible" texts.

The fourth assumption is that it is futile to have separate and discrete categories in an absolute sense since all the things that impact on the writing act, that go into spinning a web of meaning, are holistically intertwined. Thus, content is tied to organisation, and wording cannot be divorced from style.

Content

The content category stressed three considerations: focus, support, detail, and the links between them. There was a growth scheme or ladder of competence for each of these skills. Good focussing hinges on effective pre-writing. The incompetent student may not restrict the subject at all and attempt - often in futile fashion - to develop a subject. Another student may give some thought to the task but restrict herself to developing a topic, albeit with some competence. The superior student though would progress from subject, through topic to purpose translating it into theme; she would consider the constraints of time, and the type and conditions of the assignment before deciding on her intention. In brief, the superior writer will have a sharp focus.

The other aspects of content had to do with what Moffett calls elaborating and relating, "qualifying". The writer must stay on focus and at the same time marshall sufficient support and detail to make the point; at the same time, he must use techniques such as subordination and co-ordination, conjoining and embedding to link ideas within and between sentences. Relating and elaborating are tied to the rhetoric of the sentence.

Organisation

Organisation was defined as a sense of plan. It is the thread that binds the parts of the composition together. There were two levels of organisation in play. First, there was coherence or the

relationships between the different parts of the paper at the text leve; then there was cohesion or the relationship between sentences at the paragraph level. A writer's organisational schema is determined by the conventions of the discourse community within which he is operating.

Thus, in the suggestive description foci of the experiment, the apprentice writers described a scene from a fixed point or from a moving point. In the expository tasks, they used the conventions of that mode. For argumentative tasks, the writers shifted from the chronological order of narrative to the analogical order of persuasion.

Structure

Considerations of grammar and syntax have been treated elsewhere in this research report. This category focussed on composition at the sentence level. The emphasis was on getting students to see relationships between different elements of the sentence. The emphasis was on organic correction, as described by Cooper above.

Wording

Wording is terminology borrowed from Diederich. It includes not only vocabulary considerations, but what that writer calls "flavour". The writer puts his personal stamp on a piece by using language in new and exciting ways. It also includes using language correctly, and considerations of spelling.

The rating scales are all to be found in Appendix C.

CHAPTER SEVEN

LESSONS TO TEACH THE COMPONENTS OF THE SCALE

The self-contained writing programme included materials intended for both teachers and students. There were four teacher manuals matching the four components of the scale - one for each of the four units that made up the programme. These components were content, organisation, structure, and wording. There were four sets of student booklets composed of stimulus materials - usually professional models; scale descriptions of high, middle and low papers; student compositions, writing samples, and exercises.

Each teacher's manual started with an introduction which included a letter to the teacher giving an overview of the unit, a list of sequenced instructional objectives, and a chronology of assignments. The introduction was followed by ten lessons, one for each day of the two-week unit.

Each lesson in the teacher's manual followed a pattern. First, there was a summary sheet listing the objective(s), indicating the level of intensity, instructional roles, use of space, and specifying materials. After the summary sheet came the teacher materials. These were the same as for students. Then a lesson plan described the different phases of the lesson. Sometimes the lesson plan was followed by teacher supplementary materials. The rationale for these

was that the experienced teacher would only use the main line of the lesson plan; however, the less confident teacher could rely more heavily on the protocols provided. The researcher urged instructors to view the protocols as descriptive rather than prescriptive, and to follow the spirit rather than the letter of the teacher materials. Finally, each lesson ended with a sheet for teacher comments.

Teaching the Scale

Lessons were designed to teach pupils the characteristics of high, middle and low papers. Generally, the pupils worked in groups. Following Beaven (1975), at first groups were small and worked on specific tasks for 15 or 20 minutes. Pupils worked in pairs, in groups of three, and finally in groups of four - the group size becoming increasingly progressive. The goal was to foster a co-operative spirit, so groups membership changed frequently. Later, groups became larger and stayed intact for longer periods.

During the first week, students internalised the features of high, middle and low papers through the study of professional models and writing samples collected during the pilot study. At first, the teacher played the dominant role, but later students took on more and more responsibility for their own learning.

Students also distinguished between two high or low papers. Individuals in each group matched the sample with the high or low

anchor paper they had on file to decide which was better. Then ratings were compared within and between groups to decide on a final rating.

Pupils were encouraged to talk with one another, and teachers spread questions across the class to ensure wide participation and monitor progress. In short, discussion was both vertical and horizontal. The Sager lessons were "self-directing and self-correcting" to stimulate group discussion (p.86); in contrast, teachers of the experimental programme often asked for the answers to the exercises and discussed responses.

Using the Scale

During the second week, students rated papers. Lessons were constructed around a three or four phase sequence. In phase one, the class divided into groups. At first, individuals in each group worked alone to place the writing sample along a scale of quality. During phase two, the leader of each group asked for individual ratings. If members did not give the same ratings, i.e. when there was disagreement within the group, pupils had a discussion so as to reach consensus and determine a group rating. In phase three, the teachers asked for the group ratings to compare ratings between groups. If there was disagreement at the group level, discussion followed to decide on a class rating. At first, each student rated his or her own paper; then the script received peer ratings, and finally a class rating.

Students used score sheets which were collected by the researcher for analysis. Peer rating was organised round-robin fashion. Teachers distributed individual and recorder rating sheets. First, pupils rated their own scripts and recorded the score. Then, papers were circulated clock-wise within the group until all papers were marked and the scores recorded. Next, group leaders asked for ratings and a mini-conference on each paper followed. During the conference phase, the student asked questions and wrote comments. Writers could revise papers if they wished along the lines recommended. At the end of each unit, the class wrote an in-class essay as a final test. This assignment was rated by the teacher only. The instructor scored it for the particular quality or qualities, and added a score for mechanics.

Lesson 6 presents a typical lesson in some detail. There is an example of a summary sheet with explanation and commentary on the categories. This is followed by the lesson plan with discussion of the rationale. Next, comments on the lesson are given. After that, the ratings of the two classes are presented with limited analysis. Finally, an example of an edited script follows. The purpose of this rather full description is to give readers of this research report a global and comprehensive overview of a total lesson sequence from start to finish. The other lessons selected illustrate other aspects of the programme and are given abbreviated treatment.

CONTENT

Summary Sheet

FIGURE 8

SUMMARY SHEET FOR LESSON 6: CONTENT

OBJECTIVE:	(1) Discriminates between a paper rated 5 and a paper rated 4. (2) Proof-reads paper.
LEVEL OF INTENSITY:	Interactive with explicit rules.
INSTRUCTIONAL ROLES:	Main burden shared between teacher and materials.
USE OF SPACE:	Students work individually.
MATERIAL:	* 28 copies of "Bedlam". * 28 individual rating sheets.

Objective

The first item is the objective for the lesson. Each objective is an instructional objective after Gronlund (1981) rather than a behavioural objective as in programmed instruction at the training level. In short, the emphasis of the experiment was on developmental rather than mastery outcomes. Thus, the investigator would want the

pupil to tell the difference between any paper rated 4 or 5 in the content, or any other, category. Clearly, such an objective could not be mastered in one 45 or 50 minute session; rather, pupils would reinforce this distinction through subsequent individual and group activities.

A second objective in lesson 6 is student practice in proof-reading papers - theirs or other students. The programme assumed a process approach to writing: pre-writing, writing and post-writing. This second objective afforded students practice in post-writing - revising first drafts after feedback given in collaborative learning experiences. The stress was on organic correction, editing their own work, a task which many students neglect since they find it uninteresting and dull.

Level of Intensity

Cognitive psychologists point to the importance of rule structures and heuristics if learners are to become at least competent in problem-solving activity. Robinson et al (1985) stress that the aim of teaching and learning is to inculcate in pupils workable approaches for solving different kinds of problems. These strategies should become automatic through overlearning. Robinson and his associates stress that they:

> "believe that the intensity and quality of instruction changes (sic) significantly when the rule basis of the behaviour is evident to the teacher and consistently employed in instruction." (p.271)

At the lowest level, the instructor merely provides the opportunities; at the next level, the teacher may model mature behaviour; at still another level "the teacher uses the rules as a personal agenda for interacting with students." Sometimes, the rules are made explicit; at other times, they are elicited from the pupils.

Pratt (1980) distinguishes three main stages of skill learning - cognition, fixation and automation. Cognition is fundamental since skills have a knowledge base; fixation has to be achieved through constant practice, so that skills and strategies could become second nature.

In lesson 6, the rule structure is built up in systematic order. Pupils work alone during the first phase, but after that there is dialogue between and among students and teacher as the class gropes for consensus.

Instructional Roles

Instructional role describe who or what is doing the teaching. In the traditional classroom, the teacher is dominant. For example, at best the teacher will model desired behaviour, or at worst lecture; however, the responsibility for instruction can shift. Children can teach themselves as they talk or write to learn; or the materials, for example, books or audio-visual aids, may be the centre of the teaching activity. In lesson 6, the writing sample "Bedlam" and,

an earlier sample, "Lovely Room" had to be processed by students as they interacted with the material.

Use of Space

All four classrooms were equipped with a desk and a chair for each individual student, so that the arrangement and disposition of furniture were in the complete control of the teacher and students. This made different arrangements possible and facilitated students working in groups, individually, or as one class.

Materials

Each summary sheet ended with a list of materials needed for the lesson. Copies of the handouts were in the student handbooks, and of course, in the teacher manuals. Rating scales for each student were distributed by the teacher. Figures 9 and 10 present the writing materials included for lesson 6.

FIGURE 9

LOVELY ROOM

ANCHOR PAPER FOR SCRIPT RATED HIGH - 5

Instructions: Students were asked to write a description of a room in their house so as to give a clear picture of it to their classmates. They were to use the guidelines for content summarised in the scale for high and illustrated through the study of a professional model.

1: The living room (or family room as some may call it) is the most lovely room in our house. 2: My mother fixed it so pretty she doesn't want a soul to go in there if he/she is dirty or has shoes on. 3: It often reminds me of Eudora Kean High School because the creative colours that were used to decorate this room are similar to that of the school colours.. 4: The living room is not only the most beautiful room but the largest. 5: The drapes are of a goldish colour corresponding with the walls that are furnished with beige panels. 6: Against the walls in a brass shelf or wallpiece layered with crystal piece. 7: The shelf also has these special lightings that gives the ornaments such life and brightens the whole room, especially at night, to make the scenery extra special. 8: On the floor, there's this square oriental rug (my mother preferred an area rug because she didn't want to cover the beauty of the goldish-brown tiles) and on the rug the coffee table (we call it the centre table because it is in the centre of the room) is placed which is also in brass similar to the wall piece. 9: The sitting set is of maroon velvet: a couch, love seat, single; and 2 pink lamps are placed on either side of the couch and to give the room an outdoor feeling, there are some plants in the corners. 10: The best part about the living room is the view that you see from there. 11: When you open the glass door, it leads you straight to the porch which overlooks the beautiful view of the ocean. 12: You can even see St.Croix on the horizon.

FIGURE 10

BEDLAM

ANCHOR PAPER FOR SCRIPT RATED HIGH - 4

Instructions: This was another response to the assignment.

1: In my bedroom there is bedlam and chaos. 2: Let us imagine a door that looks as though it has been lacerated by bullets from an M-16 machine gun. 3: As you enter, the door begins to creak and watch out for falling objects. 4: I call them objects because you have no idea what it is or how hard it is going to hit you. 5: The room looks sort of misty like the jungles of China, and watch your step for the slippery rock formations (or the leather shoe pile) whichever you prefer. 6: Then if you look hard enough through the mist you might see what appears to be dragon flies (or roaches) that seem to smile with you. 7: Then there is the swamp (or bed as you called it), anything that is put on the top of it may never be seen again for awhile. 8: The rest of the room may appear to you that you've entered a time warp - believe me, it is a moving experience.

Teacher's Lesson Plan

Figure 11 presents the teacher's plan for lesson 6.

FIGURE 11

TEACHER'S LESSON PLAN (6)

OBJECTIVE: Discriminates between a paper rated 4 and a paper rated 5.

PHASE (1):

GETTING STARTED

* The teacher
    * states objective:
        * proof-reads for agreement errors and fragments.
        * rates "Bedlam" individually.
        * compares "Bedlam" with "Lovely Room".
    * distributes "Bedlam".
    * directs students to proof-read essay showing corrections on the script.
    * tells students to correct script.
    * asks for corrections.

(Time: 10 minutes)

PHASE (2):

RATING SAMPLE ESSAY

* Teacher

  * directs students to rate "Bedlam" HIGH, MIDDLE OR LOW.
  * cautions them to consider the content only.
  * distributes individual rating sheets.
  * asks them to fill in details on sheet.
  * calls for rating.
  * probes students to give reasons for rating.
  * and class decide on overall rating.
  * ensures that rating is recorded on the score sheet.

(Time: 25 minutes)

PHASE (3):

COMPARING

* Teacher

  * tells students to select "Lovely Room". from their study booklet. This script had been rated high in lesson 3.
  * lets each student decide which is better.
  * asks for classification and records tally
  * allows for thorough discussion.
  * and class decide on final rating.
  * gives rule.

N.B. The assumption here was that the class rating were "Lovely Room" 5, and "Bedlam" 4.

When a paper is rated HIGH for content, the student must decide how high the paper is. If it as good as Lovely Room, it gets rated 5. If it is not as good as Lovely Room, it gets a 4.

(Time: 15 minutes)

Phase 1

Robinson et al discuss a six-phase lesson sequence: getting attention, motivating new learning, establishing new learning, consolidating new learning, and linking new learning to old learning. In the programme, Lesson 6 is divided into three phases. The typical lesson plan details the moves of the teacher within and between phases.

Phase 1 in lesson 6 is the individual proof-reading phase. Essentially, pupils were applying knowledge and rules already learned. The instructor, as she listened to corrections, had the opportunity to clarify, elucidate, or correct stylistic and mechanical problems.

Phase 2 was the individual rating phase. It must be observed that by lesson 6 - the start of the second week of the unit - students should have internalised the criteria for a high, middle, and low

paper. Students were to rate for one quality only, content. The class next decided on where the paper fell on the scale of quality.

Phase 3 was the comparing phase. Discussion as to which script was better was the key to this phase. The rule was given at the end. The two scripts then became the anchor papers for papers rated 4 and 5 respectively. At this leve, a 4 paper suggested superiority; a 5 paper demonstrated superiority.

Teachers Comments

The comments of the two teachers who taught the lesson follow:

> "Interestingly enough, many students in period one rated "Bedlam" 3 or middle. For the record, although I had informed the students they would be discriminating between 4 and a 5 paper, I made no effort to prejudice the students' individual or group ratings. Most of the students justified their middle ratings by suggesting that "Bedlam" is lacking in both supporting ideas and details. However, after I asked students to identify some of "Bedlam's" supporting ideas and details, most of the class admitted they had not analysed the piece as closely as they should have prior to rating it. In retrospect, most of the students agreed that "Bedlam", in terms of content, is superior to a Middle paper, though not a 5.
>
> (Teacher: A)

> "Students came alive today. They said that "Bedlam" is the best selection; it is the only one worth discussion. The imagery is vivid - and the vocabulary more to their taste. (Even if they do not act it out they like the sophistication of "adult" sounding vocabulary).
> They also enjoyed the task of comparison after. The change in attitude was remarkable. At the beginning of the class, they were about to rebel and riot, but at the end they were actually enjoying themselves.
>
> (Teacher: B)

Ratings

Table 7 presents the ratings for "Bedlam".

TABLE 8

RATINGS OF "BEDLAM" BY CLASS

Class	Ratings				
	1	2	3	4	5
A (n = 22)			9	10	3
B (n = 24)			2	10	10

The majority of students in class A, 13 students (59%), rated the paper high. The 9 pupils who rated it middle were probably the ones who, according to teacher A, "had not analysed" the piece closely. Three pupils rated it 5, but after discussion agreed with the ranking of the experts that "Lovely Room" was better. Forty-five per cent of the pupils placed the paper correctly the first time.

Class B was almost unanimous in rating the script high - 83% of the students gave a high rating. For this group, the pupils who rated the paper middle were outliers. Teacher B's comments suggest that pupils were reacting to other variables than content. In this case, the power of the language of the piece seemed to have overshadowed the more elaborated and focussed quality of "Lovely Room".

Editing

Most student samples were included without editing - as they came - so that pupils could practise the skill. Figure 12 presents a typical example of an edited script for this lesson. The editions of this student are used as a points of reference for discussing the editions of other students.

FIGURE - 12

"BEDLAM"

AN EXAMPLE OF AN EDITED SCRIPT

1: In my bedroom there is chaos. 2: Let us imagine a door that looks as though it has been lacerated by bullets from an M - 16 machine gun. 3: As you enter the door, the door begins to creak. Watch out for falling objects. 4: I call them objects because you have no idea they are is or how hard they going to hit you. 5: The room looks sort of misty like the jungles' of China; watch your step for the slippery rock formations (or the leather shoe pile) whichever you prefer. 6: Then if you look hard enough through the mist, you might see what appear to be dragon flies (or roaches) that seem to smile with you. 7: Then there is the swamp (or bed as you call it). Anything that is put on the top of it may never be seen again for a while. 8: The rest of the room resembles a time warp. Believe me, it is a moving experience.

This student found Bedlam and chaos in sentence 1 redundant, since both words in essence are synonyms. A few other students did too, but this was not the typical response. Other students left the sentence as it is - perhaps struck by the force of the repetition, and the rhythm of the line. These saw the phrase as one natural unit where one complement reinforced the other. But many pupils found an agreement problem. These changes is to are to make the verb agree with the subject.

Sentence 2 posed no particular difficulty, but a shift in movement of sentence 3 proved problematic for this and a few other students. Here, the pupils misread the sentence, misplaced the comma, and then found it necessary to repeat the wood door to make sense of the sentence. Most students inserted a comma after the introductory dependent clause. A few students left the sentence as it is. However, most pupils placed a period after creak, fracturing the original sentence and essentially cutting it in two.

The problem with the writer of the piece seems to be that at one level, he appears to be describing for a reader not present in the room; however, at another level he seems to be giving somebody physically present a guided tour. There are problems with distance on two other occasions - sentence 5 and sentence 7. The solution of making two simple sentences of sentence 3 certainly seems to take away some of the force and immediacy of the event of entering the room.

Sentence 4 presented no editorial problems for students. In sentence 5, this student uses the semi-colon to indicate the shift in distance between writer and reader. He also misuses the apostrophe, but no other student did so. The next major revision occurs at sentence 8. This student revised for economy and inserted a period after ways, but did not spot the comma splice after me. This was certainly an advance though because he revised, and not merely edited. No other student revised this part.

ORGANISATION

The format of the organisation booklet was similar. The professional model was from Naipaul's *A House for Mr. Biswas*. At the end of lesson 2, pupils learned a rule structure for organising that kind of writing. Lessons 3 through 5 taught pupils the features of papers rated middle or low. During week 1, the teachers played the dominant role, and students worked mainly in whole class format.

During week 2, groups played the dominant role. First, pupils in groups of 3 learned to distinguish between high and low papers. Then students in groups rated their own scripts within groups. After that, groups rated for content. Finally, the class wrote an in-class composition which the teacher rated for both qualities, and then added a mark for grammar and mechanics. Lesson 8 is included here as an example to illustrate how peer evaluation worked. The summary sheet and lesson plans are presented in Appendix D. Only the student samples for rating are given here.

Materials

Figure 13 presents the first of the two student samples for lesson 8.

FIGURE 13

STUDENT SAMPLE RATED FOR TWO QUALITIES

Instructions: Students were to write about a festive occasion at which they were present. They were writing an in-class composition for their teacher under examination conditions. It had been assigned during the pilot study, but used here in lesson 8 as a sample for rating.

1: There was a lot of preparation as our family was planning a reunion. 2: However, my mother decided that the main dish would be a piece of center cut steak along with potato stuffing and peas and rice.

3: The steak was seasoned and based with gravy master, red pepper, celery and onions. 4: I placed it into a baking dish, covered it with foil paper and placed it in the oven to bake at a temperature of 400 degrees F for an hour. 5: During that time, I cooked the rice and peas and made the potato stuffing. 6: The family started to gather and my sister set the table. 7: The steak had finished baking and was placed on a serving dish. 8: Furthermore, everyone gathered around the table and sat down. 9: We talked about the old days and how everyone had grown. 10: The food was passed around and each person took what they wanted.

11: Finally, one can say the food was rather impressive due to the fact that it was garnished and that the red pepper, celery and onions that were placed on it added color to it. 12: The gravy master that was placed on it helped to add a great deal of flavor to the steak.

Figure 14 presents the second student sample for lesson 8.

FIGURE 14

STUDENT SAMPLE RATED FOR TWO QUALITIES

Instructions: This was another response to the assignment described in Figure 13.

1: The roots wedding was organised to be as natural as possible. 2: The food was to be fresh from its natural source, untampered by chemicals.

3: The ceremony took place in the lush green forest, with bride and groom dressed in clothes made of green leaves. 4: The ceremony was simple. 5: Bridge and groom embraced and jumped over a broom into the land of holy matrimony. 5: The conversion to husband and wife merely by transferring from one side of the broom to the other.

7: Melodious singing of the birds in the background supplied the music. 8: The drinking began with the husband serving the wife a jelly coconut straight from the tree, after opening it for her. 9: This was followed by all the men until everyone had a drink.

9: Next, the wife found a mango tree, from which she served her husband and herself two of its ripe juicy fruits. 10: She was followed by the other women, and so everybody had something to eat.

11: Soon the celebration was in full swing, with people wandering all over the forest. 12: There were lots of succulent fruit with varying flavors and colors. 13: Everyone was free to pick and sample. 14: A brook nearby supplied cool, pure and refreshing water. 15: The water was served by the men.

12: The most impressive points to note that absolutely no animal flesh was consumed. 13: Everything was eaten in its natural state. 14: The men served all the drinks and the women the food.

14: An unusual celebration indeed!

TABLE 9

RATINGS OF "CENTER CUT" BY GROUP (N = 5)

Group	Ratings									
	Content					Organisation				
	1	2	3	4	5	1	2	3	4	5
1 (n = 3)			1	1			1	2		
2 (n = 4)				4				3		
3 (n = 4)				4			2	2		
4 (n = 4)			1	3			4			
5 (n = 5)				4	1		2	3		

The overall picture shows that 95% of the pupils did not give this paper a high rating. Students showed more agreement for content than for organisation. For content, there was 84% agreement on a middle or average rating for the paper; for organisation, it was only 53%.

The pattern by group helps to confirm the picture. Only 3 students (16%) were outliers - one student in group 5 being the most lenient. Groups 2 and 3 had perfect agreement, each pupil rating it average. However, the picture was different for organisation. The

3 pupils in group 2 who rated it 3 were in perfect agreement, but so were the 4 pupils in group 2 who rated it 2. Group 3 were split down the middle - 50% rating it 2, and 50% rating it 3.

After discussion to determine a group rating - as distinct from individual ratings - all five groups judged "Center Cut" average (3). The two groups who were divided, groups 3 and 5, adjusted their rating to a rating of low for organisation. So that the overall picture for group ratings was 100% for content, but 60% agreement or disagreement for organisation.

Table 10 presents the group ratings for "Roots Wedding".

TABLE 10

RATINGS FOR "ROOTS WEDDING" BY GROUP (N = 20)

Group	Ratings									
	Content					Organisation				
	1	2	3	4	5	1	2	3	4	5
1 (n = 3)					2					3
2 (n = 4)					3					3
3 (n = 4)				1	3				2	2
4 (n = 4)				3	1				1	3
5 (n = 5)				1	2					3

The overall picture for "Roots Wedding" shows that all students, and so all groups, agreed that this student sample was a high paper on both dimensions. However, there was some disagreement as to how high. Eight of the 33 pupils (24%) rated it 4, thinking it only suggested superiority.

The pattern of the sample by group again mirrors the overall pattern. For content, groups 1 and 2 were in complete agreement;

they thought the paper demonstrated superiority. But the other groups were not so sure. In group 4, 3 of the 4 students (75%) gave it a 4, indeed, 3 of the 5 pupils who rated it 4 came from group 4.

For organisation, there was more agreement. Three of the 5 groups (60%) were in complete agreement, rating it 5. Most pupils in group 4 (75%) agreed with that rating, but group 3 were equally divided.

After discussion, all groups gave the paper the highest rating on both content and organisation.

STRUCTURE

For the booklet on structure, the format was different for week 1. The researcher incorporated specific content material required by the grade 12 syllabus to reflect teacher emphases. The principles of relating and elaborating ideas within the sentence (Moffett 1966 and 1983), and sentence combining strategies were employed to meet the above needs.

Lessons 1 and 2 focussed on the development of sentence sense. Working from basic sentence patterns, students expanded sentences by adding phrases and clauses. Lesson 3 was an exercise in revising sample extracts for fragments and other errors. Lessons 4 and 5 centred on subject/verb agreement problems. By the end of lesson 5, students had internalised criteria for papers rated high and low. These criteria partly included error counts in the manner of the APU scale, but also stressed sentence variety and emphasis.

During lessons 6 and 7, students working as a class or in groups reinforced sentence variety concepts through sentence combining exercises. For lessons 8 and 9, pupils rated papers in groups - for high, lesson 8; and for low, lesson 9. In lesson 10, students rated their own compositions for structure in groups of four.

The sample lesson for this unit was chosen to illustrate teacher supplementary material. It develops the following phases: establishing new learning, consolidating new learning, and applying new

learning. It gives the instructor detailed directions about what to do, moment by moment. Scripts are provided at certain points in the lesson sequence. In some ways, student responses to teacher probes are idealised: appropriate answers are always given and much material is included. Only phases 1, 2 and 3 are illustrated because of the length of the lesson, and limited space. The abbreviated lesson plan is presented in Appendix D.

Towards the end of the unit, pupils rated their own compositions for structure, organisation and content. The student samples and ratings are presented here to show how one group from one class rated a paper for the three categories. Students chose their own topics for an expository exercise.

Figure 15 presents the student sample rated for three qualities.

FIGURE 15

STUDENT COMPOSITION RATED FOR THREE QUALITIES

Instructions: This was a paper in the expository mode. Students chose their own themes and could use the library for raw material if they wished. They were to share papers with their groups.

Different Types of Fads Diet

The advocate of eating raw foods almost exclusively can find some justification in the fact that cooking might reduce the amount of vitamin A and mineral content of certain foods. On the other hand, cooking naturally improves the flavor of many foods, but also makes them easily digested. Cooking is also an important method of protecting against diseases which are transmitted by foods. Raw foods are exposed to many possibilities of contamination, and may carry the causative organism of such diseases as typhoid fever, and trichinosis. Furthermore, the utilisation of many plant protein by the body is improved by cooking.

Vegetarians follow the fad of being a vegetarian, either because they consider animal foods deleterious to health, or because they can't stand the killing of animals to supply food for man. Although I feel that it is inadvisable for an individual to get involved with fads, there are some fads which have a certain amount of truth. It has been pointed out to us that fruits and vegetables are the best source of most of the vitamins and minerals which our body require. On the other hand, the body must have a certain amount of good quality protein as is done by herbivarious animals from the plant kingdom, but large qualities of carefully selected food would have to be consumed and even then their would be danger of deficiency.

For man, animal foods such as meat, eggs, milk and dairy products improve the palatability of the diet, and provide the protein which are most efficiently needed by the body.

Dieting is also another fad which some people get involved with whether they need to diet or not. I must say there is some truth to dieting, especially for the obese person. But usually dieting ends in crash dieting and this is where the problem starts. Even the obese person has to eat properly.

Table 11 gives the ratings of one group for the composition.

TABLE 11

GROUP RATINGS OF COMPOSITION FOR 3 QUALITIES

Name	Category		
	Structure	Organisation	Content
Avarel	4	3	4
Verlyn	4	3	4
Tessa	4	4	4
Claudette	4	5	4
Germaine	4	3	4

This group of five students had been working together for some time. They disagreed on the rating for organisation, but agreed on everything else. Some comments did not match the rating allocation. For example, Averal commented that "the paragraphs flowed smoothly from each other", and Verlyin that the piece was "organised fairly well". The rating of 3 by Germaine matched her comment: "needs little work". Claudette rated her own paper 5 and said, "It stuck to the point". Finally, Tessa thought the paper was "well developed". Her rating seemed a happy compromise.

WORDING

The format of this booklet was typical of the first two booklets. During lessons 1 and 2 students analysed an extract from J. Steinbeck's, The Grapes of Wrath to establish the rule structure. In lessons 3, 4 and 5 pupils internalised the features of papers rated high, middle and low, respectively.

In lessons 6 and 7, teachers and students analysed an essay rated high. In lesson 8, students in groups rated their own papers for wording; in lesson 9, they rated for structure, and in lesson 10 they rated for all four qualities. The illustrative lesson is an example of a group exercise to tease out the features of a paper rated low. The second section gives the ratings of a composition by one class for all four qualities. The writing sample and ratings are presented here. The exercise is presented in Appendix D.

The writing sample and ratings are presented here. The exercise is presented.

FIGURE 16

STUDENT SAMPLE RATED COMPOSITION FOR ALL QUALITIES

Water, Water Everywhere

1: On Wednesday, the day before Thanksgiving, the rain fell very hard. 2: This was the ;most it has rain during this entire year. 3: While going to sleep that night, a siazable rain was falling but when my mother woke me in the morning the rain sounded like a storm. 4: When I stepped out of bed, my feet landed in a pool of water. 5: I turned on the light only to see my shoes floating across the room. 6: I tip-toed out to the living room only to find out that our entire home was flooded. 7: Our brand new carpet was soaked. 8: When you stepped on it, your foot would sink, and make a splashing sound.

9: My mother had already begun to sweep the water over the bnack porch, when she told me to sweep out the bedrooms. 10: When I walked into the bedroom I realized that many of my furniture were already destroyed. 11: The lower part of my dresser had already begun to peel. 12: The book which wrre on the lowest level of my bookshelf, were already soaked. 13: The curtains which were hanging infront of my windows and those which were infront of my closet were wet. 14: I had a basket with some clothes on the ground that too was half wet. 15: The water was about half an inch off of the ground.

16: My mother and I spent half of Thanksgiving Day not cooking, but cleaning.

TABLE 12

INDIVIDUAL RATINGS OF WRITING SAMPLE

Content n = (24)	Organisation n = (24)	Structure n = (23)	Wording n = (24)
			1 1 1
2 2 2	2 2 2 2 2	2 2 2 2 2	2 2 2 2 2
	2 2	2 2 2 2 2	2 2 2 2 2
			2 2 2 2 2
3 3 3 3 3	3 3 3 3 3	3 3 3 3 3	3 3 3 3 3
3 3 3 3 3	3 3 3 3 3	3 3 3 3 3	3
3 3 3 3 3	3 3 3 3 3	3 3	
3 3 3 3 3	3		
4	4	4	

The overall picture for content shows that 83% of the pupils gave the script a middle rating; three students (13%) failed it, and one student thought it a high paper. This last girl who rated it 4 was consistent in scoring high for all qualities. The pattern of rating for organisation showed slightly more disagreement. Sixty-seven per cent thought it was average, so four of the students who passed the paper for content failed it for organisation. A total of twenty-nine per cent failed it.

Pupils were to rate as well as comment on the structure and wording of the composition. Twelve pupils (52%) passed it for structure; 10 (43%) failed it. The comment of the girl who rated it 4 was "good grammar; no spelling errors". This casts some doubt on her analysis of the script. Most pupils gave a catalogue of errors: faulty sentences, agreement problems, spelling mistakes, bad punctuation and faulty sentence patterns. One commentator wrote with some candour: "need help!"

Only 6 students (25%) gave the paper an average score for wording. Put another way, 18 pupils (75%) failed the paper. It is to be noted that even the student who was the easiest rater of this script gave it a middle rating. Some of the comments were as follows: "too little detail," not very expressive," "this paper has a very simple style of writing," "not enough imagery words," and "a bit too plain."

CHAPTER EIGHT

ANALYSIS OF THE DATA

The post-test data were analysed to discover answers to the following three sets of questions. The first set of questions was as follows:

(1): Did the teaching programme improve the overall writing quality of the compositions of pupils in the Experimental group?

(2): Did the programme improve the quality of content in the compositions of pupils in the Experimental group?

(3): Did the programme improve the quality of organisation in the compositions of pupils in the Experimental group?

(4): Did the programme improve the quality of structure in the compositions of pupils in the Experimental group?

(5): Did the programme improve the quality of wording in the compositions of pupils in the Experimental group?

A second set of questions was concerned with the inter-rater agreement among students taught to use the scale. The second set of questions was as follows:

(6): What was the inter-rater reliability for students on the total composition score?

(7): What was the inter-rater reliability for students on the sub-scale of content?

(8): What was the inter-rater reliability for students on the sub-scale of organisation?

(9): What was the inter-rater reliability for students on the sub-scale of structure?

(10): What was the inter-rater reliability for students on the sub-scale of wording?

A third set of questions was concerned with indirect assessment of writing. The third set of questions was as follows:

(11): Was there a significant difference in performance on the English Expression STEP test between the students in the Experimental group and those in the Control group?

(12): Was there a significant difference in performance on the Mechanics of Writing STEP test between the students in the Experimental group and those in the Control group?

The final set of questions asked for attitudinal responses - the emphasis was affective rather than cognitive. The final set of questions was as follows:

(13): What was the general attitude of students about writing at the end of the programme?

(14): How did they feel especially about

(a): working in groups?

(b): rating their own papers?

(c): rating the papers of their peers?

Measures

Pre-test

All students responded to the following measures:

(1): The Sequential Tests of Educational Progress (STEP): English Expression: Form 2A.

(2): The Sequential Tests of Educational Progress (STEP): The Mechanics of Writing: Form 2A.

(3): The Daly-Miller Writing Apprehension Measure.

(4): One composition.

Post-test

Students responded to the following measures:

(5): The Sequential Tests of Educational Progress (STEP): English Expression: Form 2B.

(2): The Sequential Tests of Educational Progress (STEP): The Mechanics of Writing: Form 2B.

(3): The Daly-Miller Writing Apprehension Measure.

(4): One composition.

In addition, the students in the Experimental group rated five (5) compositions for all four (4) qualities: content, organisation, structure, and wording.

A sample of 98 students drawn from a population of 575 grade 12 students was divided into Experimental and Control groups with one Experimental and one Control group in each of two schools. The Experimental group at first comprised 50 students: the Control group comprised 38 students. Data were analysed employing analysis of co-variance techniques in order to equate the groups statistically in case of any pre-Experimental differences on three co-variates. However, only raw scores for students who completed the full battery of tests were used in the analysis. There was no significant difference between the post-test means of the two groups on the measures of English expression, mechanics of writing and writing apprehension, so the ANCOVA analyses are not reported.

Pre-test Data

Table 13 presents the means and standard deviations for the pre-test composition scores, STEP scores for English Expression and the Mechanics of Writing, and the Daly-Miller Writing Apprehension Measure.

Table 13 shows that the two groups differed before the programme was conducted. The Experimental group had a higher mean score on the writing test; there was a difference score of .51 between the means. The Experimental group also had a higher mean score on the mechanics of writing test with a difference score of 1.17. However, the Control

TABLE 13

PRE-TEST STATISTICS FOR THE EXPERIMENTAL AND CONTROL GROUPS

	Experimental Group (N=41)		Control Group (N=41)	
	Mean	S.D.	Mean	S.D.
Composition	24.24	4.72	23.73	4.93
STEP[a]	18.51	4.99	19.98	5.07
STEP[b]	22.66	5.47	21.49	5.47
Appre-hension	90.00	19.37	92.12	14.04

[a] The Sequential Tests of Educational Progress (STEP): English Expression: Form 2A.

[b] The Sequential Tests of Educational Progress (STEP): The Mechanics of Writing: Form 2A.

group had a higher mean score on the test of English expression with a difference score of 1.47 between the groups. As noted earlier, these differences were accounted for in the analysis, so that the data were analysed as if the groups had been equal before the programme started.

Neither group could be classified as low or high apprehensives. The instrument has a maximum score of 130 points, and a minimum score of 26. For students scored below or above the mean in any of the two groups. The post-test difference in means was not significant either.

Table 14 presents the means and standard deviations for pre-test composition scores and sub-scores. An examination of these data indicates that the Experimental group had higher mean scores on three dimensions of writing.

TABLE 14

PRE-TEST STATISTICS FOR COMPOSITION SCORES AND SUB-SCORES

	Experimental Group (N=41)		Control Group (N=41)	
	Mean	S.D.	Mean	S.D.
Total score[a]	24.24	4.72	23.73	4.93
Content[b]	6.14	1.61	6.20	1.40
Organisation	5.54	1.64	5.27	1.63
Structure	6.22	1.29	6.02	1.25
Wording	6.34	1.28	6.24	1.24

[a]Maximum score = 40

[b]Maximum score = 10

The Control group had a slightly higher mean score on content - an increment of .06. But the Experimental group had higher mean scores on the other categories with the following increments: organisation, .27; structure, .20; and wording, .10. Note again that these initial differences were cancelled out statistically since analysis of co-variances procedures were used.

The ANOVA test is a test of significance between group means. Differences between group means could be attributed in part to the variability within the groups themselves. Statisticians call this "error" because it represents temporary, chance and unsystematic variables not controlled for by the researcher. Differences between groups could also be attributed to lasting and systematic factors resulting from the influence of independent variables, for example, teaching method. Analysis of variance takes into account the two groups of variance. The sum of squares between groups is divided by the sum of squares within groups to give the mean square. This is then tested for significance using the appropriate table.

Briefly, in ANOVA analysis there is an attempt to control for certain variables that may be related to the variable(s) under study. Pre-experimental differences between groups could confound the results of the experiment if they are not controlled. Since it is very difficult for the researcher to use groups in an actual school setting, groups are made equal by statistical manipulation of data through analysis of co-varianced procedures after the experiment.

The scores on the control variables, obtained before the experiment, are used to adjust the scores on the post-test using regression analysis.

For example, in this investigation the actual post-test scores obtained for the dependent variable under study were adjusted to eliminate pre-test advantages or disadvantages on the pre-test variables. Before the programme started, the two groups were different in writing ability, use of English expression, and mechanics of writing. To equate the groups statistically, the actual post-test scores obtained were adjusted to compensate for the inequality of groups on those variables. For example, if a group had a higher mean score (and thus an advantage) on any of the control variables, the post-test mean of that group would be adjusted downwards, and the mean for the other group adjusted upwards to equate the two groups on that variable. After these adjustments were made, an analysis of variance was calculated from the residual data, or adjusted sums of squares, degrees of freedom, and mean squares.

Effect of the Programme

Total Score

The data used in the analysis were from 82 grade 12 students: forty-one were from the Experimental group, and 41 from the Control group. The first null hypothesis to be tested was:

> "that there would be no significant difference between the post-test total composition scores of two groups - the Experimental group using a specially-designed programme, and the Control group using the school programme - when those groups had been equated statistically for initially differences in writing ability, English expression, and mechanics of writing scores.

The criterion or dependent variable was writing ability; the control variables were previous performance on composition, English expression, and mechanics of writing measures; the independent variable was instructional programmes represented by the two groups.

The data were analysed using the SPSS/PC+ statistical software programme for the IBM computer. Sums of squares are decomposed in nine different ways. The researcher chose option 8. This option adjusts the data first for the main effects, then for the main effects and co-variates taken together, and finally for the first two plus interactions. Interactions were not considered in the analysis. The use of option 8 also produces a multiple classification analysis table (MCL) which displays unadjusted deviations, adjusted deviations for the independent variable(s), and the adjusted deviations for the independent variable(s) and co-variates from the grand mean.

Table 15 presents the Experimental and Control group post-test means for the data. The raw scores for the summarised data in this section are to be found in Appendix B.

TABLE 15

EFFECT OF THE PROGRAMME ON THE OVERALL QUALITY OF COMPOSITION BASED ON POST-TEST ADJUSTED MEANS[a]

(Unadjusted Means in parentheses)

Group		Difference	F	P
Experimental (N=41)	Control (N=41)			
26.48[b]	24.03	2.45	5.849	.018
(26.34)	(24.17)	(2.17)		

[a] Adjusted for pre-experimental differences in writing ability, STEP English expression and mechanics of writing.

[b] Maximum score 40.

The adjusted deviations for the independent variables plus co-variates were - 1.22 for the Control group, and 1.22 for the Control group. The adjusted mean for the Experimental group was calculated by adding the overall deviation to the grand mean; the adjusted mean for the Control group was calculated by subtracting the deviation from the grand mean. This was the same as making adjustments from the pre-test control variable means separately using the raw regression co-efficients.

This was done by using the regression co-efficient for each variable and multiplying it by the difference between the mean of the individual groups and the mean of the groups combined. The result was subtracted or added to the criterion mean depending on which group had a higher mean score at the beginning.

For example, the Experimental group was superior on composition at the beginning, so .08 was subtracted from 26.34, the mean of that group; but added to 24.17, the mean of the Control group. The latter group was superior on the English expression test, so .26 was added to the mean of the Experimental group and the same amount taken from that of the Control group. Finally, .04 was subtracted from the mean of the Experimental group and added to that of the Control group. This process gave the same overall result as subtraction from the grand mean.

Table 16 presents the analysis of co-variance data, source table for Table 14.

TABLE 16

ANALYSIS OF COVARIANCE OF POST-TEST TOTAL COMPOSITION SCORES FOR TWO GROUPS

		Residuals		
Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Between	1	96.598	96.598	5.849
Within	77	1271.642	16.515	
Total	81	2033.81		

[a] Significant at the .02 level

This suggested that the use of the Experimental programme resulted in more growth than that of the school programme. Thus, the null hypothesis was not accepted.

Effect of the Programme

Content

The second null hypothesis to be tested was that there would be no significant difference between the post-test content scores of two groups - the Experimental group using a specially designed programme, and the Control group using the school programme - when those groups had been equated statistically for initial differences in writing ability, English expression, and mechanics of writing scores.

The criterion or dependent variable was content; the control variables were previous performance in composition, English expression, and the mechanics of writing measures; the independent variable was instructional programmes represented by the two groups.

Table 17 presents the experimental and control group means for the data.

TABLE 17

EFFECT OF THE PROGRAMME ON THE OVERALL QUALITY OF CONTENT BASED ON POST-TESTED ADJUSTED MEANS[a]

(Unadjusted Means in parentheses)

Group		Difference	F	P
Experimental (N=41)	Control (N=41)			
6.97[b]	6.01	.96	9.584	.003
(6.93)	(6.05)	(.88)		

[a] Adjusted for pre-experimental differences in writing ability, STEP English expression and mechanics of writing.

[b] Maximum score 10.

Table 17 shows an unadjusted mean difference score of .88 favouring the Experimental group, and an even higher difference score of .96. This yielded an ^{F} ratio of 9.564 which had a ^{p} value of .003.

Table 18 presents the analysis of co-variance data.

TABLE 18

ANALYSIS OF COVARIANCE OF POST-TEST COMPOSITION (CONTENT) SCORES FOR TWO GROUPS

		Residuals		
Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Between	1	15.805	15.805	9.584
Within	77	126.979	1.649	
Total	81	142.784		

[a] Significance at the .003 level

This suggested that the use of the experimental programme resulted in more growth than that of the school programme. Thus, the null hypothesis was not accepted.

Effect of the Programme

Organisation

The third null hypothesis to be tested was that there would be no significant difference between the post-test organisation scores of two groups - the Experimental group using a specifically-designed programme, and the Control group using the school programme - when those groups had been equated statistically for initial differences in writing ability, English expression, and mechanics of writing scores.

The criterion or dependent variable was organisation; the control variables were previous performance in composition, English expression, and mechanics of writing measures; the independent variable was instructional programmes represented by the two groups.

Table 19 presents the Experimental and Control group means for the data.

TABLE 19

EFFECT OF THE PROGRAMME ON THE OVERALL QUALITY OF ORGANISATION BASED ON POST-TEST ADJUSTED MEANS[a]

(Unadjusted Means in parentheses)

Group		Difference	F	P
Experimental (N=41)	Control (N=41)			
6.236[b] (6.22)	5.496 (5.51)	.74 (.71)	4.757	.032

[a] Adjusted for pre-experimental differences in writing ability, STEP English expression and mechanics of writing.

[b] Maximum score 10.

Table 19 shows an unadjusted mean difference score of .71 favouring the Experimental group, and an even higher difference score of .74. This yielded an f ratio of 4.757 which had a p value of .032.

Table 20 presents the analysis of co-variance data.

TABLE 20

ANALYSIS OF COVARIANCE OF POST-TEST COMPOSITION (ORGANISATION) SCORES FOR TWO GROUPS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	P
Between	1	10.256	10.256	4.757	.032[a]
Within	77	166.004	2.156		
Total	81	176.26			

[a] Significant at the .03 level

This suggested that the use of the Experimental programme resulted in more growth than that of the school programme. Thus, the null hypothesis was not accepted.

Effect of the Programme

Structure

The fourth null hypothesis to be tested was that there would be no significant difference between the post-test scores for structure of two groups - the Experimental group using a specially designed programme, and the Control group using the school programme - when those groups had been equated statistically for initial differences in writing ability, English expression, and mechanics of writing scores.

The criterion or dependent variable was structure; the control variables were previous performance in composition, English expression, and mechanics of writing measures; the independent variable was instructional programmes represented by the two groups.

Table 21 presents the experimental and control group means for the data.

TABLE 21

EFFECT OF THE PROGRAMME ON THE OVERALL QUALITY OF STRUCTURE BASED ON POST-TEST ADJUSTED MEANS [a]
(Unadjusted Means in parentheses)

Group		Difference	F	P
Experimental (N=41)	Control (N=41)			
6.47[b] (6.46)	6.256 (6.27)	.22 (.19)	.696	.407

[a] Adjusted for pre-experimental differences in writing ability, STEP English expression and mechanics of writing.

[b] Maximum score 10.

Table 21 shows an unadjusted mean difference score of .19 favouring the Experimental group, and slightly higher difference score of .22. This yielded an F ratio of .696 which was not significant. The null hypothesis was not rejected.

Effect of the Programme

Wording

The fifth null hypothesis to be tested was that there would be no significant difference between the post-test scores for wording of two groups - the Experimental group using a specially designed programme, and the Control group using the school programme - when those groups have been equated statistically for initial differences in writing ability, English expression, and mechanics of writing scores.

The criterion or dependent variable was wording; the control variables were previous performance in composition, English expression, and the mechanics of writing measures; the independent variable was instructional programmes represented by the two groups.

Table 22 presents the Experimental and Control group means for the data.

TABLE 22

EFFECT OF THE PROGRAMME ON THE OVERALL QUALITY OF WORDING BASED ON POST-TEST ADJUSTED MEANS [a]

(Unadjusted Means in parentheses)

Group		Difference	F	P
Experimental (N=41)	Control (N=41)			
6.804[b] (6.73)	6.244 (6.32)	.56 (.41)	2.873	.094

[a] Adjusted for pre-experimental differences in writing ability, STEP English expression and mechanics of writing.

[b] Maximum score 10.

Table 22 shows an unadjusted mean difference score of .41 favouring the Experimental group, and a slightly higher difference score of .56. This yielded an f ratio of 2.873 which was significant, but not at the alpha set.

Table 23 presents the analysis of co-variance data.

TABLE 23

ANALYSIS OF COVARIANCE OF POST-TEST COMPOSITION (WORDING) SCORES FOR TWO GROUPS

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Between	1	3.5234	3.254	2.873
Within	77	94.466	1.227	
Total	81	126.451		

[a] Significant at the .09 level

This suggested that the use of the Experimental programme resulted in more growth than that of the school programme. Thus, the null hypothesis was not accepted.

Summary

A comparison of the Experimental and Control groups based on post-test composition mean scores yielded the following results:

(1): The difference between groups in overall writing ability was statistically significant in favour of the Experimental group;

(2): The difference between groups in overall quality of content was statistically significant in favour of the Experimental group;

(3): The difference between groups in overall quality of organisation was statistically significant in favour of the Experimental group;

(4): The difference between groups in overall quality of structure was not statistically significant;

(5): The difference between groups in overall quality of wording was statistically significant in favour of the Experimental group but not at the alpha set.

Children as Raters

Inter-reliability

One important feature of the writing programme was that pupils used a rating scale to judge their own compositions and those of their peers. Pupils worked in groups of two, three, four or five to rate papers first for one quality, then two qualities, then three, and finally all four qualities. During the post-testing period, students in the Experimental group had to rate five compositions for all four qualities of the scale. A total of 49 sets of ratings was collected. A random sample of 15 sets was drawn and an inter-reliability coefficient for total score as well as subscores for content, organisation, structure and wording calculated.

The researcher used a version of Kuder-Richardson formula 20 which is a special use of Cronbach's coefficient alpha (Ebel, 1979; Ebel & Frisbie, 1986). The formula is:

$$r = \frac{K}{K-1} \left[1 - \frac{\Sigma \sigma_i^2}{\sigma_t^2} \right]$$

k = number of independent ratings of a performance

σ_i^2 = variance from a particular rater

$\Sigma\sigma_i^2$ = sum of rater variances for all raters

σ_t^2 = variance of the sums of ratings from all raters

Estimate of Reliability

The Whole Scale

Table 24 presents the summarised data on which the calculations were passed. The raw scores for these and all other data in this section are presented in Appendix B.

TABLE 24

INTER-RATER RELIABILITY COEFFICIENT FOR THE WHOLE SCALE

(N=15)

Source of Variation	Sum of Squares	Variance
From rater scores	11557	
From essay totals	169819	2776.24
From rater totals	52291	219.76

$$r = \frac{K}{K-1} \left[1 - \frac{\Sigma \sigma_i^2}{\sigma_t^2} \right]$$

$$= \frac{15}{14} \left[1 - \frac{219.76}{2776.24} \right]$$

$$= .99$$

Estimate of Reliability

Content Component

Table 25 presents the summarised data on which the calculations were based.

TABLE 25

INTER-RATER RELIABILITY COEFFICIENT FOR THE CONTENT COMPONENT OF THE SCALE

(N=15)

Source of Variation	Sums of Squares	Variance
From rater scores	823	
From essay totals	11599	185.36
From rater totals	3643	18.88

$$r = \frac{K}{K-1}\left[1 - \frac{\Sigma\sigma_i^2}{\sigma_t^2}\right]$$

$$= \frac{15}{14}\left[1 - \frac{18.88}{185.36}\right]$$

$$= .96$$

Table 26 presents the summarised data on which the calculations were based.

TABLE 26

INTER-RATER RELIABILITY COEFFICIENT FOR ORGANISATION COMPONENT OF THE SCALE

(N=15)

Source of Variation	Sums of Squares	Variance
From rater scores	772	213.6
From essay totals	10748	213.6
From rater totals	3297	22.72

$$r = \frac{K}{K-1}\left[1 - \frac{\Sigma\sigma_i^2}{\sigma_t^2}\right]$$

$$= \frac{15}{14}\left[1 - \frac{22.72}{213.6}\right]$$

$$= .96$$

Estimate of Reliability

Structure

Table 27 presents the summarised data on which the calculations were based.

TABLE 27

INTER-RATER RELIABILITY FOR THE STRUCTURE COMPONENT OF THE SCALE

of the Scales

(N=15)

Source of Variation	Sums of Squares	Variance
From rater scores	681	
From essay totals	9417	136.16
From rater totals	2989	16.64

$$r = \frac{K}{K-1}\left[1 - \frac{\Sigma\sigma_i^2}{\sigma_t^2}\right]$$

$$= \frac{15}{14}\left[1 - \frac{16.64}{136.16}\right]$$

$$= .94$$

Estimate of Reliability

Wording

Table 28 presents the summarised data on which the calculations were based.

TABLE 28

INTER-RATER RELIABILITY DATA FOR THE WORDING COMPONENT OF THE SCALE

(N=15)

Source of Variation	Sum of Squares	Variance
From rater scores	777	
From essay totals	10875	185.84
From rater totals	3419	18.64

$$r = \frac{K}{K-1}\left[1 - \frac{\Sigma\sigma_i^2}{\sigma_t^2}\right]$$

$$= \frac{15}{14}\left[1 - \frac{18.64}{185.84}\right]$$

$$= .96$$

Summary

Inter-rater reliability co-efficients for a sample of student ratings using the analytic scale were calculated. The researcher found that:

(1): The coefficient for the total score was .99.

(2): The coefficient for the content subscore was .96.

(3): The coefficient for the organisation subscore was .96.

(4): The coefficient for the structure was .94.

(5): The coefficient for the wording was .96.

This suggests that students used the rating scale devised by the researcher consistently and reliably.

Indirect Assessment

English Expression

The hypothesis to be tested was:

> that there would be no significant difference between the post-test English expression scores of two groups - the Experimental group using a specially designed programme, and the Control group using the school programme - when those group had been equated statistically for initial differences in writing ability, English expression, and mechanics of writing scores.

The criterion or dependent variable was English expression; the control variables were previous performance in composition, English expression, and the mechanics of writing measures; the independent variable was instructional programmes represented by the two groups.

Table 29 presents the means of the two groups.

TABLE 29

EFFECT OF THE PROGRAMME ON THE OVERALL QUALITY OF ENGLISH EXPRESSION BASED ON POST-TESTED ADJUSTED MEANS[a]

(Unadjusted Means in parentheses)

Group		Difference	F	P
Experimental (N=41)	Control (N=41)			
20.07[b]	19.93	.14	.00	1.00
(20.00)	(20.00)	(.00)		

[a] Adjusted for pre-experimental differences in writing ability, STEP English expression and mechanics of writing.

[b] Maximum score 40.

Table 29 shows an unadjusted mean difference score of .00 favouring neither group, and an adjusted difference score of .14 in favour of the Control group. This yielded an f ratio of .00 which had a p value of 1.00. Thus, the null hypothesis could not be rejected.

Mechanics

The hypothesis to be tested was:

> that there would be no significant difference between the post-test mechanics of writing scores of two groups - the Experimental group using a specially-designed programme, and the Control group using the school programme - when those groups had been equated statistically for initial differences in writing ability, English expression, and mechanics of writing scores.

The criterion or dependent variable was mechanics of writing; the control variables were previous performance in composition, English expression, and the mechanics of writing measures; the independent variable was instructional programmes represented by the two groups.

Table 30 presents the means of the two groups.

TABLE 30

EFFECT OF THE PROGRAMME ON THE OVERALL QUALITY OF MECHANICS OF WRITING BASED ON POST-TEST ADJUSTED MEANS[a]

(Unadjusted Means in parentheses)

Group		Difference	F	P
Experimental (N=41)	Control (N=41)			
23.37[b]	23.73	.36	.034	.85
(23.63)	(23.37)	(.26)		

[a] Adjusted for pre-experimental differences in writing ability, STEP English expression and mechanics of writing.

[b] Maximum score 45.

Table 30 shows an unadjusted mean difference score of .26 favouring the Experimental group, and an adjusted difference score of .36 in favour of the Control group. This yielded an f ratio of .034 which had a p value of .85. Thus, the null hypothesis could not be rejected.

Free Response to the Programme

At the end of the study, the Experimental group was requested to write a letter to the researcher responding to the programme in general, but to group work in particular. This was different from the method used in the pilot study. The researcher reasoned that the letter would be a more personal and less restricting medium than the Likert-type questionnaire used in the pilot study, even though that instrument included a section for free response as well.

The questionnaire method has definite advantages as a data collection instrument and may provide for a more systematic response since students answer the same questions and analysis is more directed. However, the letter offered a stronger sense of audience and the probability of honest and sincere responses from pupils because they had met the researcher as teacher throughout the period of study. Twenty-seven letters were collected - a response rate of 55%.

In the first part of the analysis, these responses are analysed for pupil response to the programme of study. Then student comments on whether their work had improved are examined. After that, discussion centres on the effectiveness of group work as a methodology, and the pros and cons of peer and solo rating of essays. Finally, a summary relates the attitudinal data of the pilot and main studies.

General Comments

Attitude

Students' general response to the programme, as in the pilot study, was reflected in part by the vocabulary. Positive responses were signalled by such words as "excellent" - "helpful" - "tremendous help" - "remarkable idea" - "grading was wonderful work". Another reflection of positive affect was the unsolicited endorsements of the programme. Nine students recommended the programme. Gretta felt it should be "part of English education...in every high school." Jermaine said, "it was a pleasure to experience"; Ludrick that it should be instituted "at all grade levels". All felt the programme should continue so that other pupils could benefit.

Still other students admitted that at first they had had preconceived notions about the programme, but indicated a positive change in attitude after the programme started. Lencia wrote: "At first, I thought the programme would have been boring ...but, as the programme progressed I began to like it." Diahann was candid: "Before I started this programme, I really didn't like writing so I thought the programme was boring. But now I understand what the programme was about, and started to write more often, I began to like it." Gretta noted: "I must admit in the beginning this programme was boring, but as we progressed and have written compositions of our own, I began to understand and appreciate the writing programme."

Certain features of the programme, then, appeared to have brought a desirable change in the attitude of some students. Lencia did not

spell out the characteristics that appealed to her. However, Diahann noted that as she understood the intent of the programme and wrote more often, she started to like it. Gretta too thought that writing her own compositions was preferred to studying the found writing samples in the student booklets. This confirms the view that pupils should work with their own writing as soon as possible; it may not be enough to use other student material. Another factor was preparation of the students for collaborative learning in an attempt to change expectations.

This study used writing samples collected from the school before the study. These were essays written by grade 12 pupils - essays which the researcher analysed to determine the range of student writing ability and gauge expected standards. Samples were also collected before the study from selected grade 12 classes on topics suggested by the researcher. So all samples used in the study were student material written by students - except for a professional model that introduced each unit.

During the programme, pupils wrote two papers every unit - a total of eight compositions during the study. They worked with these papers to practise editorial skills, peer grading and revising. However, pupils worked for approximately six lessons per unit before using their own work! Diahann's comments suggest that perhaps the work of students should be introduced from the beginning.

One pupil did not respond positively to the programme. In a delightful and well-written letter, Kimberley said that for her the programme was "like a Rubik cube. I didn't get it at all. Nevertheless, there were times when I could see where you were coming from." For her, the materials were suitable for English or journalism majors and by implication, not for pupils reading a general English course. Kimberley found problems with rating papers and did not respond well to group work in general. However, her writing performance improved from a score of 18 (*M*=23.99) on the pre-test to a score of 26 (*M*=25.26) on the post-test. Her scores on the English Expression test increased from 20 (*M*=17.24) to 23 (*M*=20.00). Kimberley's scores on the mechanics of writing test were 30 (*M*=22.05) and 25 (*M*=23.55) score points respectively on pre-test and post-test. These data suggest that Kimberley was an above-average student by the end of the experiment.

Writing Improvement

A set of pupils said they thought their writing had improved. Ann said that as a senior she had not been writing at senior level, but noted: the programme "helped me to make a lot of changes about my writing style." Avanel was quite enthusiastic:

> "I have developed a sense of writing that I thought I could never do. Before this programme began, I didn't really like writing, but now I have just about mastered the art of writing. To me, writing is a great experience because you get to express your ideas about certain things."

Wadada too was conscious of the way the programme helped him. He wrote: "It helped me express myself more freely in writing essays and compositions." He said that his grade point average went up, that the course had made him more imaginative, and when you are more imaginative "your a better writer". Vincia said she now had confidence in writing compositions. Claudette wrote that the programme was a tremendous help. She said: "My writing has gotten better, I can spot my mistakes quicker than I use to before." Finally, Rodica found the approach "full-filling". The class, she noted, had made her more aware of her writing abilities. No student said his or her writing had not improved.

It seems then that most students can accommodate peer evaluation and group work into their learning set. However, students like Kimberley might not be altogether comfortable with the method. She might have been affected by group processes. Beaven notes that groups could spend as much as half the time on process and half on task. Kimberley suggested the problem might have been in the structure of the programme - like a Rubik cube. Could it be that the programme affected her writing more than she thought? Her gain scores certainly point to that. And if so, which aspect of the programme?

Group Work

Benefits

Bearing in mind that they said so themselves, pupils in the Experimental group reported definite advantages for group work as a sound pedagogical strategy. Durban said he "learned how to work

together as a group ... as well (as) getting to know each other." The idea that students must learn how to collaborate is important. The second point that a group interaction can improve human/social relationships was also made by Germaine. "I have learned how to <u>get</u> along with others in a group. I have learned how to express my feelings to others <u>without fear</u>." The programme certainly seemed to have affected Germaine's self-concept and confidence positively. Lencia illustrated the potential for growth when a climate of trust develops in the context of peer evaluation. She commented "I feel that honesty and trustworthiness may come about between your peers and yourself. Also, it gives you a chance to share ideas with each other ... Confidence and self-esteem has developed within me, and I highly recommend this programme to anyone who is interested in any sort of English." Ann noted that she learned a variety of things. Among other things, she "learned to cope with people who I was working with, and I learned the importance of writing for the future." Finally, some students felt the programme was good preparation for college.

<u>Disadvantages</u>

However, a few students said that they did not enjoy group work for various reasons. Difficulties centered around process and peer rating. Ann recounts her experience of group work:

> "group members were difficult to deal with. Most of the <u>burden</u> went on one or two people in the group. It was really a strain to rate four or five papers for one to four criteria with no/little help from your group work. This made me not want to work in groups any more."

This warns us that, although group work is potentially effective as a teaching/learning strategy, the teacher has to be extremely conscious of his role of seeing that the burden of responsibility for completing tasks is shared. This becomes more important when the use of rating scales is new and the process challenging.

Derick called pupils who do not pull their weight parasites.

> "Sometimes an individual becomes a parasite, getting a grade for no input on his behalf ...However, working with a group that reasons, and give input helps an individual. Justifications for given reasons will not be limited to one person but among others.

Students must also learn how to resolve disputes over ratings and work for consensus. They do so best when they learn to follow Derick's advice:

> "Rating is not easy, but once I've learned how to field out the factors, it's much easier. There are several categories to rate a paper, but once reasons for rates are known, justification can easily be made."

Kimberley thought she saw bias operating. In her vigorous and vivid style, she commented:

> "the rating part is what really stunk... What really got me steamed was when we were told to try our hand at writing a composition and put into groups to be rated...You think you paper is a middle - others too - but someone would be the sore thumb or give you a really low.

> Arguments would break out. Insults *would* fly. And boom the whole group would *blow* into a *zillion pieces* (italics added).

Dazle observed that "sometimes if you have a high paper and your friends do not want your grade to be higher, they would rate your paper even as their's or lower." Thus, although some pupils enjoy rating and consider it a learning experience, as Ann noted, it is not exciting at all times. For a few, the programme opened new vistas and possibilities. Faye found that using the scale was the most fun of all; Jerry found it very exciting." For a while, he said, "I felt like Mr. Penn or ever Professor Parris." Rodica liked "playing teacher". The opportunity to do what teachers do - to enter their sanctum sanctorum - could be very motivating for students. Some said that rating papers themselves took some of the mystery out of what teachers do. Witness Deborah who said that she had a better understanding of what a teacher looks for in grading composition; Claudette said of the experience, "I am now able to write a high paper without doubting for one minute that it is high."

Rating Own Papers

Two students made comments on the advantages of rating their own paper. Derick acknowledged that rating one's own paper was sometimes difficult, and one did not always feel confident about one's work, and would like others' opinions. Rochelle found grading her own scripts rewarding. She was able to see personally what she needed to improve. Using the scale she said, "made me able to criticise my own work and know it was constructive criticism."

Comparisons

It could be instructive to compare responses to the pilot study, the main study, and the Harris study on the categories of group work and peer rating. Comparisons must remain gross, however. In the pilot study, 14 students responded to a questionnaire; in the main study, 26 pupils gave free responses; in the Harris study, 70 students responded to a 13-item questionnaire on issues involved in peer/self-evaluation. So, the numbers were small and in one case the information was collected differently, but it is still interesting to see how students from different populations responded to the use of rating scales in groups to rate essays.

One issue involved is whose paper is being evaluated? Some students seem to have difficulties rating their friends' papers. During the pilot study, student F observed: "It is difficult rating your friend's paper. If your friends do not want your grade to be higher, they would vote your paper even as their's or lower."

Harris posed a specific question on this issue - "Do you prefer to have your paper evaluated by (a) a friend, (b) someone you hardly know?" Nineteen preferred a friend, but 44 out of 70 students preferred their paper to be rated by someone they hardly knew. Seven pupils did not care.

Still another issue is that of writing improvement. Harris asked: "Is your grade (a) higher, (b) lower or (c) the same as the final grade in your last English course? The responses were 6, 20

and 40 respectively. Student perception in the Harris study suggests that only a handful felt their writing had improved. However, pupils overwhelmingly preferred peer evaluation over self-evaluation (17,9,2); and peer evaluation over teacher evaluation. (53,13,4). In the main study, as shown, some students were conscious of growth in their writing ability. They felt that once they had internalised the criteria and judged writing samples, they were able to apply them to their own work, and "understand the basics of writing."

Still another question is who the evaluator is that judges the work. Personal feelings are involved here. In the Harris study, 39 students said they saw personal feelings operating; 29 students were not conscious of this as a factor and 2 were not sure.

Harris did not ask a specific question on the inherent and other difficulties in the use of the method. In both the pilot and main study, students felt the process was difficult though manageable and were not sure whose role it was to evaluate. Most students felt they were objective enough to assess their own papers.

Summary

An analysis of the attitudinal data confirmed the findings of previous studies. If other things are equal, peer rating could be an efficient teaching/learning strategy. The process might be difficult for some pupils, and the success of the method is closely linked to the length and kind of preparation given to students. Finally, as

Harris writes, all pupils are not comfortable with the method since the method might be incompatible with their learning style.

CHAPTER NINE

SUMMARY AND CONCLUSIONS

The pilot study sought answers to a number of problems about the effect of the experiment:

(A): Did the programme improve the writing quality of the pupils who used it?

(B): Did pupils become more proficient in the use of English expression and the mechanics of writing?

(C): What was the influence of sex differences on writing ability?

(D): What was the interaction of sex and programme?

(E): How well could teachers use the analytic scale to rate papers?

(F): Did the programme improve student attitude towards writing?

The main study investigated some of the same problems, but some changes were made as indicated.

(A): Question (A) above was changed to: Did the programme improve the writing quality of the pupils who used it on total score and each sub-score?

(B): Question (B) was unchanged: Did pupils become more proficient in the use of English expression and the mechanics of writing?

(C): Question (C) was not tested.

(D): Question (D) was not tested.

(E): Question (E) was replaced by: How well could pupils be taught to use the scale to rate papers for total score and sub-scores?

This chapter summarises the results of the significance tests and the analysis of other data. After that, the question as to how the results could be interpreted is addressed relative to the limitations on the study. Finally, the implications of the study for teaching and further research are discussed. The discussion is organised around each research question or set of questions.

Results and Limitations

Effect of the Programme

The analysis of pilot study data yielded a statistically significant difference for overall writing quality between the two groups at the .05 level of significance. A similar analysis for the main study yielded a statistically significant difference between groups at the .02 level.

How were the results obtained? Analysis of co-variance, the test recommended and detailed by Campbell and Stanley, was used to

analyse the data. However, certain assumptions must be considered and conditions met when using this procedure. The first assumption is that the samples representing the population of the sub-groups were selected at random. The second assumption is that each sample came from a normally distributed population. The third assumption is that those populations had equal variances. The final assumption is that the slopes of the regression lines used to predict criterion scores were equal.

It is very difficult to meet such conditions in actual classroom situations. Relative to this study, attempts were made to approximate as many of these conditions as possible, in some cases with doubtful success. Indeed, possible limitations on the design of the study have been pointed to throughout the report of this experiment. For example, there were problems with the selection of the pilot study sample; problems that resulted in staggering the start of the main study at one school; the problem of classes meeting at different times of the day; problems with the influence of external events, and so on. Apart from such limitations on the collection of data then, would be the possible violations of the assumptions on which ANCOVA analysis is based.

A sort of claim could be made for normality of populations for the main study. It comprised students taking general English classes at grade 12 level in the only two public high schools in the school district. The sample was drawn at random and seemed to be relatively large for this kind of study, and variance checks suggested no

significant differences in variability. Randomisation appears to be the key determinant of validity. Popham and Sirotnik (1973) note that:

> "randomisation is the sine qua non for the validity of the statistical test. We need not worry about the normality and homogeneity of variance assumptions, nor conceptualise a population and a random sampling process." (p.241)

Even so, the results of the analysis must be interpreted with caution. The results of both the pilot and main studies suggest that there was a real difference between the two groups of students, and that this difference resulted from the difference in teaching method.

Sub-scales

The pilot study post-test compositions were rated holistically, so there were no sub-scale scores. The main study papers were rated by teachers using the analytic scale devised by the researcher for total score and sub-scores. The differences between means were statistically significant at less than the .05 alpha set. The *p* value for content dimension was .003; for organisation it was .03; for wording it was .09.

The means for content showed the greatest difference. It is perhaps not easy to say why, but perhaps the order in which the units were studied helped. The unit on content was taught first, and the

unit on wording was taught last. The same limitations on interpretation mentioned above apply to the data for sub-scales as well.

Pupils as Raters

Students in the Experimental group rated five compositions, and the co-efficients obtained compared favourably with reliabilities obtained for such measures. The examples of student rating of papers in groups given in chapter 7 - for one quality, then two qualities, then three, then all four categories - were deliberately chosen to illustrate student proficiency as raters. Both sets of evidence then, from course work and end of programme examinations, suggest student proficiency as raters.

Student Attitude to Writing

Answers to this question were given extensive treatment in the last sections of chapter 4 and the final section of chapter 8. Data were analysed from self reports, and questionnaires. The assumption behind the analysis was that pupils were telling the truth, and not trying to please or displease the other parties involved. However, the tone of the letters of the main study, and the anonymity of the pilot study questionnaire point to at least a positive effect on attitude.

Implications

Teaching

The results suggest that the experimental method of teaching composition was more successful than the school method. Further, at least one teacher in the programme felt it was superior to the current method used. On the heels of the programme, the researcher was invited back to one school to consult with grade 12 teachers in the English department. The teacher who taught the programme thought that the techniques could be applied to the study of the research paper. The time teachers spent working on criteria, and discussing group processes displayed a degree of enthusiasm which holds promise for the future. The programme received favourable comment at Council of Teachers of English meetings, and caught the attention of education officials.

The invitation was indicative of the willingness of teachers to co-operate and come together on their own time to talk about writing. It indicated that teachers were philosophically ready to entertain a new concept of roles - willing to allow students to teach themselves, and participate in the evaluation process. It suggested that they were ready to try new methodologies like collaborative techniques which demand interactive learning, whereas in the traditional classroom the teacher is dominant. It pointed to teachers being ready to invest time and effort in learning new techniques. The teachers at this school certainly seemed to be ready.

The right kind of teachers seems to be necessary for the success of the programme. The techniques should not be imposed on all

teachers, but selected teachers could try out the method. If results are satisfactory, there could be a systematic incorporation of more classes, but this might call for wider educational policy decisions. What are some of the other implications for teaching?

First, there is the need for teacher-training in inter-personal skills and group processes. Collaborative learning, as we have seen, demands a wide range of skills. Here the speech area of the Humanities division of the university could help. Teachers need to feel comfortable with the technique, and training sessions could alleviate initial apprehension. Students also must be made ready to accept new roles, and a climate of trust developed in the classroom. Above all, they must be taught to collaborate. It was pointed out that placing students in groups is only the first step. They must be taught to ask the right questions. In short, they must learn to talk as well as talk to learn.

The second consideration is the need to restructure the English curriculum from grade 9 to 12, so as to introduce collaborative techniques as early as possible. A developmental sequence of writing abilities arranged in a spiral curriculum would help to take pressure from the top of the school where there is a rush to complete course work. The absence of the pressure of exit examinations, and reduced competition for limited student time would help teachers consolidate the kind of skills necessary for the institution of a successful programme. Improvement in writing takes time; learning group processes also take time. Studies suggest that teachers could at first spend

half their time on process and half on task.

A third consideration is the side-effects for both teacher and pupils. Teachers would meet to talk about writing and discuss evaluation techniques in in-service training sessions. Students would talk about their papers using the terminology of the field and develop meta-linguistic skills. Teachers would have more time to monitor the progress of individual students; students find renewed motivation to perform at what can approach their optimum potential.

The request of the school mentioned earlier suggests that the need for new methods is present. The model in current use certainly does not appear to work for some students - at least for the corpus of recent high school graduates seeking admission to the University of the Virgin Islands. The failure rate on the English placement test - especially on the essay component - is high. It is reasonable to assume that one factor affecting results is teaching method. The usual response to this failure rate is that the better students go off to study on the mainland United States; the rest attend the local university. If this is true, the current method is not working for those who are left behind.

Finally, the confirmation of the Sager findings on a new population, the indication that teachers are willing to incorporate student grades into summative assessment, the new ways of seeing developed in pupils who were involved in the experiment, the application of criterion-referenced evaluation of essays - all suggest

implications for further research.

Research

Suggestions for further research centre on two foci:- the need for a longitudinal study of this research problem, and for replication studies.

Most researchers would agree that growth in writing abilities takes time, and some skills seem to take longer to develop than others. For instance, one result of this study was that the two methods of teaching structure - the school method and the experimental method did not result in significant differences between groups. Put another way, one method was as good as the other. Would a longitudinal study have made a difference?

More importantly, with the increasing interest in developmental studies, it is important that researchers chart the progress of students over time. There seems to be a paucity of such studies in the field of writing research as the same studies come up in the literature, again and again. In consequence, one suggestion for further research is a longitudinal study at the high school level.

A researcher may wish to do a panel study, selecting a sample of students at the start of their high school career (grade 9) and chart their progress through grade 12. This cohort of students could be tested at specific intervals, say at the end of each semester, to study changes over time as writing tasks change according to audience

and function. The study could be replicated or modified to test its efficacy across a student's educational career.

The study suggests many possibilities for replication. Indeed, the need for such replication is evident. The experiment described in this report represented one sample of a large population of possible experiments; the subjects represented one sample of a large population of possible subjects. So one important question is whether the relationship between the experimental programme and writing quality is a true relationship.

From one viewpoint, the results of earlier research suggest that it is, but these studies were conducted in the mainland United States on population samples which were culturally and economically different from those in the Virgin Islands, a United States territory. In the same way, the population of students in the neighbouring British Virgin Islands is essentially different from that in the USVI. Indeed, one facet of the rationale for this study was to replicate the Sager study on the culturally and socio-economically different population in St.Thomas and St.John. This study should be replicated then to test the validity of findings. First, there could be a series of fairly literal replications.

First, the study could be replicated using a different sample of grade 12 students. Second, the study could be replicated at another grade level, say grade 6 in public schools, since junior high school students exit at this level. Third, a researcher might wish to conduct the study on a different population - private, instead of

public schools. Or certain minor changes could be made to the design.

A researcher could change the order of the dimensions of writing under study. The category of _content_ had the highest _p_ value. Was it because this was the first focus of attention? What would happen if _structure_ were treated first?

Or an investigator could focus on interaction effects. Would there be an interaction between sex and programme for a different sample of students? Would there be a significant difference in writing ability between male students and male teachers on the one hand, and female students and female teachers on the other?

Still another area of concern is whether the programme suited "slower" learners rather than _faster_ learners. One teacher made this suggestion, and it might be interesting to test the hypothesis. Again, the researcher might opt for the use of student samples only, eliminating professional models entirely. But the study could be modified or extended in other various ways.

For instance, different measures could be used. Instead of the STEP objective measures used in this investigation, and the researcher-made composition probes, the researcher could use other measures. It would certainly be instructive to test the results using holistic scoring procedures rather than analytic. For instance, the use of

a 4-point scale with points 1 and 2 failing, and points 3 and 4 passing could eliminate the middle category and side-step the central tendency of raters.

Thus, regardless of the kind of replication employed, such studies would help to verify or modify the results of this empirical study.

APPENDIX A

APPENDIX A

TABLE 31

SOURCE TABLE FOR TABLE 2

Analysis of Covariance of Post-test Composition Scores with Co-variate Composition Scores

Residuals

Source of Variation	Degrees of Freedom	Sums of Squares	Means Square	F
Between	1	26.25	26.25	4.06[a]
Within	39	252.27	6.47	
Total	40	278.52		

[a] Significant at the .05 level.

TABLE 32

SOURCE TABLE FOR TABLE 3

Analysis of Covariance of Experimental and Control Grade - Twelve Students for Post-test Grammar Scores

Residuals

Source of Variation	Degrees of Freedom	Sums of Squares	Means Square	F
Between	1	19.02	19.0	3.39[a]
Within	39	218.17	5.49	
Total	40	237.19		

[a] Significant at the .07 level.

TABLE 33

SOURCE TABLE FOR TABLE 4

Effect of Pre-test Mechanics Scores on Post-test Composition Scores

Source of Variation	Degrees Freedom	Sums of Squares	Mean Square	F
		Residuals		
Between	1	17.33	17.33	2.71[a]
Within	39	249.08	6.39	
Total	40	266.41		

[a] Significant at the .10 level

TABLE 34

SOURCE TABLE FOR TABLE 5

Analysis of Covariance of Post-test Grammar Scores with Scores Covariate Pre-test Grammar Scores

Source of Variation	Degrees Freedom	Sums of Squares	Mean Square	F
		Residuals		
Between	1	31.20	31.20	1.60
Within	44	856.74	19.47	
Total	45	887.94		

TABLE 35

SOURCE TABLE FOR TABLE 6

RATER "R":WHOLE SCALE (N=3)

SOURCE	SUMS OF SQUARES	VARIANCE
36 Essay Scores	6114	25.2
12 Essay Totals	18076	62.3
3 Raters Totals	69746	

Using Cronbach's alpha and the above data

TABLE 36

SOURCE TABLE FOR TABLE 6

RATER "R" : CONTENT (N=3)

SOURCE	SUMS OF SQUARES	VARIANCE
36 Content Scores	333	2.91
12 Content Totals	955	5.75
3 Rater Totals		

Using Cronbach's alpha and the above data

$$r = 3/2\ (1-2.91/5.75)$$

$$= .74$$

TABLE 37

SOURCE TABLE FOR TABLE 6

RATER "R" : ORGANISATION (N=3)

SOURCE	SUMS OF SQUARES	VARIANCE
36 Organisation Scores	392	2.29
12 Organisation Totals	1146	5.25
3 Rater Totals	4374	

Using Cronbach's alpha and the above data

$$r = 3/2\ (1-2.29/5.25)$$

TABLE 38

SOURCE TABLE FOR TABLE 6

RATER "R" : STRUCTURE (N=3)

SOURCE	SUMS OF SQUARES	VARIANCE
36 Structure Scores	421	2.17
12 Structure Totals	1235	4.58
3 Rater Totals	4739	

Using Cronbach's alpha and the above data

$$r = 3/2\ (1-2.17/4.58)$$

TABLE 39

SOURCE TABLE FOR TABLE 6

RATER "R" : WORDING (N=3)

SOURCE	SUMS OF SQUARES	VARIANCE
36 Wording Scores	426	2.07
12 Wording Totals	1256	4.67
3 Rater Totals	4814	

Using Cronbach's alpha and the above data

$$r = 3/2\ (1-2.07/467)$$
$$= .84$$

APPENDIX B

APPENDIX B

TABLE 40

RAW SCORE DATA FOR MAIN STUDY

PRE-TEST

1	2	3
38	20	28
16	17	17
21	15	21
29	16	17
23	10	21
19	7	8
34	19	26
29	21	26
22	10	24
31	22	26
19	12	17
26	23	20
22	14	22
26	19	25
23	21	19
18	12	14
19	16	22
21	19	22
19	17	27
33	25	22
26	27	23
20	18	16
23	10	19
24	20	21
28	25	25
19	28	32
30	13	24
23	16	21
21	19	25
23	24	31
20	18	29
24	19	18
32	24	29
22	19	13
21	18	26
24	22	16
22	21	25
24	18	16
19	16	10
21	22	23
27	15	20
24	13	18
20	18	21
35	17	23

POST-TEST

1	2	3	4	5	6	7
32	8	8	8	8	21	29
23	6	5	6	6	21	13
19	4	4	5	6	22	26
23	5	5	6	7	20	22
20	4	6	5	5	17	20
15	4	3	4	4	13	17
30	8	8	7	7	28	31
27	7	7	7	6	25	25
23	5	4	7	7	19	20
32	9	7	8	8	25	28
21	7	5	5	4	17	20
30	8	8	7	7	22	22
22	6	6	5	5	15	27
32	9	9	7	7	24	33
25	8	4	6	7	21	26
20	4	5	5	6	16	13
22	5	5	6	6	16	22
24	6	6	6	6	22	22
20	4	3	7	6	18	25
30	7	7	8	8	24	22
32	5	5	6	6	22	28
21	4	4	6	6	14	19
26	6	7	7	6	15	21
22	6	5	5	6	15	25
26	6	6	7	7	23	28
19	6	4	5	4	26	25
18	4	4	5	5	18	22
22	7	4	5	6	20	20
25	7	5	6	7	22	24
32	8	8	8	8	24	22
26	7	7	6	6	16	25
26	7	5	7	7	24	26
26	8	6	6	6	25	36
19	3	4	6	6	17	19
23	6	3	8	6	17	24
24	6	5	6	7	19	15
25	6	4	7	8	19	31
21	4	5	6	6	14	25
28	7	7	7	7	11	19
24	4	6	7	7	25	25
26	7	7	6	6	22	20
20	4	4	6	6	19	17
24	7	5	6	6	23	20
29	7	8	7	7	28	19

TABLE 40 (Continued)

28	12	11
22	18	27
21	12	21
21	7	17
23	8	16
29	15	29
17	13	19
26	10	15
25	19	32
24	12	19
21	10	18
17	8	12
30	22	26
17	15	16
24	27	28
21	24	29
29	12	17
24	17	24
33	24	33
25	14	21
24	19	15
28	25	29
24	15	22
25	24	23
30	17	22
26	16	26
18	20	32
25	18	27
27	24	27
21	20	23
19	12	27
12	24	26
29	19	24
31	14	22
27	17	19
23	21	28
22	20	25

21	5	4	5	7	19	11
24	6	5	6	7	27	30
24	6	5	6	6	19	23
24	7	5	6	6	21	25
24	6	6	6	6	17	16
25	7	5	7	6	23	36
15	4	3	3	5	21	17
13	4	3	3	3	18	18
20	6	4	5	5	20	19
26	6	5	7	8	16	23
30	7	8	8	7	23	28
18	4	3	5	6	16	20
27	8	7	6	6	22	28
15	4	3	4	4	24	26
35	9	9	8	9	21	28
27	8	7	6	6	20	29
27	8	6	7	6	19	30
27	7	6	7	7	14	20
30	7	8	8	7	27	31
32	8	8	8	8	10	16
24	9	8	8	9	20	24
29	8	6	7	8	27	25
32	8	8	8	8	13	27
29	8	7	7	7	18	24
26	7	7	6	6	18	26
34	9	8	9	8	27	27
26	7	5	7	7	23	35
27	7	6	6	8	20	33
35	9	9	8	9	23	30
35	9	9	8	9	21	26
23	6	6	5	6	18	21
21	5	6	5	5	11	23
29	8	7	7	7	13	19
29	8	8	7	6	19	20
24	7	5	6	6	16	18
34	9	7	8	10	25	28
30	8	8	7	7	19	23

TABLE 41

PRE-TEST AND POST-TEST MEANS AND STANDARD DEVIATIONS

Variable	Grand Means	Standard Deviations
Pre-tests		
Grammar	17.24	5.05
Mechanics	22.05	5.43
Apprehension	91.06	16.85
Essay		
Total	23.99	4.80
Content	6.17	1.50
Organisation	5.40	1.63
Structure	6.12	1.27
Wording	6.29	1.25
Post-tests		
Grammar	20.00	4.28
Mechanics	23.55	5.10
Apprehension	90.23	14.68
Essay		
Total	25.26	5.01
Content	6.50	1.59
Organisation	5.87	1.68
Structure	6.37	1.20
Wording	6.52	1.25

TABLE 42

CORRELATION MATRIX BASED ON POPULATION DATA

	PRE-TEST							POST-TEST		
	Essay	Step	Step	Essay	Content	Organ.	Structure	Wording	Step	Step
Group	.0536	-.1458	.1077	.2179	.2753	.2114	.0817	.1754	.0000	-.0072
Essay		.2933	.2617	.4315	.3977	.4120	.3857	.3071	.3333	.3137
Step			.5894	.4557	.4515	.3336	.3981	.4130	.4718	.3950
Step				.3943	.4207	.3215	.3677	.2633	.4145	.5516
Essay					.8858	.8882	.8800	.8378	.2867	.3622
Content						.7651	.6618	.6202	.3182	.3880
Organisation							.6774	.5918	.2176	.3006
Structure								.8044	.2400	.3437
Wording									.2382	.4008
Step										

TABLE 43

RAW SCORES FOR STUDENT RATINGS : TOTAL SCORE

ESSAYS				
A	B	C	D	E
15	4	12	17	10
14	5	11	12	12
15	6	15	12	11
10	8	11	10	10
15	4	17	12	11
14	5	14	12	13
16	4	13	12	10
18	4	14	14	12
16	5	11	11	12
15	7	18	16	12
15	6	14	13	10
15	5	16	15	14
16	4	18	13	13
13	5	19	11	12
14	5	15	13	12

TABLE 44

RAW SCORES FOR STUDENT RATINGS : CONTENT

ESSAYS				
A	B	C	D	E
4	1	3	5	3
4	1	2	2	4
4	1	2	4	3
2	1	2	2	2
4	1	4	3	3
4	2	4	4	3
5	1	3	3	3
5	1	3	4	3
4	1	2	2	3
3	2	5	4	3
4	2	5	4	3
3	2	4	5	3
5	1	4	4	4
2	2	5	3	3
4	1	5	4	4

TABLE 45

RAW SCORES FOR STUDENT RATINGS : ORGANISATION

ESSAYS				
A	B	C	D	E
4	1	3	4	3
3	2	4	4	4
5	1	4	3	3
2	2	2	2	2
4	1	5	3	2
4	1	5	3	4
4	1	3	2	2
5	1	2	4	2
3	1	5	4	2
4	1	4	3	3
3	1	4	3	4
5	1	4	3	3
2	1	5	2	3
4	1	5	4	4

TABLE 46

RAW SCORES FOR STUDENT RATINGS : STRUCTURE

ESSAYS				
A	B	C	D	E
4	1	3	4	2
4	1	1	1	2
4	3	4	2	3
3	3	3	3	2
5	1	4	3	3
2	1	2	2	2
4	1	3	4	2
4	1	4	2	3
3	1	4	4	2
4	2	4	3	3
2	2	4	3	2
5	1	4	3	4
2	1	5	2	3
4	1	5	3	3
4	2	3	3	2

TABLE 47

RAW SCORES FOR STUDENT RATINGS : WORDING

A	B	C	D	E
3	1	3	4	2
3	1	4	3	3
2	1	5	3	2
3	2	4	3	4
2	1	4	3	3
4	1	3	3	4
3	1	4	3	3
4	1	5	4	4
4	2	3	4	4
5	2	4	5	4
5	1	1	3	2
4	1	4	4	3
4	1	5	4	3
5	1	4	3	3
2	1	2	2	2

APPENDIX C

APPENDIX C

FIGURE 17

QUESTIONNAIRE

Dear Student:

You are about to end your writing programme and we would like you to answer a few questions. Think about your answers carefully as they will help us to adjust the programme for the next class.

---

(1) How did you like each of the following. Circle your choice. The lowest rating is (not at all). The highest rating is 5 (very much).

A:	Writing letters	1	2	3	4	5
B:	Receiving letters	1	2	3	4	5
C:	Rating papers	1	2	3	4	5
D:	Working in groups	1	2	3	4	5
E:	Using booklets	1	2	3	4	5
F:	Writing assignments	1	2	3	4	5

Give reasons for each choice

A: ______________________________

____

____

B: ______________________________

____

____

C: ______________________________

____

____

D: ____

____

____

E: ____

____

____

F: ____

____

____

(2) Do you think your writing has improved? ____________

If yes, say in what ways.

____

____

____

____

If no, say why not.

____

____

____

____

(3) Write anything else you wish to say about the programme.

____

____

____

____

____

FIGURE 18

WRITING APPREHENSION MEASURE

DIRECTIONS: Below are a series of statements about writing. There are no right or wrong answers to these statements. Please indicate the degree to which each statement applies to you by circling whether you (1) strongly agree, (2) agree, (3) are uncertain, (4) disagree or (5) strongly disagree with the statement. While some of these statements may seem repetitious, take your time and try to be as honest as possible. Thank you for your cooperation in this matter.

1: I avoid writing ... ...

2: I look forward to writing down my ideas. ... ...

3: I have no fear of my writing being evaluated. ... ...

4: I am afraid of writing essays when I know they will be evaluated. ... ...

5: Taking a composition course is a very frightening experience. ... ...

6: Handing in a composition makes me feel good. ... ...

7: My mind seems to go blank when I start to work on a composition. ... ...

8: Expressing ideas through writing seems to be a waste of time. ... ...

9: I would enjoy submitting my writing to magazines for evaluation and publication. ... ...

10: I like to write my ideas down. ... ...

11: I feel confident in my ability to clearly express my ideas in writing.	...	...	...	...	...	...
12: I like my friends to read what I have written.	...	...	...	...	...	...
13: I am nervous about writing.	...	...	...	...	...	...
14: People seem to enjoy what I write.	...	...	...	...	...	...
15: I enjoy writing.	...	...	...	...	...	...
16: I never seem to be able to clearly write down my ideas.	...	...	...	...	...	...
17: Writing is a lot of fun.	...	...	...	...	...	...
18: I expect to do poorly in composition classes even before I enter them.	...	...	...	...	...	...
19: I like seeing my thoughts on paper.	...	...	...	...	...	...
20: Discussing my writing with others is an enjoyable experience.	...	...	...	...	...	...
21: I have a terrible time organising my ideas in a composition course.	...	...	...	...	...	...
22: When I hand in a composition I know I'm going to do poorly.	...	...	...	...	...	...
23: It's easy for me to write good compositions.	...	...	...	...	...	...
24: I don't think I write as well as most people do.	...	...	...	...	...	...
25: I don't like my composition to be evaluated.	...	...	...	...	...	...
26: I'm no good at writing.	...	...	...	...	...	...

FIGURE 19

RATING SCALE: HIGH

5

THE PAPER SHOWS EVIDENCE OF CAREFUL THOUGHT

---

* FOCUS

    * There is a sharp focus.

        * The writer limits the subject to one controlling idea.

    * The writer stays on focus.

        * The writer shows control and there is no irrelevant material.

* SUPPORT

    * Supporting ideas are well chosen and sufficient to support the writer's purpose.

* DETAIL

    * The detail is sufficient to develop each supporting idea or main point.

* LINKS

    * Focus, supporting ideas and detail fuse in a chain of relationships to give total picture and a sense of completeness.

        * Supporting ideas are directly linked to the controlling idea.

        * The detail is directly linked to each supporting idea.

FIGURE 20

RATING SCALE: MIDDLE

THE PAPER SHOWS SOME EVIDENCE OF THOUGHT

* FOCUS

    * There is a general focus.

    * The writer develops a TOPIC rather than a main idea.

* SUPPORT

    * There are some supporting ideas but these are not always carefully selected.

* DETAIL

    * Supporting ideas are developed in some detail but these are not always fully explained.

    * Usually the detail is enough to give a general picture.

* LINKS

    * There is a chain of relationships but the links are somewhat loose.

    * Some sentences are not linked by transitional devices.

FIGURE 21

RATING SCALE: IDEAS

LOW

THERE IS NOT MUCH EVIDENCE OF THOUGHT

* FOCUS

    * There is no real focus.

      or

    * The writer does not develop a focus.

* SUPPORT

    * There are hardly any points to support the controlling idea.

    * The writer lists ideas without developing them.

* DETAIL

    * There is little or no elaboration.

    * The writer does not stay long enough on any idea to make the point.

* LINKS

    * There is no common thread holding the paper together.

FIGURE 22

RATING SCALE: ORGANISATION

HIGH

THE PAPER IS WELL ORGANISED AT BOTH LEVELS:

THE WHOLE PAPER AND AT THE PARAGRAPH LEVEL

THE WHOLE PAPER:

* A controlling idea holds the paper together.
* A master plan gives order to the material.
* The writer sticks to the plan and puts cues in the script to make it easy for the reader to follow.
* The paper has three main divisions: an introduction, development and conclusion.
* Ideas are given space in relation to their importance.

THE PARAGRAPH:

* The main idea in each paragraph is directly linked to the controlling idea in the first paragraph.
* Each paragraph division is natural, and forms a distinct unit of thought.
* The detail is directly linked to the topic sentence.
* The flow between paragraphs is smooth and natural.
* Transitions help to tie the ideas in the paragraph together.

FIGURE 23

RATING SCALE: ORGANISATION

MIDDLE

THERE ARE PROBLEMS WITH EITHER THE PAPER AS A WHOLE OR THE PARAGRAPHING

THE WHOLE PAPER:

* There is a controlling idea, but the writer is not always in control.

* If there is a master plan, the writer does not always follow the plan and some parts are out of order.

* The paper has three (3) main divisions, but some parts are out of balance.

THE PARAGRAPH:

* All main points are not linked to the controlling idea.

* Some divisions are faulty and come at the wrong places.

* Some detail is not well chosen and not connected to the main points.

* The flow between paragraphs is not always smooth or natural.

FIGURE 24

RATING SCALE

ORGANISATION: LOW

THERE ARE PROBLEMS WITH THE PARAGRAPH AS A WHOLE AS WELL AS THE PARAGRAPHING

THE WHOLE PAPER:

* No controlling idea holds the paper together.

* If there is a master plan, the writer does not follow it and the reader's expectations are not fulfilled.

* The paper may have three main divisions, but it is lop-sided and some ideas are given too much space in relation to their importance.

THE PARAGRAPHS:

* Main ideas are not linked to the controlling idea and exist apart from it.

* The writer divides the paper at random and the sections are not separate units of thought.

* The detail in some paragraphs is not relevant and the links between sentences are weak or non-existent.

* the flow between paragraphs is not smooth or natural.

FIGURE 25

RATING SCALE: SENTENCE STRUCTURE

HIGH

ASSESSED ON FIRST TWENTY-FIVE LINES

---

* Demonstrates competence in the correct use of sentence patterns taught in class.
* Makes no gross errors in sentence sense - especially fragments and run-ons.
* Masters the mechanics of agreement within the sentence.
* Varies sentence patterns, length and order of the elements.

FIGURE 26

RATING SCALE: MIDDLE

ASSESSED ON THE FIRST 25 LINES

* The writer has

    * some a basic understanding of sentence patterns and uses most of them correctly.

    * has some agreement errors but not more than 5 in all.

    * some sentence variety but there is an unwillingness to experiment.

    * a few errors in sentence sense and the paper has a few fragments or run-ons.

FIGURE 27

RATING SCALE: LOW

ASSESSED ON THE FIRST 25 LINES

* The writer has

  * serious problems with sentence patterns and the paper is riddled with errors.

  * many problems in sentence sense - especially fragments and run-ons.

  * problems with the agreement rules learned in class.

FIGURE 28

RATING SCALE: WORDING

HIGH

The writer uses words in new and exciting ways.

* The paper has suggestive power.
    * Words and phrases
        * are chosen for their vivid qualities.
        * are generally specific and concrete.
        * are used correctly in context.
* There is a variety of sensory detail.
* The writer uses imagery to help the reader share the experience.
* There are no gross spelling errors, especially in high frequency words.

FIGURE 29

RATING SCALE: WORDING

MIDDLE

There are some descriptive words and phrases, but the paper as a whole does not suggest vivid pictures to the reader.

---

* Some words paint pictures, but they are not very sharp.
    * The writer uses stale and overworked expressions.
    * words are so general that they have lost their force.
    * there is some sensory detail but words are not used in exciting ways.
* Sometimes words are used incorrectly, inappropriately or needlessly repeated.
* The description is somewhat overdone and seems unnatural.
* There is the occasional spelling error.

FIGURE 30

RATING SCALE: WORDING

LOW

A PAPER RATED LOW IS DULL AND LIFELESS

* Words and phrases usually has no image-making power.
    * There
        * are too many general and abstract words.
        * is too little sensory detail.
* The English is too informal and thus inappropriate for the context of school writing.
* Words are generally used carelessly, incorrectly or are otherwise unsuitable.
* Spelling errors are frequent.

FIGURE 31

ESSAY

TIME: 50 minutes

Carnival is an important yearly event in The Virgin Islands. Hundreds of tourist visit the islands for this occasion.

Imagine that you have a relative overseas who has never been here. Write a letter persuading the person to pay you a visit at carnival.

CARNIVAL

(1) Select one aspect of carnival.

(2) Describe it in sufficient detail to give a clear picture.

(3) Make your description exciting so that it will come alive to the reader.

GOOD LUCK

FIGURE 32

ESSAY

TIME: 50 MINUTES

There is something special about Christmas in many countries. Write a composition with the title:-

---

WHAT MAKES CHRISTMAS SPECIAL FOR ME

(1) State specifically what makes Christmas special for you.

(2) Describe it in sufficient detail to give a clear picture.

(3) Make your description exciting so that it will come alive to the reader.

GOOD LUCK

FIGURE: 33

ENGLISH COMPOSITION

INDIVIDUAL

RATING SHEET

CONTENT

NAME OF RATER: ____________________

GROUP NUMBER: ____________________

GROUP NAMES: (1) ____________________ (3) ____________________

(2) ____________________ (4) ____________________

(5) ____________________

STUDENT SCRIPTS		INDIVIDUAL RATINGS				
NAMES	CATEGORIES	LOW		MIDDLE	HIGH	
(1) ____________	Content	1	2	3	4	5
(2) ____________	Content	1	2	3	4	5
(3) ____________	Content	1	2	3	4	5
(4) ____________	Content	1	2	3	4	5

FIGURE: 34

ENGLISH COMPOSITION

INDIVIDUAL

RATING SHEET

CONTENT/ORGANISATION

NAME OF RATER: ____________________

GROUP NUMBER: ____________________

GROUP NAMES: (1) ____________________ (3) ____________________

(2) ____________________ (4) ____________________

(5) ____________________

STUDENT SCRIPTS			INDIVIDUAL RATINGS			
NAMES	CATEGORIES	LOW		MIDDLE	HIGH	
(1) ____________	Content Organisation	1 1	2 2	3 3	4 4	5 5
(2) ____________	Content Organisation	1 1	2 2	3 3	4 4	5 5
(3) ____________	Content Organisation	1 1	2 2	3 3	4 4	5 5
(4) ____________	Content Organisation	1 1	2 2	3 3	4 4	5 5

FIGURE: 35

ENGLISH COMPOSITION

RECORDER

RATING SHEET

STRUCTURE/ORGANISATION/CONTENT

NAME OF RATER: ____________________

GROUP NUMBER: ____________________

GROUP NAMES: (1)____________________ (3) ____________________

(2)____________________ (4) ____________________

(5) ____________________

STUDENTS SCRIPTS				INDIVIDUAL RATINGS		
NAMES	CATEGORIES	LOW		MIDDLE	HIGH	
(1) ____________	Structure	1	2	3	4	5
	Organisation	1	2	3	4	5
	Content	1	2	3	4	5
(2) ____________	Structure	1	2	3	4	5
	Organisation	1	2	3	4	5
	Content	1	2	3	4	5
(3) ____________	Structure	1	2	3	4	5
	Organisation	1	2	3	4	5
	Content	1	2	3	4	5
(4) ____________	Structure	1	2	3	4	5
	Organisation	1	2	3	4	5
	Content	1	2	3	4	5

FIGURE: 36

ENGLISH COMPOSITION

RECORDER

RATING SHEET

CONTENT/ORGANISATION/STRUCTURE/WORDING

NAME OF RATER: ____________________

GROUP NUMBER: ____________________

GROUP NAMES: (1) ____________________ (3) ____________________

(2) ____________________ (4) ____________________

(5) ____________________

STUDENTS SCRIPTS				INDIVIDUAL RATINGS		
NAMES	CATEGORIES	LOW		MIDDLE	HIGH	
(1) ____________	Content Organisation Structure Wording	1 1 1 1	2 2 2 2	3 3 3 3	4 4 4 4	5 5 5 5
(2) ____________	Content Organisation Structure Wording	1 1 1 1	2 2 2 2	3 3 3 3	4 4 4 4	5 5 5 5
(3) ____________	Content Organisation Structure Wording	1 1 1 1	2 2 2 2	3 3 3 3	4 4 4 4	5 5 5 5
(4) ____________	Content Organisation Structure Wording	1 1 1 1	2 2 2 2	3 3 3 3	4 4 4 4	5 5 5 5

APPENDIX D

APPENDIX D

FIGURE 37

LESSON 8: SUMMARY SHEET

ORGANISATION

OBJECTIVE: Rates writing samples.

LEVEL OF INTENSITY: Interactive with explicit rules.

Mainly student - student discussion.

INSTRUCTIONAL ROLES: Burden on materials and students.

USE OF SPACE: Students

* form groups of four.
* use horse shoe format.
* face one another.

The teacher

* circulates among groups.
* helps with group processes.

MATERIALS: Samples essays:

* "A Center Cut"
* "The Roots Wedding"

FIGURE 38

TEACHER'S LESSON PLAN (8)

PHASE (I):

GETTING STARTED

* Students
    * Choose own groups.
    * Arrange furniture.
    * Determine roles.
    * Select
        * leader.
        * recorder.

* Teacher sets the stage:
    * says students will have a go at rating papers.
    * gives class directions:
        * each student will
            * work alone.
            * determine rating.
            * fill in details at top of individual rating sheets.

        * each leader will
            * ask for individual ratings.
            * let students justify rating.
            * discuss one group rating.
        * The recorder will list specific points raised about the composition.
* Teacher
    * distributes scripts and score sheets.
    * students work on "Center Cut".

(TIME: 5 minutes)

PHASE (2):

GROUP RATING

* Teacher
    * sets group work in train.
    * urges students to use scale descriptions and anchor papers to help them decide on ratings.
    * monitors and facilitates the process.

* asks each student to decide rating.
* directs each group to decide rating.
* THE SAME PAPER IS RATED FOR <u>CONTENT</u> WITHOUT CONSENSUS.

(TIME: 20 minutes)

PHASE (3):

* CLASS RATING

* Teacher

  * asks each leader to report.
  * requests each recorder to read points made.
  * compares group ratings.
  * discusses mismatches to resolve discrepancies.

(TIME: 10 minutes)

THE PROCESS IS REPEATED

(TIME: 15 minutes)

Summary Sheet

Figure 11 presents the summary sheet for lesson 8.

FIGURE 39

SUMMARY SHEET FOR LESSON 8: STRUCTURE

OBJECTIVE:	(1) Studies variety in sentence structure. (2) Rates paper for structure. (3) Distinguishes between a paper rated 4 and a paper rated 5.
LEVEL OF INTERACTION:	The flow of discussion is between teacher and student, rather than between student and student.
USE OF SPACE:	Whole class format.
INSTRUCTIONAL ROLES:	Shared between teacher and students.
MATERIALS: *	Scale for HIGH.
*	SUSAN.

FIGURE 40

STUDENT SAMPLE FOR RATING

Instructions: Pupils were to describe a person eating something. They were to describe the process paying special attention to content, organisation and structure. They were to make sure that their classmates got a clear picture.

1: I am sitting in a far corner of the lunch room. 2: Susan is gracefully eating her lunch which consists of sardine, sweet potato, peaches and a bun. 3: She is now watching around her. 4: Her hands are now moving towards her mouth as she takes a small bite of her bun. 5: Her hand is in her lap and she is using her right hand to guide the fork to her mouth. 6: She is now moving her fork into her left hand as she puts her hand on her right hip. 7: Susan is now taking her napkin and wiping her mouth in a slow graceful motion, like the movie stars on television. 8: She dips her left hand into her shirt pocket and takes out a mint. 9: She carefully unwraps it and plops it into her mouth, then she puts the wrapper into the tray.

10: Now joining into the conversation with her friends she giggles, like that of a little child. 11: Her laughter makes you want to smile also. 12: There is now a sense of warmth: her posture is erect. 13: She slowly takes the mint out of her mouth and makes a quick examination of the piece of candy, and then places it back in her mouth. 14: Susan is easily rising from her seat as the sun does on a beautiful spring morning. 15: She picks up her tray and like a swan over a clear pond, glides over to the garbage bin. 16: Emptying her tray, her fork drops and she bends or rather stoops down and picks it up. 17: Dispensing her tray and utensils into the window she is smiling at the lady in the window. 18: Letting go of the tray, she turns around and is heading towards the table. 19: Telling her friends good-bye now, she is picking up her bag and adjusting the straps on her shoulder; she strolls through the door.

KEN

FIGURE 41

TEACHER'S SUPPLEMENTARY MATERIALS

PHASE: (1)

INTRODUCING
NEW LEARNING:

(A) GETTING STARTED.

1: Tell class the lesson will look closely at sentence variety.

2: Explain that a writer gets sentence variety in many ways. These include adjusting

a: openings.
b: length.
c: structure.
d: types.
and e: the position of certain elements.

3: Tell pupils you will illustrate from the close study of SUSAN.

4: Point out that they will want to adopt the techniques for analysing their own work.

(TIME: 5 minutes)

PHASE: (2)

ESTABLISHING
NEW LEARNING:

(B) SENTENCE ANALYSIS.

1: Analyse sentence 1.

2: Call on a student to read the sentence.

3: Ask another student for the basic sentence pattern.

4: Write the pattern on the board:

(I am sitting.)

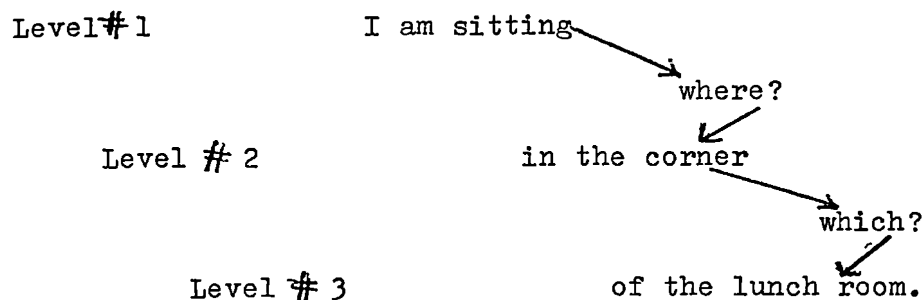
5: Tell the class that the sentence is an example of a B.S.P.i.

S + V = B.S.P.1.

SCRIPT

T: Now, tell us how this sentence is expanded?
I am sitting where?
S: In a corner.
T: Which corner?
S: Of the lunch room.
T: So how is the sentence - the B.S.P. - expanded?
S: By the addition of phrases.
T: Yes. Notice the importance of asking the right questions of the key words. We asked questions of which words?
S: Sitting and corner.
T: O.K. Somebody come and expand the chart.

N.B. STUDENT WRITES ON THE BOARD



T: Good. Sentence 1 is a simple sentence. It has one main verb only, but notice that the parts of the sentence are not of equal rank. Let us do sentence 2 and then compare the two sentences.

5: Ask for basic sentence pattern:

Level 1 Susan is eating lunch.

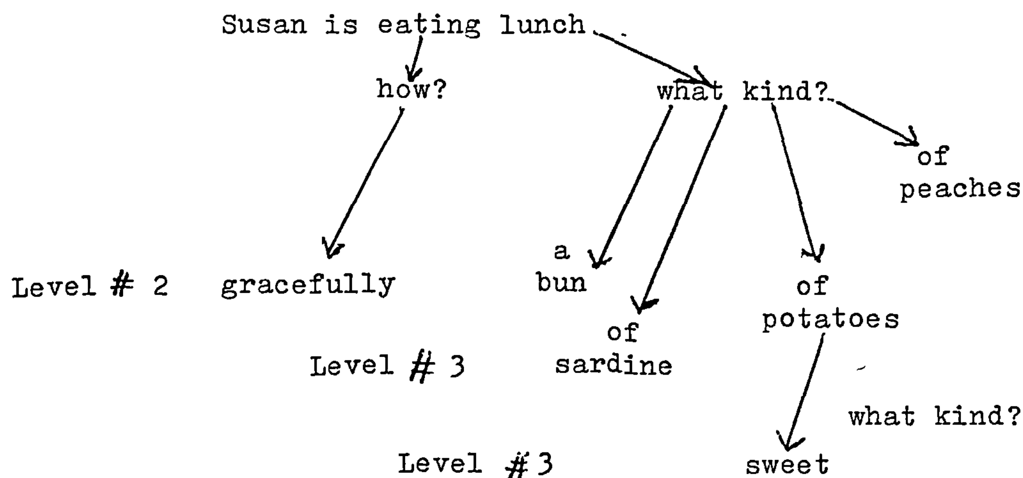
6: Explain that this is a B.S.P.2

S + V + O = B.S.P.2

7: Ask leading questions. For example:

a: Whose lunch?
b: Eating how?
c: Which word tells more about lunch?

8: Call on another student to chart sentence.
Stress hierarchical relationships.



SCRIPT

T: How is the sentence expanded?
S: By a word and a clause.
T: Read the clause.
S: "Which consists of sardine, sweet potato, peaches and a bun."
T: What is the main verb for the clause?
S: Consists.
T: What is the subject?
S: Lunch and which.
T: Yes, which is the subject of the clause? The subject is a pronoun. We can take out which and put in lunch. How will it read?
S: The lunch consists of sardine, sweet potato, peaches and a bun.
T: Good. Now let us compare sentences 1 and 2. Look at the openings. Are they different?
S: Yes.
T: How?
S: I and Susan are different.
T: How are the sentences alike?
S: The subject word comes first.
T: Are they the same length?
S: No, the second is longer.
T: Do they have the same structure?
S: No, the basic sentence patterns are different.
T: How?
S: Sentence 1 has phrases only. Sentence 2 has clauses. Sentence 2 has a list of things.
T: Notice commas separate the items in the list. Sentence 2 is called a complex sentence. It has one main clause and one dependent clause.
What about the order of the elements?
S: Nothing unusual.

(TIME: 15 minutes)

PHASE: (3)

CONSOLIDATING NEW LEARNING:

(A) GIVING PRACTICE.

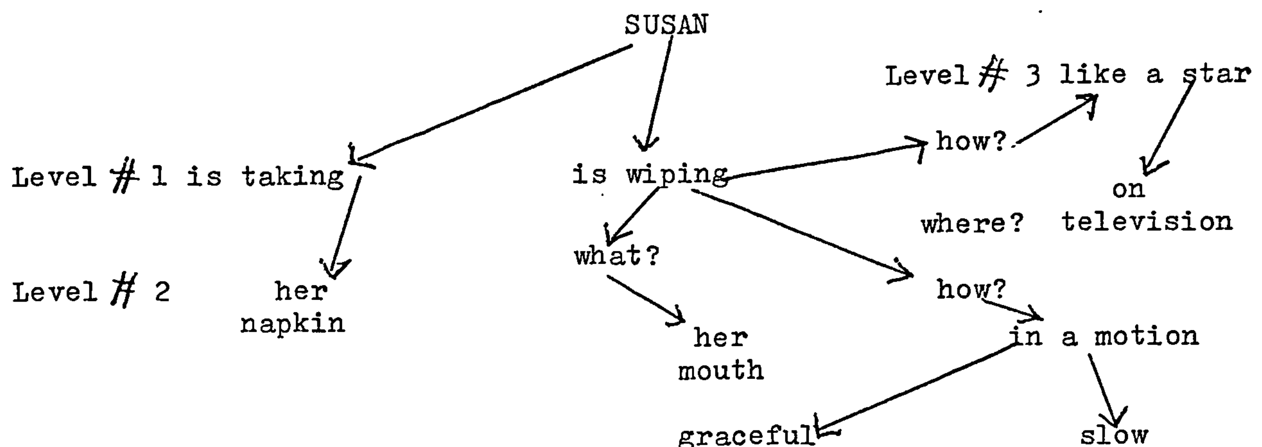
1: Analyse sentence 7.

2: Ask for the basic sentence pattern.

(S + V + O = B.S.P.2)

3: Call on students to do a chart in their books.

ONE POSSIBLE CHART



4: Ask for kind of sentence. Point out that this is a simple sentence: the same subject has 2 verbs - a compound verb.

5: Compare the 3 sentences in detail. Discuss each category: subjects, length, structure, etc.

6: Ask students to chart sentence 9. Use the same Procedure as for 7.

7: Discuss differences among the sentences.

8: Take each category in turn.

9: Match paragraph 1 against the rating scale for high.

NOTE: This paragraph is an overall HIGH paper for structure. But it is not perfect. Notice the subject word comes first in every sentence although the words themselves change frequently.

10: Ask students to rearrange sentences 5 and 9 moving the subject from the first position.

11: Ask for some sentences.

FIGURE 42

SUMMARY SHEET FOR LESSON 4: WORDING

OBJECTIVE:	Knows characteristics of a paper rated low.
LEVEL OF INTERACTION:	Vertical and horizontal. Free response between student and student, and teacher and student.
INSTRUCTIONAL ROLES:	Burden shared between teacher and student.
USE OF SPACE:	Whole class format.
MATERIALS:	HANDOUT * 5 WATER
	HANDOUT * 6 RATING SCALE FOR LOW
	HANDOUT * 7 EXERCISE

FIGURE 43

STUDENT WRITING SAMPLE

Instructions: This assignment was given after a spate of bad weather on the island. Students were to write a composition describing a scene during this bad weather. They were to pay special attention to wording, but the composition was to be rated in groups for all four qualities. They were to write comments on its structure and wording.

1: For the past several weeks, the sun had been shining brightly over the horizon every day. 2: It had been making the entire island of St.Thomas extremely hot and adding perspiration to everyone's body. 3: Well on Monday of last week all that changed.

4: On Monday the rain started to fall and at first I loved it. 5: I have always enjoyed the aroma that comes from the earth when it gets wet by rain. 6: The smell that the rain gives off makes me feel as though I can have it for supper, but I no I cannot.

7: I had hoped that it would have stopped by the end of the day but instead it carried on throughout that day and many days to follow. 8: Realising that the rain was not stopping, I began to fear that it would spoil my thanksgiving because the rain had brought a lot of disasters around my neighbourhood. 9: It filled up my backyard with water and mud. 10: It dug up out street and it kept many of my neighbours inside and had not allowed them to go out shopping for their groceries. 11: But in as much as I had taught that it would have ruined my thanksgiving day because my family and friends who would normally come to visit at these times would not have been able to do so, it did not. 12: As a matter of fact this was one of my better thanksgivings even though it was one of my most rainy ones.

FIGURE 44

WEEK: 1 ENGLISH: 12

LESSON: 4 HANDOUT: 5

HOMEWORK

EXERCISE

A PAPER RATED LOW IS DULL AND LIFELESS

There is one bright spot in Perl's paper, but before and after that the paoer fails for wording. Answer the following questions carefully. The exercise will help you to avoid this kind of writing.

(1) The Bright Spot

(A): Aroma is a well-chosen word. Write down all the reasons why you think it is well chosen.

(B): List all the words in this paragraph that create the image of eating.

(C): Although this section is a fair attempt at effective writing, there is room for improvement. For example, there are no details about smell. Add sensory details and make whatever other changes you think necessary to help the reader feel the experience.

(2) PARAGRAPH 1:

The first paragraph is not vivid. For example,

"the sun had been shining brightly"

is a worn-out statement. We hear it over and over again. It is so stale that it has no image-making power. "Extremely hot" does not help the reader much either.

(A): Revise this section to show how hot the sun was. You may wish to develop an image, or change the verbs to more effective ones.

(B): "Adding perspiration to everyone's body" is an odd statement. Make this part more effective.

(3) PARAGRAPH 3

The last section is not much better than the first. In sentences 7 and 8, she merely mentions that the rain was falling all the time. There is no sensory or concrete detail to make the point.

(A): Add sensory and concrete detail.

(4)

(B): Make a list of verbs in this section that you think are somewhat effective. Say why they are effective.

(C): Which are the least effective? Give reasons.

(5) Sentences 9 and 10 attempt to show how much rain fell. Improve sentence 9 to make it more effective.

"It filled up my backyard with water and mud."

(A) Use sensory words suggesting sound and touch. These questions should help you with this section:

* How was the water flowing?
* What colour was the water?
* Did she go outside? How did the mud and water feel?
* Did the water settle anywhere? What shape was the puddles?

(6) Some words are inappropriate or not spelled correctly. Correct them.

FIGURE 45

TEACHER'S LESSON PLAN

PHASE: (1)

INTRODUCING
NEW LEARNING:

* The teacher
    * checks for homework.
    * states objective:
        * students will
            * know criteria for a paper rated LOW.
            * share their responses to the exercise.
    * warns students to expect a variety of responses since people are different and will respond differently.

(TIME: 5 minutes)

PHASE: (2)

ESTABLISHING
NEW LEARNING:

* The teacher
    * stresses the essential features of a paper rated LOW.

    T: A PAPER RATED LOW IS DULL AND LIFELESS.
    * Explores one question at a time.

        THE BRIGHT SPOT
        * focusses on question 1.
        * calls on students across the class and lists reasons why aroma is well chosen.

* asks for words that develop the image of eating.

* calls on different students to read their rewrites.

(TIME: 15 minutes)

PARAGRAPH 2

* asks students to read changes in (A) and (B).

* compares a few responses.

PARAGRAPH 3

* asks for responses.

* discusses verb lists.

(TIME: 20 minutes)

PHASE: (3)

REVIEWING
NEW LEARNING:

* The teacher

  * collects homework.

  * distributes scale for LOW.

  * discusses scale.

(TIME: 10 minutes)

BIBLIOGRAPHY

ASSESSMENT OF PERFORMANCE UNIT (1982): Language Performance in schools, secondary survey report no.1. London: HMSO.

BARNES, D. (1982): From communication to curriculum. Bucks, England: Penguin.

BARON, J.N. (1984): Writing assessment in Connecticut: A holistic eye towards identification and an analytic eye toward instruction. Educational Measurement, 3, 1, 27-38.

BEAN, J.C. (1979): A method of peer evaluation of student writing. College Composition. 30, 301-302.

BEAVEN, M.H. (1977): Individualised goal setting, self-evaluation and peer evaluation. In: C.R. Cooper and L. Odell (Ed.), Evaluating Writing: describing, measuring, judging (pp.135-156).

BERKENKOTTER, C. (1981): Understanding a writer's awareness of audience. College Composition and Communication, 32, 388-399.

BORG, W.L., and GALL, M.D., (1983): Educational research (4th ed.). New York: Longman.

BOYD, W. (1924): Measuring devices in composition, spelling and arithmetic. London: Harrap

BRACHT, G.H., and GLASS, G.V. (1968): The external validity of experiments. American Educational Research Journal. 5, 437-74.

BRELAND, H.M. (1983): The direct assessment of writing skill: A measurement review (College Board Report No.83-6, ETS. Report No. 83-32). Princeton, NJ: Educational Testing Service.

BRELAND, H.M. (1983): Linear models of writing assessment. E.T.S. research report. Princeton, NJ: Educational Testing Service.

BRITTON, J. (1982): Language and learning. Middlesex: Penguin.

BRITTON, J. (1977): The development of writing abilities, 11-18. Urbana, 111.: National Council of Teachers of English.

BRITTON, J.N., MARTIN, N.C., and ROSEN, H. (1966): Multiple marking: An account of an experiment. London: HMSO.

BROOKS, C., and WARREN, R.P. (1972): Modern Rhetoric (Shorter 3rd ed.). New York: Harcourt Brace Jovanovich.

BRUFFEE, K.A. (1984): Collaborative learning and the 'conversations of mankind'. College English, 46, 635-652.

CALKINS, L. (1980): Children's rewriting strategies. Research in the teaching of English. 331-341.

CAMPBELL, D.T. and STANLEY, J.C. (1966): Experimental and quasi-experimental design for research. Boston, MA: Houghton Mifflin.

CARIBBEAN EXAMINATIONS COUNCIL (1979): Guidelines to markers. Bridgetown: Barbados.

CHAPMAN, C.O., FYANS, L.J., and KERINS, C.T. (1984): Writing assessment in Illinois. Educational Measurement. 3, (1), 24-26.

CONRY, R., and JEROSKI, S. (1980): The British Columbia assessment of written expression. Report prepared for the learning assessment branch of the Ministry of Education.

COOPER, C. (1975): Measuring growth in writing: Research round-up. English Journal. 64, 111-120.

DALY, J.A., and MILLER, M.D. (1975): The empirical development of an instrument to measure writing apprehension. Research in the teaching of English 9 (Winter) 242-249 (E.J. 137933).

DAVIS, B.G., SCRIVEN, M., and THOMAS, S. (1981): The evaluation of composition instruction. Pt.Reys, CA: Edgepress.

DEPARTMENT OF EDUCATION (1989): V.I. public enrollment as of September, 1988. Charlotte Amalie, St.Thomas: U.S. Virgin Islands.

DEPARTMENT OF EDUCATION (1988): Organisational chart as of September, 1988. Charlotte Amalie, St.Thomas: U.S. Virgin Islands.

DEPARTMENT OF EDUCATION (1989): V.I. public enrollment as of December, 1988. Charlotte Amalie, St.Thomas: U.S. Virgin Islands.

DIEDERICH, P. (1974): Measuring growth in English. Urbana, 111.: National Council for the Teachers of English.

DENISON, W.F. (1978): Research report: The assessment of performance unit - where is it leading? Durham and Newcastle Research Review, 40, 31-36.

DIXON, J. (1967): Growth through English. Urbana, 111.: National Association for the Teaching of English.

DOOKAN, I. (1974): A history of the Virgin of the United States. Essex, England: Caribbean University Press in association with Bowler Publishing Company.

DUSEL, J.W. (1955): Determining an effective teaching load in English. Illinois English Bulletin, 43, 1-19.

EBEL, R.L. (1979).: Essentials of educational measurement. (3rd ed.) Prentice-Hall, Englewood Cliffs: New Jersey.

EBEL, R.L., and Frisbie, D.A. (1986): Essentials of educational measurement. (4th ed.). Englewood Cliffs, New Jersey: Prentice-Hall.

EDE, L.S. (1979): On audience and composition. College Composition and Communication, 30, 291-295.

EDE, L.S., and LUNSFORD, A. (1984): Audience addressed/audience invoked: Composition in theory and pedagogy. College Composition and Communication, 35,

EDE, L.S. (1984): Audience: An introduction to research. College Composition and Communication, 35, 140-154.

ELBOW, P. (1981): Writing with power: Techniques for mastering the writing process. New York: Oxford University Press.

ELBOW, P. (1987): Closing my eyes as I speak: An argument for ignoring audience. College English, 49.

EVANS, P., BROWN, P., and MARCH, M. (1977): Criteria for the evaluation of student writing grades 7 & 8: A handbook. Toronto: Ontario Institute for Studies in Education.

FLOWER, L. (1979): Writer-based prose: A cognitive basis for problems in writing. College English. 41, 19-37.

FLOWER, L., and HAYES, J.R. (1981): A cognitive process theory of writing. College Composition and Communication, 32, 365-387.

FLOWER, L. (1981): Problem-solving strategies for writing (2nd ed.). New York: Harcourt Brace Jovanovich.

'ORD, B.W. (1973): The effects of peer editing/grading on the grammar-usage and theme-composition ability of college freshmen. Ed.D. dissertation, University of Oklahoma, 1973. Dissertation Abstracts International (1973): 73-15321.

'OX, L. (1981): What to do when grammar exercises no longer help. College Composition and Communication, 32, 332-333.

'REEDMAN, S.W. (1972): Some reasons for the grades in compositions. English Journal, 71, 86-89.

;ERE, A.R. (1987): Writing groups: History, theory and implications. Carbondale: Southern Illinois University Press.

;RIMM, N. (1986): Improving students' responses to their peers' essays. College Composition and Communication, 37, (1), 91-94.

;RONLUND, N.E. (1981): Measurement and evaluation in teaching (4th ed.) New York: Macmillan.

IARTOG, P. (1944): The marking of English essays. London: Macmillan.

IARRIS, C.E. (1986): Effective strategies in peer evaluation of writing. (ERIC document Reproduction Service No. ED 272 897).

IAIRSTON, M. (1982): A contemporary rhetoric (3rd ed.). Boston, MA: Houghton Mifflin.

IAWKINS, T. (1976): Group enquiry techniques for teaching writing. College English, 37, 637-646.

IERMANN, F., and WILLIAMS, P. (1984): Writing assessment in Maryland. Educational Measurement, 3.

JACKO, C.M. (1974): A study of the triad as an instructional mode employing predetermined roles in a small group. Unpublished doctoral dissertation. University of Pittsburg.

KATSTRA, J., TOLLEFSON, N., and GILBERT, E. (1987): The effects of peer evaluation, attitudes towards writing, and writing fluency of grade 9 students. Journal of Educational Research, 80 , 168-172.

KLINE, C.R., and MMERING, D. (1977): Formal fragments: The English minor sentence. Research in the teaching of English, 11, 87-109.

KROTHWOHL, (1974): Taxonomy of educational objectives, Handbook 11: Affective domain.

KROLL, B.M. (1984): Writing for readers: Three perspectives on audience. College Composition and Communication, 35, 172-185.

LAGANA, J.R. (1973): The development, implementation, and evaluation of a model for teaching composition which utilises individualised learning and peer grouping. Ph.D. dissertation, U. of Pittsburg. Dissertation Abstract International: 73-04127.

LAMBERG, W. (1980): Self-provided and peer-provided feedback. College Composition and Communication, 31, 63-75.

LEWIS, G.K. (1972): The Virgin Islands. Northwestern University Press: Evanston.

LONDON ASSOCIATION FOR THE TEACHERS OF ENGLISH: (1965) Assessing composition. Blackie.

LONG, R.C. (1980): Writer-audience relationships: Analysis or invention? College Composition and Communication, 31, 221-226.

MEREDITH, V.H. (1984): Writing assessment in South Carolina. Journal of Educational Measurement, 3, 19-21.

MINSKY, M. (1975): A framework for representing knowledge. The psychology of computer vision, Ed. P. Winston. New York: McGraw-Hill.

MITCHELL, R., and TAYLOR, M. (1979): The integrating perspective. College English, 41, 247-271.

MOBERG, G. (1984): Writing groups: New techniques for writing without drills (3rd ed.). New York: The Writing Consultant.

MOFFETT, J. (1968): Teaching the universe of discourse. Boston, MA: Houghton Mifflin.

MULLIS, I.V.S. (1984): Scoring direct writing assessments: What are the alternatives? Educational Measurement, 3, 16-18.

ONG, W.J. (1975): The writer's audience is always a fiction. Modern Language Association, 90, 9-21.

PARK, D. (1982): The meaning of audience. College English, 44, 247-257.

PARK, D. (1986): Analysing audiences. College Composition and Communication, 37, 478-488.

PFISTER, F.R., and PETRICK, J. (1980): A heuristic model for creating a writer's audience. College Composition and Communication, 31 213-220.

PIERSON, B. (1967): Peer and teacher correction: a comparison of the effect of two methods of teaching composition in grade nine English classes. Ph.D. dissertation, New York University.

POPHAM, W.J., and SIROTNIK, K.A. (1973): Educational Statistics: Use and interpretation (2nd ed.). New York: Harper and Row.

PROTHEROUGH, R. (1983): Encouraging writing. Methuen, London.

PRATT, D. (1980): Curriculum design and development. New York: Harcourt Brace Jovanovich.

PURVES, A.C. (1984): The teacher as reader: An anatomy. College English, 46, 259-265.

QUELLMALZ, E.S., CAPELL, F.J., and CHOU, C. (1982): Effects of discourse and response mode on the measurement of writing competence. Journal of Educational Measurement, 19, (4).

ROBINSON, G.R., ROSS, A.J., and WHITE, F. (1983): Curriculum development for effective instruction. Toronto: Ontario Institute for Studies in Education.

ROSEN, H. (1973): Written language the sense of audience. Educational Research, 15.

ROTH, R.G. (1987): The evolving audience: Alternatives to audience accommodation. College Composition and Communication, 38, 47-55.

RUSSELL, C. (1985): Peer conferencing and writing revision: A study of relationships. (ERIC document Reproduction Center No.ED 260 392).

SACHE, P.P. (1984): Writing assessment in Texas: Practices and Problems. Journal of Educational Measurement, 3, 21-23.

SAGER, C. (1973): Improving the quality of written composition through pupil use of a rating scale. Ed.D. dissertation, University of Michigan, Ann Abor, Mich.: University Microfilms. Order no. 73-23605.

SATTERLEY, D. (1981): Assessment in schools: Theory and Practice in education. Oxford: Basil Blackwell.

SEQUENTIAL TEST OF EDUCATIONAL PROGRESS: Handbook for series 11. (1971). Princeton, NJ: Educational Testing Service.

SHAUGHNESSY, M. (1977): Errors and expectations. New York: O.U.P.

SNOW, R. (1974): Representative design and quasi-representative designs for research on teaching. Review of Educational Research, 44, 265-291.

SPENCER, E. (1983): Writing matters across the curriculum. Edinburgh: Scottish Council for Research in Education.

STANFORD, G.S. (1979): Classroom practices in teaching English: How to ease the paper load. Urbana, 111.: National Council for Teachers of English.

THE DAILY NEWS: Saturday, August 20, 1988, pp.1,2. St.Thomas, Virgin Islands.

THE DAILY NEWS: Friday, March 24, 1988, p.2. St.Thomas, Virgin Islands.

TURNBULL, C. (1981): A brief history of Virgin Islands education. Virgin Islands Educational Review 1 (1), 20-21, 24.

TURNBULL, C. (1988): The evolution of the Virgin Islands board of education. Social studies newsletter, p.4.

WALZER, A.E. (1985): Articles from the California divorce project: A case study of the concept of audience. College Composition and Communication, 36, 150-158.

WHITE, J. (1986): The assessment of writing: Pupils aged 11 and 15. Windsor: NFER-Nelson.

WITTE, S.P., and FAIGLEY, L. (1981): Coherence, cohesion, and writing ability. College Composition and Communication, 32, 189-204.

WIENER, H.S. (1986): Collaborative learning in the classroom: A guide to evaluation. College Composition and Communication, 48, 52-61.

WILKINSON, A., BARNSLEY, G., HANNA, P., and SWAN, M. (1980): Assessing language development. Oxford University Press.

WINTEROWD, W.R. (1975): Contemporary rhetoric: A conceptual background with readings. New York: Harcourt Brace.