DOCUMENT RESUME

ED 146 202 TM 006 582

AUTHOR McDonald, Frederick J.; And Others
TITLE The Effects of Classroom Interaction Patterns and Student Characteristics on the Acquisition of Proficiency in English as a Second Language. Final Report. Volume I.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY New Jersey State Dept. of Education, Trenton.
REPORT NO ETS-PR-77-5
PUB DATE May 77
NOTE 339p.; For related documents, see TM 006 580, 581, and 583 ; Tables are marginally legible due to small type

EDRS PRICE MF-$0.83 HC-$18.07 Plus Postage.
DESCRIPTORS Academic Achievement; Achievement Gains; Adult Basic Education; *Adults; Age Differences; Day Schools; Educational Background; *English (Second Language); Ethnic Origins; Interaction; Language Instruction; *Language Proficiency; Language Tests; Observation; Oral English; Performance Factors; Predictor Variables; Sex Differences; Spanish Speaking; Statistical Analysis; *Student Characteristics; Student Teacher Relationship; Teacher Behavior; Teacher Characteristics; Teacher Qualifications; Teaching Skills; Teaching Styles; *Teaching Techniques
IDENTIFIERS Oral Proficiency Test

ABSTRACT

Teachers and students in the English as a second language classes at the West New York (New Jersey) Adult Learning Center participated in a study to determine which patterns of classroom interaction and student characteristics were most highly related to the acquisition of oral proficiency in English. Differences in teaching styles and performances were related to the acquired proficiency of the students, who varied considerably in age, education, and previous experience with English. Teachers used a variety of instructional techniques; however, the audio-lingual and silent method were the principal techniques used. Student proficiency was measured at two points in time, and the intervening instruction was observed daily. Measures of speaking facility were either direct measures of proficiency, such as the Oral Proficiency Test developed for this project, or other measures of knowledge of English. Statistical analyses were used to relate differences in teacher behavior to gains in students' oral proficiency. Results indicated that teaching style did affect proficiency, particularly for certain types of students. This volume describes the research methodology, presents details of the statistical analyses, and summarizes the results. (Author/MV)

PR-77-5

THE EFFECTS OF CLASSROOM INTERACTION PATTERNS AND STUDENT CHARACTERISTICS ON THE ACQUISITION OF PROFICIENCY IN ENGLISH AS A SECOND LANGUAGE

FINAL REPORT

VOLUME I

By

Frederick J. McDonald
Meredith K. Stone
Allen Yates

May 1977

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

FINAL REPORT
VOLUME I

7

THE EFFECTS OF CLASSROOM INTERACTION PATTERNS AND STUDENT CHARACTERISTICS ON THE ACQUISITION OF PROFICIENCY IN ENGLISH AS A SECOND LANGUAGE

Frederick J. McDonald

Meredith K. Stone

Allen Yates

Educational Testing Service
Princeton, New Jersey

A project conducted by Educational Testing Service for the West New York, New Jersey Board of Education and funded by the New Jersey State Department of Education

PREFACE

This report describes a study of the teachers and students in the West New York Adult Learning Center. The purpose of the study was to find out which patterns of classroom interaction and students' characteristics were most highly related to the acquisition of oral proficiency in English by adults who were learning English as a second language.

The report is written in two ways. The major portion, Chapters Three and Four, are written primarily for those interested in the details of the statistical analysis. Chapters One, Two, the first part of Chapter Three, and Chapter Five have been written so that the reader who does not have a statistical background may understand the methodology and results of the study. Chapter Five is in fact a summary of the entire study without the detailed explanations of the methods presented particularly in Chapters Three and Four. Chapters One and Five were written by the Project Director, Chapter Two and the description of the Observation System in Chapter Three by Meredith Stone, Associate Project Director, and the analysis of the observational data in Chapter Three and Chapter Four by Allen Yates, Research Statistician.

We wish to express our appreciation to Arthur Von Schalscha, Director of Continuing Education, Kathleen Durnin, Coordinating Teacher, Mariluz Garcia and Robert Layton, Tester-Observers, Diane Cappucilli, Secretary, and the 12 participating teachers for their assistance and cooperation.

We also wish to thank Joanne Farr for her dedicated and outstanding performance as Project Secretary.

TABLE OF CONTENTS

Page

Page

Page

Appendices Listed Below Appear
in
VOLUME II
of
The West New York Project Final Report

APPENDIX A STUDENT MEASURES

APPENDIX B TEACHER QUESTIONNAIRE

APPENDIX C OBSERVATION LEXICON

APPENDIX D DATA PROCESSING

APPENDIX E CLASSROOM INTERACTION FACTOR PATTERN AND FACTOR INTERCORRELATION MATRICES

CHAPTER ONE

THE DESIGN OF THE STUDY

INTRODUCTION

This report describes the work conducted over one year in the West New York Adult Learning Center to determine the relation between teaching strategies, methodologies, and performances and the acquisition of facility by the students at the Center in speaking English as a second language. The West New York Adult Learning Center provides a training program in which adults learn to speak English as a second language. The adults attend classes either during the day or in the evenings. Instruction is provided at three different levels adapted to the proficiency in English which the students have already acquired.

The purpose of this project was to determine which kinds of teaching performances (sometimes called competencies) were directly associated with differences in acquired proficiency in English on the part of the students. The methodology used in the study related differences in teaching styles and performances among the teachers to differences in the acquired proficiency of their students.

Acquired proficiency in English in this case meant acquisition of the ability to speak English. The Center does not attempt, except indirectly, to improve students' abilities to write or read English. The instruction is directed primarily to stimulate acquisition of facility in speaking English, and the procedures used by the teachers rely heavily on oral discourse between teachers and students. The students themselves vary considerably in age, previous education, and previous experience with English.

In instructing the students the teachers use a variety of techniques and methods, though two methods, the audio-lingual and the "silent way," are the principal methods used. Some teachers adhere to one method; others choose among methods.

Relatively little formal research has been done on the efficacy or efficiency of specific teaching procedures, though there are advocates of one or another method. The approach taken in this study was to look at all procedures used by the teachers and attempt to identify those which, irrespective of method, were most highly related to gains in acquired proficiency in speaking English. It is conceivable that some teaching performances are highly effective irrespective of the general method in which they are embedded. Or, it may be that a cluster of teaching performances characteristic of one method may be more effective than those of another method. A third possibility is that some procedures or methods may be more effective at one stage in the learner's acquisition of speaking proficiency, and others are more effective at later stages.

The general methodology used in the study (which will be described in greater detail in the following pages) required us to measure student proficiency at two points in time, and to observe daily the intervening instruction. The measures of speaking facility were either direct measures of proficiency, such as the Oral Proficiency Test developed specifically for this project, or were other measures of knowledge of English. The observations provided a continuous record of both teacher and student behavior during class sessions and were made between the two points of assessment of the students' speaking proficiency.

CONCEPTUAL BASIS OF THE METHODOLOGY

The methodology used in this study is built on two major hypotheses and a few minor assumptions. The first of these hypotheses is that among a group of teachers of English as a second language are some who are more effective in producing speaking proficiency in English than others. This hypothesis does not imply that the less effective teachers are necessarily incompetent or have no success in facilitating the acquisition of English as a second language. It does imply that some teachers are more skilled than others in helping their students make greater gains in acquiring the ability to speak English proficiently.

This hypothesis is an old idea--in any profession or craft there seems to be practitioners who are recognizably more effective than other practitioners. These more effective practitioners presumably possess skills which account for their relatively greater efficacy. The purpose of the methodology is to identify these more effective practitioners and to identify what it is that they do which accounts for their greater success in producing proficiency in speaking English.

The second hypothesis was that a teacher's performance, that is what the teacher does, accounts for the acquisition of proficiency. Teachers ask questions, provide models of desired, correct, or appropriate responses, and indicate to the student when he or she has made an appropriate or an inappropriate response in speaking. These and similar actions of the teacher elicit the desired speaking responses, shape or modify them, or stimulate the processes by which the students attempt to generate English statements.

The teachers' attitudes toward the students and toward the language being learned and toward the language and culture of the students are undoubtedly important factors in stimulating the students to try to learn English and in their responsiveness to instruction. Differences among the teachers in these characteristics are small and probably would be unrelated to differences in the students' learning. The teachers in the Center are carefully selected for this work; they work closely together; and the atmosphere of the Center is warm, friendly and supportive. The staff knows the students well and accommodates instruction to their needs.

But the actual instruction is mediated by the kinds of performances in which the teacher speaks English and helps the student to speak English. It is, therefore, presumed that the strongest relations between what the teacher does and student learning will be found by looking at those performances which are directly instructional in character.

Differences in the characteristics of the students may also account for differences in their acquired proficiency. Some may have more aptitude for learning languages. Or the differences in their achievement of proficiency may be related to their previous experience with English or their educational attainments or their socioeconomic background and current status. The competing hypothesis to the one that attributes differences in proficiency to differences in teaching styles is that it is such differences as these among the students which account for their differences in learning.

In this study information was gathered on a variety of the students' characteristics which were statistically related to differences in proficiency in speaking English achieved by the end of the school year. This study thus provides data on the degree to which specific teaching performances or combinations of them account for differences in students' acquired proficiency and on the degree to which differences among the students themselves account for their acquired proficiencies.

Finally, the assumption was made that correlations between teacher performances and gains in student learning are a reasonable basis for ascertaining which teacher performances are most likely to influence students' acquisition of English as a second language. We are well aware that correlation does not necessarily imply causation but we are also aware that where correlations are found causal connections of some kind may be inferred, subject to verification by careful experimental study. This study, therefore, was designed to find out if such correlations existed and their relative size, and from such correlations to conceptualize a hypothetical picture of teaching performance likely to be associated with greater gains in acquiring proficiency to speak English as a second language.

Of the two hypotheses mentioned above, the first is the more tenuous. While it seems reasonable to hypothesize that there are varying degrees of teaching effectiveness, it is possible that in selecting a particular sample of teachers one may not find differences among the teachers either in the effects they produce ~~or in their~~ teaching skills. Thus, a study of this kind inevitably runs a risk that the particular sample of teachers involved will not produce the necessary information. But practical experience suggests that in almost any group of teachers one can find teachers who are more effective than other teachers, and it was on the basis of this practical experience that this study was conducted.

The second hypothesis is a theoretical one. It postulates that the learner is directly affected by those stimuli which impinge on him or her. But the learner processes these stimuli in some way so it is equally reasonable to assume that characteristics of the learner are also directly related to how much the learner achieves. For the latter reason, we have gathered as much information as was possible in the context of this study about the characteristics of the learner to determine the extent to which these characteristics might influence learning prior to and independent of any influence of teacher performance.

It is conceivable that some teaching performances in conjunction with certain characteristics of students will provide either highly effective or relatively ineffective combinations related to the amount of learning achieved by the students.

The final assumption is no more than a statement of what is implied by finding an association or correlation between two sets of events. One can neither draw strict inferences about a causal relation between the events nor can one dismiss the possibility of one. The analytic methods used in this study examined the hypothesis that teaching performance must be taken into account in order to better predict student learning, with the apparent implication that teacher performance influences how much students learn. However, experimental studies have to be done to test the validity of any causal hypotheses derived from a correlational study such as this.

These hypotheses and assumptions outlined above are the underpinning of the methodology. The essence of the methodology is to relate variation in teaching performance to differences in the degree to which students achieve proficiency. If correlations are found between measures of teaching performance and measures of student proficiency, such relations become the basis for establishing hypotheses about causal connections between teaching performance and student learning.

The above paragraphs suggest the limitations of this study as well as its potential value. At best, from a study of the type conducted, one can examine the hypothesis that certain types of teachers' actions or activities make a difference in the accuracy with which we can predict students' learning. The data reported do not represent proof that there is a direct connection. One can also estimate how much of a role any given teaching performance plays in the prediction of students' learning, but the same caution in making causal inferences applies here also.

The multivariate statistical methods used in the study are appropriate for data of the kind gathered. Factor analysis and canonical discriminant function analysis were used to reduce both student-performance data and teacher-performance data to their underlying dimensions. Multiple regression, canonical correlation, and factor analytic methods were used to relate the dimensions of teacher performance to those of student performance. In the following chapters the basic descriptive and correlational data among a variety of student performance and teacher performance variables are presented. The analysis then moves on to multivariate studies of classroom interaction. It concludes with an analysis of how student achievement and classroom interaction variables relate to each other. In these analyses account is taken of students' background characteristics and their proficiency in English at the beginning of the data-gathering phase of this study.

THE MEANING OF PROFICIENCY IN SPEAKING ENGLISH

There are obviously degrees of proficiency in speaking English, even among native speakers. Adults learning English as a second language will also acquire skill in the language to varying degrees. But what is meant by "different degrees of proficiency"? The meaning of proficiency needs to be operationally defined before undertaking a study of the kind described here.

The operational definition of proficiency used in this study was derived by an analysis of the goals of the West New York Adult Learning Center. These goals are to facilitate the acquisition of English as a second language so that:

(1) the adult learners can understand conversational English; (2) they will be able to communicate in English in ordinary situations so that they are adequately understood; and (3) they will acquire the basic structures of the language so that they are likely to continue to grow in proficiency.

The meaning of these goals is best understood by thinking about how the learners who come to this Center will use English. These adults are immigrants to the United States who have been in this country varying amounts of time. The majority of them have lived in a community where speaking Spanish is the norm. Most, but not all students at the Center, come from Spanish-speaking countries, principally Cuba. They wish to acquire sufficient proficiency in English so that they can communicate with non-Spanish speaking people in stores, in clinics and hospitals, in schools, in places of employment, and in interactions with government agencies of one kind or another. In other words, they wish to become, and the goal of the school is to help them become, functionally proficient in speaking English. The aim is to help them achieve sufficient facility in English so that they are in fact bilingual for the practical purposes of everyday living.

Three levels of performance in speaking English related to this goal may be distinguished. The first level is acquisition of the language such that the person understands ordinary communications to her or him; for example, a person is asked simple questions such as occur in everyday conversation and is able to understand the question being asked even though they cannot always provide a full or accurate answer. A person who has attained this level of proficiency is able to understand most simple communications, but cannot respond adequately and with facility.

A second level of proficiency is represented by the learner being able to respond to questions or to make statements about himself or herself, what they plan to do, what they think on practical matters, and the like. To communicate at this level, a person must have acquired the basic structures of the English language.

A person has attained the third level of proficiency when he or she can generate questions and statements on his or her own, can extend discourse through a series of statements or questions, and in speaking uses more complex structures.

Each of these levels of proficiency may have one or more of three characteristics: (1) the person spoken to may give evidence of understanding the language spoken to him but does not respond with facility or accuracy or completeness; (2) the person may both understand and use appropriate structures but may make errors in the use of the language; (3) the person may both understand and respond with appropriate structures and use them accurately.

Thus, there are two underlying concepts by which proficiency has been described. One of these concepts describes the level of language usage available to the individual. On this dimension performance ranges from sufficient usage to comprehend what is being heard to the ability to generate relatively complex structures in extended discourse.

The other dimension is that of the accuracy of the form of the communication. Accuracy means that a person uses English sentences which are structurally correct and (by implication) has also used words correctly.

The West New York Adult Learning Center defines the kinds of structures that are to be acquired at each level of instruction. The acquisition and use of these structures defines operationally what is meant by proficiency. Proficiency means facility and accuracy in the use of these structures.

The Oral Proficiency Test

An important aspect of this study was the development of an oral proficiency test which was administered to students at the Center in the spring of 1976. This test (which will be described in a later chapter) utilized three formats for eliciting the speaking of English. In one part of the test the examiner asked the student questions about himself or herself to which the student may respond. The purpose of this procedure was to see if the student could comprehend the question or statement, and could respond appropriately. In the second part of the test the students were presented with pictures of two events and asked to describe what was happening. The purpose of this procedure was to estimate how well students could generate language freely. In the third part the learner was presented with cartoons from a Spanish newspaper and was asked to explain the cartoon. The purpose of this procedure was to estimate how easily the student could move from idiomatic Spanish to equivalent English statements. We assume that part of the process of acquiring language facility is to acquire the ability to transform concepts expressed in one language into equivalent concepts expressed in a second language. This third assessment format was designed to estimate how well these students could make such transformations.

The students' responses were scored on three characteristics: (1) comprehension; (2) use of structure; and (3) correctness. Comprehension meant that the student showed by the response that he or she understood the question but may have or may not have responded accurately. Use of structure meant that in responding the student used the appropriate semantic and syntactical forms of English but may or may not have made errors in the other parts of the response. Correctness meant that the student spoke one or more English sentences which were appropriate and were grammatically accurate in every respect. Scores on these dimensions for each student were the indicators of achieved proficiency in speaking English.

Other Measures of the Effects of Instruction

Although the goal of the Adult Learning Center is to develop functional proficiency in speaking English, other developments in the use of language may occur as a consequence of the instruction provided. The Center gives relatively little attention to formal instruction in reading, and then mainly at the highest level of instruction. It is possible that acquiring facility in speaking the language, seeing English words written on the board, reading papers and magazines and being able to decode the words because of the language instruction, the learner may acquire greater facility in reading the language.

We, therefore, used a functional reading literacy test as one measure of increased language facility. A set of items had been developed in another project that measured the ability of English-speaking adults to perform functional reading tasks. A functional reading task, for example, is reading labels of bottles, reading signs and directions, reading instructions on forms and similar tasks which one meets in everyday life that require some reading skill. This test was administered to the students at the Center twice in the year, once as a pretest and again as a posttest, and changes in performance were calculated. The purpose of this analysis was to determine if the learners had improved their ability to read as a beneficial side effect of their improvement in speaking English.

Another measure, a decoding test, was used on the presumption that one aspect of acquiring language proficiency is to acquire decoding skills in the language being learned. This decoding test had been used in the past to measure the decoding skills of English-speaking children but is usable with adults because a speaker of another language learning to speak English has to acquire the same set of skills, even though they may acquire them in a different way. Since the graphemes of English and Spanish are highly similar, the adult learning English is acquiring new or different phoneme-grapheme correspondences. The teachers may give some attention to these differences to improve understanding and pronunciation; for the students may detect them on their own. In either case we would expect some improvement in decoding skill. The decoding test was also administered as a pre- and posttest and estimates were made of the amount of change in this performance skill.

We presumed that one of the consequences of constant exposure to the phonology of English would be an increase in these skills. We also presumed that any increase in decoding skills was a beneficial consequence because the learner will have acquired a set of skills that can be transferred to a variety of situations and which should help him or her continue to develop proficiency in speaking English.

Two other measures of proficiency were also used, the John Test and the Morano Test. The Center had been using these tests for several years to estimate their students' language proficiency. They were administered as pretests in this study and scores on them were correlated with scores on other tests.

There were two reasons for using the John and Morano Tests. One of these was psychometric: to obtain information on the reliability and validity of these tests. Such information is necessary to determine whether or not the procedures used to assign students to levels of instruction are accurate and effective. The other reason was that the set of five measures used in the study (the John Test, the Morano Test, the Oral Proficiency Test, the functional reading Literacy Test, and the Decoding Test) could be analyzed to study the extent to which language proficiency was a multi-factor skill.

To recapitulate the main ideas in this section: the principal measure of the effects of instruction is the scores on the Oral Proficiency Test. We asked the question: to what extent do teachers who use different methods of teaching differ in the degree of oral proficiency their students have acquired as measured by this test? We also asked: to what extent has the instruction had other effects such as the acquisition of simple reading and decoding skills? We also asked if two other measures of knowledge and proficiency in the English language (John and Morano Tests) were related to these measures (Oral Proficiency, Decoding and Literacy Tests).

THE MEASUREMENT OF TEACHING PERFORMANCE

In this study we were concerned with how teachers teach English as a second language. The words, "how they teach", imply that we wanted an accurate description of how the teachers organized the classes for instruction, what materials they used, how they interacted with the students, and the content that they taught. The best way to compile a description of this kind is to observe what the teachers do as they teach class. We therefore assigned observers to make daily observations of each of the teachers in the classes of the Adult Learning Center.

The method of observation (described in a later chapter) was developed by observing the teachers for a period of several months. The purpose of these preliminary observations was to familiarize ourselves with how the teachers taught.

A category system which described the teaching activities was constructed from this information. This category system was then tried out systematically and further refined. The final product was a set of categories and a method for observing that provided descriptions of the activities typically occurring in the classrooms (unusual or infrequent activities were noted by the observer when they occurred). During the actual observation the observer checked continuously those categories which described what the teachers and students were doing.

Theoretical preconceptions did not determine what we should or would observe in a class. The categories, however, do include descriptions of teaching performances associated with two different theories of language instruction, but the reason that the system includes these categories is that behavior relevant to them had been observed in the classes of the Center. These two methods are the audiolingual and the "silent way". The basic elements of the audiolingual method require the teacher to model appropriate speech, elicit students' responses, and give corrective feedback. Different teachers combine these elements in different ways; for example, some teachers do relatively little modeling whereas others follow regularly a sequence of model, student practice, and corrective feedback. The "silent way" method relies heavily on non-verbal cues from the teacher to elicit speaking responses on the part of students.

Since these two methods are used by different teachers in the Center, the category system reflects what we observed them doing, and contains categories to fully describe either method. It is possible, therefore, to study three problems: (1) do the teachers adhere to a method (such as the audiolingual) or vary among themselves in how they use it; (2) is there any evidence that one method is more effective than another, either for all students or certain kinds of students; (3) are there elements in either method which are particularly effective?

These two methods are anchoring frames for teaching styles. The methods are not used mechanically or in a rote fashion. One of the goals of this research was to study the actual teaching styles used by the teachers in the Center to see how either of these two methods was used in practice and how elements of both may have been used by a particular teacher.

The observational data were gathered after the first testing of the students and was terminated shortly before the second testing. The logic of this method is that changes in student performance from the first testing to the second testing should be related to what the teachers did in the intervening period.

ANALYSIS OF THE DATA

Because the data are complex and particularly rich, they may be analyzed in a variety of ways. An important problem to be solved was to reduce the number of categories of the observational data. Several different analytic procedures were used to uncover the dimensionality of these data. The day-school data could be reduced to nine factors, and the night-school data to eight. To make the analyses for both groups comparable, nine factors were used in all analyses of the observational data. The factors were readily interpretable and correspond to our judgments of what was occurring in the classes.

The student achievement data were analyzed in two ways. First, the usual psychometric analyses were performed such as item analyses and interval consistency reliability estimation. Second, scores on all the tests were intercorrelated with each other and with measures of student characteristics. Factor analyses were also performed on these data. The purpose of these factor analyses was to find out if there was an underlying structure in the data; for example, we wanted to know if the measures of proficiency in English were measuring the same or different dimensions.

The major data-analytic aim of the study was to relate the data on teaching performance to the student achievement data. The measures of achievement focused upon proficiency in speaking English, decoding skill and functional reading skill. Regression methods were used to relate achievement to classroom interaction experiences. In all analyses the students' initial level of achievement and background characteristics were taken into account. Only those students' data were used who were tested at the beginning and end of the study period.

The following chapters present each of these analyses. The second chapter contains the description of the program, information on the tests used--how and why they were constructed, the students' performances on them, their reliabilities and intercorrelations, and related information--as well as student and teacher background information. The third chapter describes how the observational system was constructed and used and the analyses of these data. The fourth chapter presents the results of relating student achievement to teaching performances.

We conclude in the fifth chapter with a statement about what appear to be effective teaching performances and practices which should be regarded as hypotheses. This study, because of its exploratory correlational design, could not confirm in the strict scientific sense that any teaching performance directly affected achievement. We could infer, however, that it was likely that it did because of our statistical control for students' background characteristics and initial level of achievement. Replications of this study with other teachers or systematic experimentation is needed before definitive prescriptions about effective teaching performances can be made. When drawing conclusions about effective teaching practices it should be remembered that there may be other practices which may be as effective but which were not observed being used by the teachers in this study. If these qualifications are kept in mind, the results of this study can be used to improve our understanding of the teaching of English as a second language.

CONCLUSION

This chapter described the purposes and methods of the study conducted on teaching performance and student learning at the West New York Adult Learning Center. The following chapters will describe in detail the methods used and the results of the study. The methodology of the study provided information about the acquisition of speaking proficiency in English and about how English as a second language was taught to adult learners. The goal of the study was to determine what relations existed between how teachers taught and how much students learned.

CHAPTER TWO

STUDENT AND TEACHER MEASURES

DESCRIPTION OF THE PROGRAM

Day-School Classes

Fourteen day school classes taught by six different teachers were available for study. (A 15th class had to be dropped from the study in February when the teacher left for another job.) The majority of the classes met for an hour-and-a-half a day, five days a week; however, two classes met only one hour a day, one intermediate class met for an hour-and-a-half but only three times a week, and an advanced class met for an hour-and-a-half twice a week. These differences were taken into account in the analysis.

All classes met at the Adult Learning Center which is located in an office building in downtown West New York on the main bus line, although many of the students lived within walking distance. The Center operated on the West New York school system calendar and all school vacations, holidays and snow days were observed.

Night-School Classes

A sample of six teachers from the 22 available was chosen according to the following criteria: (1) they did not also teach in the day school; (2) their classrooms had a sufficient number of "representative students" as defined by the student background information; (3) there was a range among classes in student competence level and (4) there was a range among teachers in previous training and experience. Five of the six classes met for a two-hour session three evenings a week. The sixth class met only two evenings a week.

All classes were held in Memorial High School which is convenient to public transportation, although again many students lived within walking distance and many others drove to class. Night-school classes were also held according to the public school calendar.

STUDENT MEASURES: BACKGROUND INFORMATION

In addition to the battery of tests (discussed below) which were administered to assess student learning, each student supplied background information which provided a concrete description of the total sample and allowed us to test for comparability among subsets of the students and to investigate the relation between certain background variables and present learning.

Student Information Sheet

The background information collected on the Student Information Sheet was as follows:

1. Sex
2. High school diploma
3. Age
4. Education level
5. Time in United States
6. Study of English in former country
7. Study of English in United States
8. Country of origin
9. Occupation in former country
10. Occupation in United States

This information was coded (see Appendix A), keypunched and added to the students' file on the computer.

Descriptive Statistics

Day School

Background information was collected for 148 day-school students. Table 2.1 presents the descriptive statistics for the total group and by level for the day-school pretest sample and the matched sample. The matched sample contains

TABLE 2.1

Day School Student Background Information --
Pretest and Matched Samples: Descriptive Statistics

	Pretest Total Freq.	Pretest Total Percent	Pretest Freq. by Level 1	Pretest Freq. by Level 2	Pretest Freq. by Level 3	Matched Total Freq.	Matched Total Percent	Matched Freq. by Level 1	Matched Freq. by Level 2	Matched Freq. by Level 3
Number	148		52	42	54	81		24	28	29
Sex: M	54	36	22	21	11	26	32	9	13	4
F	94	63	30	21	43	55	68	15	15	25
Diploma: Yes	63	43	17	15	31	32	40	8	10	14
No	83	56	34	27	22	49	60	16	18	15
Age: Mean	45		46	44	44	46		48	45	45
Range	19-73		21-69	19-70	24-73	19-70		21-69	19-70	24-67
Education: Mean Years	10		9	10	12	10		9	10	11
Range	3-17		3-16	4-16	4-17	4-16		4-16	4-16	4-16
Time in U. S. Mean Years	6		5	6	7	6		6	6	7
Range	1-24		1-15	1-22	1-24	1-24		1-15	1-17	1-24
Former English Mean Years	1.08		.58	.83	1.83	1.17		.83	.75	1.83
Range	0-12		0-5	0-6	0-12	0-12		0-5	0-2	0-12
English in U. S. Mean Years	1.08		.50	1.41	1.33	1.17		.58	1.25	1.50
Range	0-12		0-1.5	0-12.0	0-3.0	0-3.0		0-1.5	0-2.5	0-3.0
Former Country										
Columbia	8	5	4	4	0	3	4	1	2	0
Cuba	123	83	42	35	46	73	90	22	25	26
Dominican R.	3	2	2	0	1	1	1	0	0	1
Ecuador	2	1	1	0	1	1	1	1	0	0
Peru	4	3	3	0	1	0	0	0	0	0
Other	8	5	0	3	5	3	4	0	1	2

Occupation*	Pretest T	Pretest 1	Pretest 2	Pretest 3	Matched T	Matched 1	Matched 2	Matched 3
1. Foreman	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
2. Craftsman	7/4	3/1	3/2	1/1	1/0	1/0	0/0	0/0
3. Semi-skilled Worker	14/44	7/15	7/16	0/13	6/20	4/4	2/7	0/9
4. Laborer	0/1	0/0	0/0	0/1	0/0	0/0	0/0	0/0
5. Household Worker	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
6. Personal Service	5/11	2/3	2/5	1/3	2/7	0/2	2/5	0/0
7. Fireman/Policeman	0/1	0/0	0/1	0/0	0/1	0/0	0/1	0/0
8. Professional	38/4	6/0	12/0	20/4	23/1	5/0	8/0	10/1
9. Technician	5/2	1/0	1/0	3/2	4/1	0/0	1/0	3/1
10. Farmer	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
11. Farm Worker	2/0	2/0	0/0	0/0	1/0	1/0	0/0	0/0
12. Business Owner	2/0	2/0	0/0	0/0	1/0	1/0	0/0	0/0
13. Manager/Official	2/1	0/0	1/0	1/1	1/1	0/0	1/0	0/1
14. Office Worker	24/9	8/0	8/4	8/5	19/7	7/0	6/4	6/3
15. Salesperson	8/7	5/1	1/3	2/3	4/6	1/0	1/3	2/3
16. Housewife	28/32	12/14	4/2	12/16	13/17	3/5	4/2	6/10
17. Unemployed	0/29	0/15	0/9	0/5	0/19	0/12	0/6	0/1
18. Student	9/1	1/1	3/0	5/0	6/1	1/1	3/0	2/0

*Previous country/USA
See Appendix A for definitions of categories.

those students for whom we had complete test data from both pre- and posttest administrations. Class level was designated by Center personnel: level 1 = beginner, level 2 = intermediate, level 3 = advanced. (Students are assigned to classes on the basis of their John Test scores.) As is evident from the sizes of these two samples (148 vs. 81), there was a 45 percent attrition rate between November and June for day-school students.[1]

While examination of the descriptive statistics in Table 2.1 indicated that the two samples were similar, this assumption was tested statistically. Conventional t-tests were run on each variable to detect any significant differences between the students that dropped out (N = 67) and those that remained (N = 81). None of the differences were significant. While there was a trend for the students who were younger, housewives, semi-skilled workers or unemployed to drop out, the proportions within categories remained similar in the pretest and matched samples.

We can thus state with confidence that in spite of the high attrition rate and the resulting smaller sample on which to do our analyses, the day-school matched sample is representative of the type of students who attend classes at the Center.

Night School

Background information was collected for 181 night-school students. Analysis of these demographic data supplied the criteria for selecting a sample of night-school students. This procedure assured that the sample chosen would be representative of the types of students attending night-school classes.

---

[1]Since complete test data were required for our analyses, this figure may be slightly inflated.

Table 2.2 presents the descriptive statistics on the background variables for the night-school sample. Since there was so little attrition between the pretest (N = 46) and matched (N = [illegible]) samples, only information for the matched sample is given.

Comparison of Day and Night-School Samples

An examination of Tables 2.1 and 2.2 reveals a number of differences between the day and night-school samples. The percentage of males is higher for the night-school, (44 percent vs. 36 percent), fewer students have their high school diplomas (17 percent vs. 43 percent), the average age is younger (42 vs. 45). In addition, night-school students on the average have studied English in the United States six months longer (1.67 years vs. 1.17 years) than day-school students, and only 61 percent come from Cuba as compared to 90 percent of the day-school students. Also, predictably more housewives and the unemployed attend daytime classes.

STUDENT MEASURES: TESTS

Three different measures were used to assess students' learning: a set of Literacy Items, an Aural Decoding Test and an Oral Proficiency Test. In addition, student scores for two measures regularly given by the Center, the John and Morano tests, were entered into the data base. The development and administration of each measure is discussed separately below.

Literacy Items

The set of 50 Literacy Items measures listening and reading comprehension. The set was adapted from 170 items originally developed under the Right-to-Read

TABLE 2.2

Night School Student Background Information --
Matched Sample: Descriptive Statistics

	Frequency	Percent	Frequency By Level 1	2	3
Number	41		21	14	6
Sex: M	18	44	8	9	1
F	23	56	13	5	5
Diploma: Yes	7	17	3	2	2
No	34	83	18	12	4
Age: Mean	42		44	42	37
Range	17-64		17-64	22-56	30-45
Education: Mean Years	10		7	11	12
Range	3-18		3-18	6-16	8-16
Time in U. S.					
Mean Years	5.9		5.5	5.3	8.8
Range	1-15		1-15	1-13	4-13
Former English					
Mean Years	1.17		1.00	1.08	2.0
Range	0-8		0-7	0-8	0-5
English in U. S.					
Mean Years	1.67		1.17	2.41	1.75
Range	0-5		0-3	0-5	0-4
Former Country					
Argentina	2	5	0	1	1
Columbia	3	7	0	2	1
Cuba	25	61	15	7	3
Ecuador	4	10	3	1	0
Guatemala	2	5	0	2	0
Other	5	12	3	1	1

Occupation*	Total	1	2	3
1. Foreman	0/0	0/0	0/0	0/0
2. Craftsman	4/7	3/4	1/2	0/1
3. Semi-skilled Worker	4/24	1/15	3/8	0/1
4. Laborer	1/2	0/1	0/1	1/0
5. Household Worker	0/0	0/0	0/0	0/0
6. Personal Service	3/2	2/1	0/0	1/1
7. Fireman/Policeman	0/0	0/0	0/0	0/0
8. Professional	6/0	1/0	3/0	2/0
9. Technician	1/0	1/0	0/0	0/0
10. Farmer	0/0	0/0	0/0	0/0
11. Farm Worker	0/0	0/0	0/0	0/0
12. Business Owner	0/0	0/0	0/0	0/0
13. Manager/Official	1/1	0/0	1/1	0/0
14. Salesperson	1/0	1/0	0/0	0/0
16. Housewife	7/1	6/0	1/1	0/0
17. Unemployed	1/0	1/0	0/0	0/0
18. Student	3/0	2/0	0/0	1/0

*Previous country/USA

See Appendix A for category definitions.

project.[2] All 170 items were ranked in order of difficulty and then reviewed for appropriateness for this population. Items deemed appropriate were then pilot tested with an ESL student with little or no English proficiency and an advanced student with proficiency. The test was then revised to include more items at both extremes, and at the same time shortened to a total of 50 items. Parallel instructions in both Spanish and English were written. Instructions on how to indicate the answers were given in either language (depending on student level of proficiency) to insure that students understood how to do the task. The question for each item was read to all students in English; they then read the item and indicated the answer to the question by circling the appropriate word or sentence. Testing was stopped when the student answered five consecutive items incorrectly. A copy of the Literacy Items can be found in Appendix A.

Aural Decoding Test

This test measures decoding skills. Part I is concerned with recognition of syllables and root words while Part II tests for phoneme-grapheme correspondence. The test was originally developed for the Beginning Teacher Evaluation Study.[3] Tests at two difficulty levels were pilot tested with the above mentioned ESL students and the results indicated that the more difficult test should be used. Instructions were modified such that the item stem was read to the student, rather than the student reading it to himself. Again parallel instructions were prepared in both English and Spanish; however, test items were read only in English. A copy of the Decoding Test can be found in Appendix A.

[2] Murphy, R. T. Adult Functional Reading Study. Educational Testing Service. Princeton, New Jersey, 1973.

[3] Conducted for the California Commission for Teacher Preparation and Licensing by ETS. Frederick J. McDonald, Project Director.

Administration of Literacy and Aural Decoding Measures

The tester-observers were trained how to administer the Literacy Items and the Aural Decoding Test and practiced giving them to several people at the different levels who were drawn from the Learning Center's waiting list.

The tests were administered to small groups of students in an unused classroom by the tester-observers. The bilingual tester-observer tested all the beginning students so that the directions could be given in Spanish and students would therefore not be penalized for lack of comprehension of spoken English as far as knowing how to indicate their answer.

All daytime students were pretested at the Center between November 5th and 25th, 1975 and posttested between April 5th and May 18, 1976 on these two measures. Night-school students were pretested on both measures between April 5th and 29th, 1976 and posttested on the Literacy Items between June 1st and 11th, 1976.

Oral Proficiency Test

It was originally planned to give an oral proficiency test as a pretest. Once the project began in October we searched the ETS Test Collection for available tests. We found none which led us to believe that such tests as have been reported to exist are available only in "fugitive" documents. We next discussed with P. Woodford of ETS the systems he had used in evaluating the language-speaking qualifications of Peace Corps Volunteers. We found that these systems involved an interview between a volunteer and an expert in the language. This expert followed an unspecified structure from easy language responses to the most difficult, but adapted the interview as he or she proceeded. The evaluation of the volunteer was solely the product of the expert judgment of the interviewer. Such a system while satisfactory for the purposes of the

Peace Corps was not satisfactory for the purposes of this study. For this study it is necessary to have an instrument that can detect a range of differences in speaking ability which in turn means that the instrument must be similarly structured for every person to be interviewed.

We next reviewed the work of R. Politzer of Stanford, an expert in bilingual education and the teaching of English as a foreign language. Politzer had developed an oral proficiency test, but it had been used with elementary school children and its content was obviously inappropriate for use with adults.

We also reviewed two tests developed by specialists in teaching English as a second langauge. Psychometric data were not available for these tests, such as their reliability. But a more important consideration led to the decision that these tests would not be appropriate. Their content did not match the objectives of the Learning Center precisely enough.

It was apparent at this point that it would be necessary to build a test specifically for this project. Two problems had to be solved. First, we needed to obtain a complete and concrete description of the objectives of each of the teachers in the Center. Second, we needed to see what similarities there were among these objectives, and how the teachers differed in terms of the objectives they were trying to achieve.

A preliminary set of items was developed by Kathleen Durnin, Meredith Stone and Patricia Elias based on lists of objectives supplied by the Learning Center teachers. Each teacher provided us with a list of 20 objectives. After a review of these statements of objectives it was obvious that we would need to build criterion-referenced scales related to the objectives and find a way of creating a test from which the performance of students could be tracked with respect to the objectives.

The teachers' lists were combined and overlapping objectives eliminated producing a list of 132 objectives across seven teachers. Each teacher then rated the 132 objectives on the following: (1) whether or not it was something they taught; (2) how important they felt it was for proficiency; (3) how much emphasis they gave it in their class; and (4) how difficult they thought it was to learn. The Proficiency Test would be constructed using those objectives on which there was high agreement among teachers teaching at the same level.

The work described above was carried on from October to December. By December it was apparent that the oral proficiency test could not be built to the specifications agreed upon in time to use it as a pretest. We decided, therefore, to use the Oral Proficiency Test as a posttest and to include analyses which would relate other indices of initial proficiency (John and Morano Tests, student background information) to the Oral Proficiency scores obtained later.

During February and March the project director and associate project director developed several series of items based on representative examples from the teachers' objectives for each level. These items were in a structured conversational format and arranged in order of difficulty by objective (based on semantic and syntactic complexity). In addition, questions concerning two magazine "action" pictures and translation of three Spanish cartoons were used to elicit English responses. The first two components of the test (structured conversation and action pictures) were designed to elicit increasingly complex forms of the verb phrases in terms of tense and prepositional and adverbial qualifiers. The third component (cartoons) was designed to assess the students' ability to go from idiomatic Spanish to idiomatic English. Items were reviewed by the coordinating teacher and then pilot tested by the coordinating teacher and associate project director with students on the waiting list. Items were modified or deleted until a 60 item test requiring approximatley 20 minutes to administer emerged.

The purpose of the Oral Proficiency Test was twofold: to assess both comprehension and the ability to generate appropriate English sentences. Therefore, the test was scored for comprehension, use of appropriate structure, correctness and number of prompts for each of the items. Again, instructions were given in either Spanish or English, although test items were given only in English. A copy of the Oral Proficiency Test and its score sheet appear in Appendix A.

Administration of the Oral Proficiency Measure

The coordinating teacher and the associate project director trained the tester-observers to administer the test. Students from the Center who were not part of the sample were subjects for practice test sessions. Each test was administered individually in a separate room and tape recorded to allow for review or verification of scoring. Daytime students were tested between May 26th and June 11th, 1976. As additional testers were needed to complete the night-school testing, two other teachers were also trained and practiced giving the test. Night-school students were tested between June 1st and June 10th, 1976. Again beginning students were assigned to bilingual testers so that directions could be given in Spanish.

John Test

This oral proficiency test was developed by Linda Kunz at Hunter College. It consists of eight pictures about which the student is asked 22 questions.

In addition to the comprehension score the student is also rated on fluency, structure, pronunciation and vocabulary by the tester. A copy of the test and its score sheet are in Appendix A.

This test is regularly given to each student when applying to the Center and again at the end of each year. Student scores are then used for placement. Thus scores from this test may have been obtained from day-students anywhere between April, 1975 and September, 1975. However, the night-school students were all tested (or retested) between March 18th and April 12th, 1976. Despite this variation in testing dates and the fact that many, if not most, of the students had been given the test at least once before, scores for the John Test were included in the analyses because they were found to be highly related to other indices of initial status.

Morano Test

This test is a paper and pencil test of recognition of correct use of English grammar. It has 50 items each consisting of three sentences expressing the same idea, only one of which is grammatically correct. The student is instructed to read the items and indicate the correct sentence. A copy of the test is in Appendix A.

This test is also regularly given by the Center to its applicants. Although the scores are considered during placement, they are not given anywhere near the same weight as the John Test scores. The day students' scores were obtained between April, 1975 and September, 1975, while all night-students were given the test between March 18th and April 12th, 1976. The Morano Test scores were included in the analyses because they were related to other indices of initial status and allowed us to control for differences in proficiency prior to the period of this study.

Descriptive Statistics

Table 2.3 presents information concerning each of the aforementioned tests. The group (day or night-school sample), number of students, and time column-headings are self-explanatory. The mean ($\bar{X}$) is the average score of the students taking the test. The range gives the lowest and highest scores obtained by the group of students. The standard deviation (SD) is a measure of the variation in scores within the group of students. The reliability coefficient is a measure of the internal consistency of an individual's responses to all the items on the test. You will note that the reliabilities are consistently high, with the one exception of Aural Decoding II which is in the mid 70's.

Due to the possibility of selective attrition in the day-school sample, t-tests were computed to detect significant differences in pretest scores between those students who remained in the sample (N = 81) and those who dropped out (N = 34-67 depending on the test). No significant differences were found. There was a slight trend for those who initially scored higher on the John and Morano tests to drop out; however this trend was not found for the Literacy or Aural Decoding Measures.

Table 2.4 presents a comparison of scores for each test, pre- and post, between all students taking the test (Total) and the Matched samples for both day and night-school students. An examination of the mean, range and standard deviation figures listed indicates that in the cases where there are differences, these differences never amount to more than one or two points. This result confirms that our day and night-school matched samples are representative of their respective larger groups. Thus all further discussion of analyses will concern only data for the matched samples.

TABLE 2.3

STUDENT TESTS: DESCRIPTIVE STATISTICS

Instrument	Who	When	N	$\bar{X}$	Range	SD	Reliability
Literacy	Day	November (Pre)	148	29.82	3-49	12.52	.96
	Day	April (Post)	119	37.91	11-49	7.22	.89
	Night	March (Pre)	45	34.73	4-49	9.70	.93
	Night	May (Post)	43	36.81	11-48	9.98	.94
Aural Decoding I	Day	November (Pre)	148	26.72	6-38	7.40	.88
	Day	April (Post)	120	28.42	13-37	5.33	.81
	Night	March (Pre)	45	29.47	14-37	6.90	.88
II	Day	November (Pre)	148	47.84	31-57	5.28	.74
	Day	April (Post)	120	52.45	36-59	4.63	.77
	Night	March (Pre)	45	49.33	47-59	4.86	.73
Total	Day	November (Pre)	148	74.57	49-93	11.03	.89
	Day	April (Post)	120	80.88	56-95	7.90	.83
	Night	March (Pre)	45	78.80	60-96	9.47	.87
Proficiency Comprehension	Day	June (Post)	113	30.48	2-58	14.66	.96
	Night	June (Post)	43	27.65	3-53	16.21	.97
Correctness	Day	June (Post)	113	11.83	0-45	8.84	.92
	Night	June (Post)	43	10.49	0-29	8.65	.92
Structure	Day	June (Post)	113	14.53	0-41	10.16	.93
	Night	June (Post)	43	12.09	0-33	10.01	.94
Prompts	Day	June (Post)	113	9.31	0-26	4.86	.72
	Night	June (Post)	43	7.70	0-19	4.47	.71
John	Day	November (Pre)	115	37.27	0-70	20.54	.87
	Night	March (Pre)	46	30.54	0-65	21.07	.86
Morano	Day	November (Pre)	118	26.82	3-50	11.47	.93
	Night	March (Pre)	46	26.30	7-46	10.46	.92

TABLE 2.4

Comparison of Test Statistics for Total (TOT) and Matched (MAT) Samples for Day (D) and Night (N) Schools

Instrument	Who	Time	N	$\bar{X}$	Range	SD
Literacy	D-TOT	Pre	148	30	3-49	13
	D-MAT	Pre	81	30	3-48	12
	D-TOT	Post	119	38	11-49	7
	D-MAT	Post	81	37	11-48	8
	N-TOT	Pre	45	35	4-49	10
	N-MAT	Pre	41	34	4-48	10
	N-TOT	Post	43	37	11-48	10
	N-MAT	Post	41	37	11-48	10
Aural Decoding I	D-TOT	Pre	148	27	6-38	7
	D-MAT	Pre	81	26	6-38	7
	D-TOT	Post	120	28	13-37	5
	D-MAT	Post	81	28	13-37	6
	N-TOT	Pre	45	29	14-37	7
	N-MAT	Pre	41	29	14-37	7
Aural Decoding II	D-TOT	Pre	148	48	31-57	5
	D-MAT	Pre	81	48	31-56	6
	D-TOT	Post	120	52	36-59	5
	D-MAT	Post	81	52	36-58	5
	N-TOT	Pre	45	49	37-59	5
	N-MAT	Pre	41	49	37-59	5
John	D-TOT	Pre	115	37	0-70	21
	D-MAT	Pre	81	39	0-69	20
	N-TOT	Pre	46	31	0-65	21
	N-MAT	Pre	41	30	0-65	21
Morano	D-TOT	Pre	118	27	3-50	11
	D-MAT	Pre	81	25	0-43	10
	N-TOT	Pre	46	27	7-46	10
	N-MAT	Pre	41	25	7-46	11
Proficiency Correctness	D-TOT	Post	113	12	0-45	9
	D-MAT	Post	81	13	0-33	9
	N-TOT	Post	43	10	0-29	9
	N-MAT	Post	41	10	0-27	8

RELATION OF STUDENT BACKGROUND INFORMATION TO TEST SCORES

The intercorrelations of student test scores, both pre- and post, with each other and with coded student background variables (see Appendix A) are presented in Table 2.5 for the day school and Table 2.6 for the night school. For the day-school sample with an N = 81 a correlation above .22 is significant at the .05 level and a correlation above .28 is significant at the .01 level. Examination of the test intercorrelations in Table 2.5 indicates that all of the tests are significantly correlated with each other with the exception of the Aural Decoding II posttest. The subscores of the Proficiency Test, correctness and comprehension, were highly correlated (r = .86) as were the correctness and use of structure scores (r = .97).

The relative magnitudes of the correlations suggest that while the John and Morano tests tap quite similar skills (r = .78), the Literacy and Decoding Tests appear to measure some different competencies. The Proficiency Test correctness score falls in between with correlations of .68 and .63 with the John and Morano Tests, correlations of .51 and .60 with Aural Decoding I, and correlations of .47 and .59 with the Literacy Test. The Proficiency Test comprehension subscore, however, has a high correlation with the John Test (r = .85).

A number of background variables were significantly related to test scores. While no relation was found between sex and test scores, a negative relation was found with age: younger students do better on the Decoding, Literacy and Proficiency Tests. Time spent in the U. S. was not related, while having a diploma was related only to the fall Literacy Test score. A higher status code (see Appendix A) for job in former country was related to all test scores except Decoding II, while a higher code for job in U. S. was related only to Decoding I and Literacy in the spring.

TABLE 2.5

Intercorrelations of Fall and Spring Tests and
Student Background Information
Day School -- Matched Sample (N = 81)

	D_1-F	D_2-F	L-F	J-F	M-F	D_1-S	D_2-S	L-S	P-CT	P-CP	P-ST
Decoding_1 - F	–										
Decoding_2 - F	.50	–									
Literacy - F	.65	.41	–								
John - F	.56	.29	.77	--							
Morano - F	.55	.25	.59	.78	–						
Decoding_1 - S	.73	.40	.48	.42	.45	–					
Decoding_2 - S	.33	.27	.22	.10	.15	.41	–				
Literacy - S	.47	.31	.71	.64	.51	.47	.16	–			
Proficiency											
Correct	.60	.26	.59	.68	.63	.51	.22	.47	–		
Comprehension	.70	.39	.75	.85	.73	.63	.21	.61	.86	–	
Structure	.63	.30	.63	.74	.68	.54	.23	.51	.97	.90	–
Sex	.07	.10	.17	.17	.02	.03	-.06	-.02	.12	.19	.13
Age	-.32	-.15	-.30	-.13	.05	-.27	-.06	-.28	-.18	-.23	-.21
Time in U. S.	-.15	.05	.18	.13	-.07	-.13	-.08	.00	.05	.06	.04
Diploma	.21	.08	.22	.11	.09	.22	.17	.09	.12	.18	.16
Former Job	.25	.17	.39	.33	.33	.29	.18	.22	.24	.32	.27
Job in U. S.	.13	.08	.21	.20	.19	.26	.16	.24	.13	.20	.17
C-Origin	.05	-.12	.02	-.07	-.15	.00	.09	.04	-.01	-.04	.01
Years Education	.21	.12	.33	.18	.16	.23	.15	.23	.22	.23	.25
Former Eng.	.16	.19	.25	.26	.17	.08	.13	.12	.12	.14	.16
Eng. in U. S.	.28	.13	.41	.51	.43	.25	.03	.31	.17	.37	.23

$p \leq .05$ for $r \geq .22$
$p \leq .01$ for $r \geq .28$

Country of origin was not related to test scores. Years of education was related only to Literacy in the fall (although Decoding I - Fall approaches significance), but to all but Decoding II in the spring. The study of English in former country was related only to John and Morano test scores, while study of English in the U. S. was related to Decoding I, Literacy and Proficiency scores as well.

For the night-school sample with an $N = 41$, a correlation of .30 or above is significant at the .05 level while a correlation of .39 is significant at the .01 level.

Table 2.6 indicates that once again all of the tests are significantly correlated with the exception of Aural Decoding II. The three subscores of the Proficiency Test are again highly related (r's = .83, .87 and .98). As with the day-school, the John and Morano Tests were highly correlated ($r = .77$). However, correlations between the Morano and the Literacy Test scores ($r = .71$ and .77) were much higher for this sample than for the day-school sample. Correlations of all these tests with the Proficiency correctness score were also high ($r = .73$ to .78).

Only five of the ten background factors were related to any of the test scores. Age was again negatively related to Decoding I and Literacy. Country of origin was significantly related to the Proficiency comprehension subscore but not to either the correctness or use of structure subscores.

Years of education was related to all test scores but Aural Decoding II. The study of English in former country was related to Aural Decoding II and Morano Tests for the pretests, but not to any of the posttests, while study of English in the U. S. was related only to the Morano pretest and to two Proficiency subscores, comprehension and use of structure.

TABLE 2. 6

Intercorrelations of Winter and Spring Tests and
Student Background Information
Night School —Matched Sample (N = 41)

	D_1-W	D_2-W	L-W	J-W	M-W	L-S	P-CT	P-CP	P-ST
$Decoding_1$ - W									
$Decoding_2$ - W	.25	-							
Literacy - W	.74	.14	-						
John - W	.64	.10	.71	-					
Morano - W	.67	.36	.71	.76	-				
Literacy - S	.74	.26	.80	.72	.77	-			
Proficiency - S									
Correct	.71	.21	.73	.78	.78	.73	-		
Comprehension	.73	.29	.73	.73	.79	.79	.83	-	
Structure	.72	.24	.73	.77	.81	.74	.98	.87	-
Sex	-.11	.05	-.06	.08	-.09	-.02	-.14	-.09	-.15
Age	-.42	-.23	-.31	-.18	-.17	-.41	-.21	-.20	-.17
Time in U. S.	.03	-.11	.17	.27	.15	.15	.22	.22	.22
Diploma	.29	.20	.12	.26	.22	.23	.18	.24	.14
Former Job	.13	.13	.00	.04	.07	.08	-.16	.07	-.08
Job in U. S.	.13	.12	.16	.19	.27	.26	.25	.23	.23
C-Origin	.22	-.05	.14	-.00	.04	.26	.27	.31	.24
Years Education	.56	.29	.45	.51	.48	.65	.40	.61	.45
Former Eng.	.22	.35	.12	.27	.30	.26	.27	.26	.25
Eng. in U. S.	.29	.09	.25	.16	.32	.21	.25	.36	.32

$p \leq .05$ for $r \geq .30$

$p \leq .01$ for $r \geq .39$

PROCESSING OF STUDENT DATA

Following each administration of the tests, the test booklets or coded answer sheets were sent to ETS for processing. These booklets or answer sheets were processed and data sets were made for use as input to the computer. These data were sorted by the student identification number and were edited for errors. All errors were corrected and the corrected data were re-edited until no errors were found.

Scores for each test were derived by comparing the item responses with the answer key for the test and counting the number of correct responses. Item analysis, preliminary item and score correlations, and preliminary summary statistics were then generated.

After the final test administration for each group of students (day and night), the pre- and post data were matched. (Day and night data were kept on separate data sets and were analyzed separately). More analyses including score correlations were made.

Student background data and class attendance data were collected and sent to ETS. These data were processed and data sets were made which were sorted and edited and used as input to various preliminary analyses.

The student background and attendance data were then merged with the matched spring and fall data for more analyses. These data were eventually merged with the teacher observation data for the final analyses. Please see Appendix D for details of the data processing.

TEACHER BACKGROUND INFORMATION

Both day and night school teachers filled out a questionnaire asking for the following information: age; sex; undergraduate institution; undergraduate major; graduate work institution; graduate work major(s); number of credits completed; degree(s) completed; number of years teaching; grade level of teaching experience; number of years teaching ESL part-time (night-school adults); number of years teaching ESL full-time to children, to teens, to adults; number of ESL seminars; number of ESL workshops; usefulness of undergraduate experiences; usefulness of graduate experience; usefulness of ESL seminars and workshops; percent of teaching which is audiolingual, silent-way, and other; and a statement describing their teaching methods and philosophy. A copy of the Teacher Questionnaire appears in Appendix B.

Table 2.7 shows the descriptive statistics for these information items for each group of teachers. The two groups differ on a number of dimensions. The majority of the day-school teachers are women, while only half of the night-school teachers are. All six day-school teachers have completed some graduate work, three have completed a Master's degree, and one is working on a second degree; while only three of the night-school teachers have done graduate work—all three completed a Master's degree and two of them have completed a second Master's degree. Night-school teachers on the average have had more experience teaching, but this apparent difference is due primarily to one teacher who has taught for 26 years. Most of the day-school teachers' experience with ESL has been teaching adults, while most of the night-school teachers also teach ESL to children or teens during the day as well as to adults in the evening. Day-school teachers have attended a few

TABLE 2.7

Teacher Background Information

	DAY SCHOOL (N = 6)*	NIGHT SCHOOL (N = 6)*
Age	$\bar{X}$: 35 Range: 24-45	$\bar{X}$: 35 Range: 23-55
Sex	5 Females; 1 Male	3 Females; 3 Males
Undergraduate Institution	Jersey City State: 2 Douglas: 1 Ladycliff: 1 Montclair State: 1 Seton Hall: 1	Jersey City State: 3 N.E. Missouri: 1 Saint Peter's: 1 St. Elizabeth's: 1
Undergraduate Major	Elementary Ed.: 2 Spanish/Ed.: 2 Philosophy: 1 Italian: 1	Elementary Ed.: 4 History: 2
Graduate Work Institution (1st. MA)	Jersey City State: 2 Fairleigh Dickinson: 1 Montclair State: 1 Rutgers: 1 Seton Hall: 1	Jersey City State: 3
Graduate Work Major	Elementary Ed.: 1 ESL: 1 ESL and Ed.: 1 ESL and Reading: 2 Guidance and Personnel: 1	Elementary Ed.: 1 Reading: 2
Number of Credits Completed	N = 6; $\bar{X}$: 30, Range: 9-47	N = 3, $\bar{X}$: 32, Range: 30-34
1st Master's Degree Completed	N = 3	N = 3
Graduate Work Institution (2nd MA)	Rutgers: 1	Jersey City State: 1 Montclair State: 1 Fairleigh Dickinson: 1
Graduate Work Major	Italian Lang. & Lit.: 1	ESL: 2 Elementary Ed.: 1
Number of Credits Completed	N = 1, $\bar{X}$: 27	N = 3, $\bar{X}$: 26, Range 15-34
2nd Master's Degree Completed	N = 0	N = 2

** The N for each category = 6 unless otherwise stated.

TABLE 2.7 (Continued)

	DAY SCHOOL*	NIGHT SCHOOL*
Number of Years Teaching	$\bar{X}$: 5 Range: 1-18	$\bar{X}$: 10 Range: 1-26
Teaching Experience Level	Elementary: 3 High School: 1 Adults: 2	Elementary: 4 Junior High: 2
Years ESL at Night, Adults	N = 2, $\bar{X}$: 8, Range: 3-12	N = 6, $\bar{X}$: 4, Range: 1-6
Years ESL Children Full Time	N = 1, $\bar{X}$: 4	N = 3, $\bar{X}$: 4, Range: 1-7
Years ESL Teens Full Time	N = 0	N = 2, $\bar{X}$: 4, Range: 3-4
Years ESL Adults Full Time	N = 6, $\bar{X}$: 2, Range 1-4	N = 0
Number of ESL Seminars	$\bar{X}$: 4, Range: 0-8	$\bar{X}$: 1, Range: 0-5
Number of ESL Workshops	$\bar{X}$: 17, Range: 4-30	$\bar{X}$: 13, Range: 2-25
Usefulness Undergraduate Experience**	$\bar{X}$: 4, Range: 2-5	$\bar{X}$: 3, Range: 1-4
Usefulness of Graduate Experience**	N = 6, $\bar{X}$: 3, Range: 1-4	N = 3, $\bar{X}$: 2, Range: 1-3
Usefulness of ESL Seminars and Workshops**	$\bar{X}$: 1, Range: 1-2	$\bar{X}$: 1, Range: 1
Percent Audio-Lingual Method	N = 2, $\bar{X}$: 62%, Range: 60-65%	N = 6, $\bar{X}$: 60%, Range: 10-90%
Percent Silent-Way Method	$\bar{X}$: 66%, Range: 30-99%	$\bar{X}$: 35%, Range: 10-60%
Percent Other Methods	N = 5, $\bar{X}$: 14%, Range: 1-30%	N = 1, $\bar{X}$: 30%

*The N for each category = 6 unless otherwise stated.

**Very useful = 1, not at all useful = 5.

TABLE 2.7 (Continued)

	DAY SCHOOL*	NIGHT SCHOOL*
Concepts Underlying Method & Philosophy:		
Audiolingual	N = 2	N = 4
Cognitive	N = 1	N = 0
Conversational	N = 0	N = 1
Counseling-learning	N = 1	N = 0
Ecletic	N = 3	N = 0
Flexible-adjust to class	N = 3	N = 2
Group work	N = 0	N = 1
Individualize	N = 1	N = 0
Listening and speaking	N = 1	N = 1
Need for English	N = 0	N = 1
Peer teaching	N = 0	N = 2
Reading and Writing	N = 1	N = 1
Silent way	N = 4	N = 4
Situational	N = 3	N = 3
Student dominated	N = 4	N = 0
Student responsibility	N = 2	N = 1
Variety	N = 2	N = 1

*The N for each category = 6 unless otherwise stated.

more ESL seminars and workshops. Both groups consistently rate their undergraduate experience as not very useful, their graduate experience as more useful, and the ESL seminars and workshops as very useful.

While only two day-school teachers say they use the audiolingual method of teaching, they estimate that some 60-65 percent of their teaching falls in this category; all of the night-school teachers state that they use the audiolingual method, but their estimates range from 10-90 percent. All of the teachers in both groups state that they use silent-way methods, but the day-school teachers' estimates are from 30-90 percent while the night-school teachers' estimates fall between 10 and 60 percent. With respect to the 17 concepts given as underlying their teaching methodology and philosophy, only two are substantially different for the two groups: day-school teachers more frequently state their belief that teaching style should be ecletic and that students should dominate classroom interaction.

A discussion of how the Teacher Questionnaire information relates to other analyses appears in Chapter Three.

CHAPTER THREE

THE OBSERVATION SYSTEM

DESCRIPTION OF THE OBSERVATION SYSTEM

Development of the Observation System

Teachers' methods, materials, and interaction with students were coded with an in-class observation system. This system was based on what the ESL teachers and students at the Center actually do. To develop the system, the project director, associate project director and research assistant observed in several classrooms and then discussed their observations with the coordinating teacher.

The associate project director and research assistant then took on primary responsibility for developing the system and spent much of the first three weeks of the project observing in Center classrooms and talking with the teachers. The last week in October each teacher was videotaped teaching a 30 minute class.

On the basis of the in-class observations, the videotapes and discussions with the project director and coordinating teacher, we developed a preliminary form of the observation system. This first system was tried out in the Center classrooms by the research assistant. After discussions among the project director, associate project director, research assistant and consultants, repeated modifications were made and a lexicon of definitions prepared. During this time actual observations were discussed to clarify definitions of categories; categories were expanded or eliminated; videotaped and in-class examples were considered in developing the final categories. An important consideration in developing the system was to make sure that the categories would describe the differences between the two major teaching strategies used in the Center. The final draft version was used in preliminary observations of night-school classes as well as day-school classes to insure its appropriateness to both settings.

Description of the Observation System

The result of the development work described above was a categorical observation system which allowed for sequential coding of classroom behavior. A reduced copy of the Observation Coding Sheet is shown in Figure 3.1. A complete lexicon and a sample observation are in Appendix C.

Three superordinate categories--context (instructional design), materials, and strategy (method of instruction)--describe the classroom setting within which the teacher and student behaviors are recorded. Each of these categories is subdivided; e.g., the context can be drill, writing, explanation, dictation, etc., and each subdivision has a numerical code. These categories are coded initially and recorded again only if they change during the observation period.

The first group of teacher behavior categories--questions, serial redirects, direction, models, writes on board, explanation, and other--are discrete instructional behaviors which usually initiate a teacher-student(s) interaction. Where these behaviors are carried out nonverbally, they are coded with an "N" rather than a check mark. The next column--class, group, individual--signifies to whom this behavior is directed.

The first group of student behavior categories--answers, free response, practice, writes on board, reads, chooses not to respond, asks question, (participates in) conversation, student-to-student feedback, and other--are those behaviors which either follow the teacher's initial behavior or initiate an interaction on the part of the student. Conversation and student feedback are coded with an "S" if they occur in Spanish rather than English.

The next three categories--positive, corrective, negative--describe the possible types of teacher feedback. Here again, since nonverbal feedback is an important component of "silent way" instruction, an "N" is used to indicate a nonverbal response.

The second group of teacher behavior categories--models, prompts, asks to repeat, repeats, explanation, writes on board, direction, question, and other--designate response behaviors on the part of the teacher. Again, they are coded to indicate whether they are given verbally or nonverbally.

The second group of student behavior categories--student models, student prompts (these two usually follow a direction from the teacher), answers, free response, practice, writes on board, reads, chooses not to respond, asks question, (participates in) conversation, student-to-student feedback and other--describes those student behaviors given in response to the teacher's response to the student's initial behavior or response. Again, an "S" is used to code those responses which were given in Spanish.

The comment column allows the observer to indicate what the "other" behavior coded on that line is or to note some unusual classroom occurrence.

Use of the Observation System

The observation system was designed for in-class use. When the observer first entered the room, he or she took a few minutes to fill out the top of the observation sheet, indicating the date, teacher, time, number of students, observer, class set-up (diagram) and language structure being taught. During this orientation time it was possible to code the three superordinate categories and then quickly proceed to accurate coding of the classroom interaction behavior.

By way of illustration, a possible classroom sequence and its codes follows--the sequence has been coded on an observation form in Figure 3.1. Reading from left to right, the context is drill (1), they are using Cuisenaire rods as materials (1), and the instructional strategy is question and answer (1).

FIGURE 3.1

Observation Code Sheet, with an Illustration of a Possible Classroom Behavior Sequence

DATE —— TEACHER —— TIME —— NO. OF STUDENTS 8 OBSERVER ——

CLASS SET UP ⊤ STRUCTURE simple picture

	CONTEXT 1	MATERIAL 2	STRATEGY 3	TEACHER BEHAVIOR: QUESTION 4	SERIAL REDIRECT 5	DIRECTION 6	MODELS 7	WRITES ON BOARD 8	EXPLANATION 9	10	CLASS/GROUP/IND. 11	STUDENT BEHAVIOR: ANSWERS 12	FREE RESPONSE 13	PRACTICE 14	WRITES ON BOARD 15	READS 16	CHOOSES NOT TO RESP. 17	ASKS QUESTION 18	CONVERSATION 19	S/S FEEDBACK 20	21	POSITIVE 22	CORRECTIVE 23	NEGATIVE 24	25
1	1	1	1	✓							✓	✓										✓			
2																									
3																									
4							✓				1			✓											
5																									

	TEACHER BEHAVIOR: MODELS 26	PROMPTS 27	ASKS TO REPEAT 28	REPEATS 29	EXPLANATION 30	WRITES ON BOARD 31	DIRECTION 32	QUESTION 33	ALTERNATE RESP. 34	STUDENT MODELS 35	STUDENT PROMPTS 36	STUDENT BEHAVIOR: ANSWERS 37	FREE RESPONSE 38	PRACTICE 39	WRITES ON BOARD 40	READS 41	CHOOSES NOT TO RESP. 42	ASKS QUESTIONS 43	CONVERSATION 44	S/S FEEDBACK 45	46	47	COMMENT 48
1			N									✓											
2		N										✓											
3	✓					✓								✓									
4	✓																						
5																							

- 54 -

The original observation sheet was 8 1/2" by 14"

First teacher behavior (TB_1): "How many rods will you take?" [Code line 1, columns 4 and 11: TB_1 - question $\longrightarrow$ individual].

First student behavior (SB_1): "I will take three rods." [Code line 1, column 12: SB_1 - answer].

Second teacher behavior (TB_2): "Good." [Code line 1, column 22: TB_2 - positive feedback] and signals that the student should repeat what he has said [Code line 1, column 28: TB_2 - asks to repeat - N (nonverbal)].

Second student behavior (SB_2): "I will take three rods." [Code line 1, column 37: SB_2 - answers].

Teacher: While the student is repeating the sentence, the teacher holds up a finger for each word. When he is finished she indicates by pushing the first two fingers together that he should use the contraction. Since the interaction is still with the same student, the observer drops down a line on the right-hand side of the sheet. [Code line 2, column 27: TB_2 - nonverbal prompt].

Student: "I'...ill take three rods." [Code line 2, column 37: SB_2 - answers].

Teacher: "<u>I'll</u> take three rods." [Code line 3, column 26: TB_2 - models] and writes contraction on board [Code line 3, column 31: TB_2 - writes on board].

Student: "I'll take three rods." [Code line 3, column 39: SB_2 - practice].

Teacher: "O.K., everyone. <u>I'll</u> take three rods." [Code line 4, columns 7 and 11: TB_1 - models $\longrightarrow$ class].

Students: "I'll take three rods." [Code line 4, column 14: SB_1 - practice].

Observer Training

Using the videotapes of the Center teachers, the associate project director and research assistant trained the coordinating teacher and both tester-observers on the system. In-class training was accomplished by placing two observers in the same classroom and comparing their interaction codes for agreements and disagreements. Those categories where the highest disagreement occurred were in several cases further modified and in others better clarified to improve inter-observer agreement. Once these final modifications were completed, reliability studies were begun.

Observation Schedule

Day School

The coordinating teacher drew up a schedule for observations which allowed for a 20-minute observation of each class on four different days of the week. For those classes meeting only two or three times a week, observations were scheduled for each meeting. While the coordinating teacher was originally scheduled to observe in each classroom at least once a week, this became impossible due to other job commitments. She did, however, occasionally substitute for one of the tester-observers on a random basis. (See the latter portion of this chapter on the reliability of the observations.) Observation of the day-school classes began January 19th and continued through April 1, 1976.

Several factors affected the actual number of observations made; e.g., school holidays and closing as well as teacher and observer illness. The number of observations per class and observer are presented in Table 3.1.

TABLE 3.1

Number of Classroom Observations: By Class and Observer

Day Classes	Observer 1	Observer 2	Observer 3	Total
0221	8	0	21	29
0233	9	0	6	15
0312	1	1	16	18
0313	15	0	1	16
0323	15	0	14	29
0412	15	0	15	30
0422	17	2	13	32
0433	16	0	15	31
0511	16	0	15	31
0522	15	1	13	29
0621	15	1	17	33
0711	16	1	14	31
0721	14	0	14	28
0731	16	0	14	30
TOTAL	188	6	188	382

Night Classes	Observer 1	Observer 2	Observer 3	Total
0901	5	3	4	12
1203	7	2	2	11
1402	5	4	5	14
1601	5	1	4	10
1701	7	3	4	14
2102	4	1	4	9
TOTAL	33	14	23	70

Note that for all but three of the classes (two of which met only two or three times-a-week), we have an average of 30 observation sessions per class, giving us a data base of some 600 minutes of observation per class.

Night School

A similar observation schedule was drawn up for the night school. Here, however, since class meetings were held for two hours, three times-a-week, one observer observed during the first half of a class and a second observer observed during the second half. Night-school observations were made by the same tester-observers and the coordinating teacher. Observations were made at two different times during the semester: between February 5th and March 17th, 1976, and again between May 3rd and May 25th, 1976.

Again, the number of observations per class and observer are listed in Table 3.1. For the six night-school classes we had an average of twelve observations per class, giving us a data base of approximately 200 minutes per class.

Processing of Observation Data

As classroom observation data were being collected, the data were being transferred and coded onto forms which were sent to ETS. These data were processed and data sets were made for use as input to the computer. An editing procedure similar to the one used for student data was used on these data.

Data for analyses were then created and this process is described in detail in the section of this chapter entitled, "Analysis of Behavioral Observations." Many analyses including summary statistics, correlations and factor analyses were completed. Factor scores were then created for each teacher and were placed on each student's record (the appropriate student and teacher were matched) for the final analyses. Please see Appendix D for the detail of the data processing.

Descriptive Statistics Pertaining to the Observational Data

Table 3.2 gives the observation item numbers, category labels, codes and means for day and night-school classes. The category means may be interpreted as the proportion of observation episodes[1] in which the event was observed in the day-school or night-school classes. However, these will add to 1.00 only when the behaviors within a group are mutually exclusive and exhaustive.

Note that the means for day and night-school classes in most instances are highly similar. The categories where this is not the case are: night-school teachers are more likely to work without materials (item 2); day-school teachers are more likely to use corrective feedback (item 12), prompts (item 15), and their actions are more likely to be nonverbal (item 23); night-school students' initial behavior is more likely to be answering or questioning (item 9), while these behaviors are more likely to be successive behavior for day-school students (item 20).

Overall, the similarity between means for the remaining categories indicates that similar amounts and the same kinds of behavior were observable in both the day and night-school classes. A detailed description of how the observation data were analyzed follows.

---

[1] An episode is defined as a sequence of behaviors between the teacher and any particular student. It may be initiated by either teacher or student and ends when the teacher addresses or responds to another student.

TABLE 3.2

Item Categories for Classroom Observation

Item	Number of Categories	Category Labels	Codes	Day School Means+	Night School Means+
1 Context	2	1 Drill	DRILL	.99	1.00
		2 Other	OTHR	.01	.00
2 Materials	7	1 No Materials	NO MATLS	.53	.65
		2 Rods	RODS	.11	.12
		3 Cartoons	CARTOONS	.02	.00
		4 Pictures	PICTURES	.05	.05
		5 Mimeo Sheets	MIMEO	.23	.17
		6 Objects	OBJECTS	.04	.00
		7 Sight Words	SIGHTWDS	.02	.00
3 Strategy/Model	5	1 Question & Answers	Q&A	.94	.99
		2 Free Response	FREE RES	.03	.01
		3 Repetition	REPEAT	.00	.00
		4 Directed Dialogue	DIRECTED	.02	.00
		5 Discussion	DISCUSSN	.01	.00
4 Initial Teacher Behavior	2*	2 Question	TB1-QUES	.32	.40
5 Initial Teacher Behavior	2**	2 Models	TB1-MODL	.10	.11
6 Initial Teacher Behavior	2**	2 Writes on Board	TB1-WOB	.05	.06
7 Initial Teacher Behavior	4	1 Serial Redirect	TB1-SR	.01	.02
		2 Direction	TB1-DIR	.18	.18
		3 Explanation	TB1-EXP	.04	.09
		4 Other	TB1-OTHR	.03	.01
8 Object of Teacher Behavior	3	1 Class	CLASS	.29	.31
		2 Group	GROUP	.01	.00
		3 Individual	INDIVDL	.70	.69
9 Initial Student Behavior	2*	2 Answers	SB1-ANSR	.35	.45
10 Initial Student Behavior	2*	2 Practices	SB1-PRAC	.11	.11
11 Initial Student Behavior	8	1 Free Response	SB1-FREE	.09	.05
		2 Writes on Board	SB1-WOB	.01	.02
		3 Reads	SB1-READ	.09	.12
		4 Chooses Not to Respond	SB1-CNOT	.00	.01
		5 Asks Question	SB1-AQ	.12	.06
		6 Conversation	SB1-CONV	.01	.03
		7 Student-Student Feedback	SB1-SFBK	.02	.04
		8 Other	SB1-OTHR	.02	.03

69

70

TABLE 3.- (continued)

Item	Number of Categories		Category Labels	Codes	Day School Means+	Night School Means+
12 Corrective Feedback	2*	2	Corrective	FB-CORR	.50	.30
13 Quality of Feeback	4	1	OK	FB-OK	.07	.02
		2	Positive	FB-POS	.05	.06
		3	Negative	FB-NEG	.01	.01
		4	Other	FB-OTHR	.00	.00
14 Successive Teacher Behavior	2*	2	Models	TB2-MODL	.22	.22
15 Successive Teacher Behavior	2*	2	Prompts	TB2-PROM	.47	.20
16 Successive Teacher Behavior	7	1	Asks Student to Repeat	TB2-ATR	.16	.07
		2	Repeats	TB2-REP	.04	.07
		3	Explanation	TB2-EXP	.07	.07
		4	Writes on Board	TB2-WOB	.06	.09
		5	Direction	TB2-DIR	.02	.02
		6	Alternate Response	TB2-ALT	.02	.00
		7	Other	TB2-OTHR	.01	.00
17 Successive Teacher Behavior	2*	2	Question	TB2-QUES	.09	.12
18 Other Student Behavior	2*	2	Student Models	SB2..SMODL	.09	.05
19 Other Student Behavior	2*	2	Student Prompts	SB2-SPROM	.05	.04
20 Successive Student Behavior	2*	2	Answers	SB2-ANSR	.75	.40
21 Successive Student Behavior	2**	2	Writes on Board	SB2-WOB	.02	.02
22 Successive Student Behavior	8	1	Free Response	SB2-FREE	.02	.01
		2	Practices	SB2-PRAC	.22	.17
		3	Reads	SB2-READ	.02	.05
		4	Chooses not to Respond	SB2-CNOT	.00	.00
		5	Asks Question	SB2-AQ	.04	.01
		6	Conversation	SB2-CONV	.00	.00
		7	Student-Student Feedback	SB2-SFBK	.04	.03
		8	Other	SB2-OTHR	.01	.00
23 Observer Comments	7	1	General Comment	C1-GENLC	.02	.02
		2	Student Response in Spanish	C2-SRSPN	.03	.04
		3	Teacher Action Non-verbal	C3-TA-NV	.10	.01
		4	Backwards Buildup Exercise	C4-BLDEX	.01	.00
		5	Teacher Reads	C5-READG	.00	.00
		6	Teacher Response in Spanish	C6-TRSPN	.00	.00
		7	Comment on Back of Summary Sheet	C7-OTHR	.07	.04

+Mean number of occurrences per episode. In most instances these may be interpreted as proportion of episodes in which the event was observed. They will only add to 1.00 when groups of behaviors are mutually exclusive and exhaustive.

*Binary items are designated appropriately as having two categories, but only the "behavior present" category is labeled and scored for analysis.

**Some of the binary items which could be repeated in rapid succession were simply counted instead of making a separate data entry for each instance.

ANALYSIS OF BEHAVIORAL OBSERVATIONS

The system of recording classroom behaviors employed in this study yielded a 23-item response record for each episode of interaction observed in the classroom. Each such episode could be initiated either by the teacher or by a student, and could continue through an extended sequence of dyadic interchanges. Provision was also made for recording participation of other students in the basic pattern of teacher/student interactions. Every interchange in each sequence of interaction (episode) was initially recorded and coded individually; these individual instances were then aggregated by summing to get one data-record per episode. An episode began at the initiation of any new interaction either by the teacher or by a student.

Methods of Organizing Dyadic Chains of Interaction

As seen in Table 3.2, some of the things a teacher could do to initiate an interaction with a particular student, group of students, or the class were: ask a question, model correct usage, or give directions (use of the Observation lexicon in Appendix C in conjunction with Table 3.2 will be helpful throughout this section). A student could initiate interaction by asking a question, for example, or could continue an episode of interaction initiated by the teacher by answering a question, following instructions (e.g., to read or to write on the board), practicing, etc. If the sequence of interaction continued, it could lead to further teacher behavior in the form of corrective feedback, modeling, prompting, questioning, etc.; which could lead, in turn, to the student's second attempt to answer correctly, to more practice, etc. Long chains of cyclical, dyadic interaction could thus be, and indeed were, coded as "successive" behaviors in each episode of classroom interaction.

For purposes of data analysis we chose to treat each distinct interaction sequence or classroom episode as the basic unit. This means that all teacher and student behaviors after the initiation of any particular interaction sequence were aggregated. We thus retained only the distinction between initiating and subsequent behaviors. By noting only the length of each chain of subsequent interactions we forewent the possibility of dealing in any detailed fashion with sequential analysis of classroom interaction episodes, but we also avoided the problem of handling chains of varying length. By maintaining the distinction between initiating behaviors and subsequent consequences, however, we left some opportunity for the most basic aspects of sequential organization among classroom behaviors to emerge in our analyses.

Had we not aggregated across successive interchanges within each dyadic classroom episode, we would have been left to deal with the original data in its basic categorical form; i.e., 23 multicategory items as listed in Table 3.2. Each of the original behavioral interchanges was recorded as a 23 item "word" with a total of 81 (2+7+5+2+2+2+4+3+2+2+8+2+4+2+2+7+2+2+2+2+2+8+7) categories. Even if we neglect the fact that more than one teacher or student behavior can occur at once (e.g., model and explanation), each interaction record could represent any one of 154,350,000 (2x7x5x7x3x10x5x10x3x10x7) possible different patterns of interaction. We could thus summarize the data from all individual instances of interaction in an eleven-way contingency table. Such a table would be mostly empty, however, since only 20,000 instances of behavioral interaction were observed during the entire 150 hours of classroom observation in this study.

Even our aggregated data must be viewed as having some of the features of multicategory data since many interaction chains were very short in length. Likewise, the categories in each of the 11 partitioned subsections of Table 3.2 are likely to be mutually exclusive and exhaustive; for instance, the occurrence of one category of teacher behavior such as modeling generally precludes the occurrence of certain other categories of teacher behavior. There were many dependencies among the categories in each subsection of Table 3.2 which had to be taken into account in our analysis of the data just as though it were pure multicategory data.

Objectives of the Analysis of Classroom Interaction

Our main interest in the analysis of the behavioral observation data is to find, if they occur, fairly stable and distinct patterns of classroom interaction. That is, we want to find out if certain teacher behaviors give rise to or are associated with certain specific student behaviors. The possibility of more-or-less stereotyped chains of interaction is thus being entertained, along with the possibility that these are characteristic of certain contexts, materials, or classroom groupings. From the point of view of analyzing categorical data we are interested in the patterns of association shown between all pairs of multicategory items (subsections) given in Table 3.2; i.e., in all two-way faces of the eleven-way contingency table alluded to earlier.

Because categories of each item tend to be mutually exclusive and exhaustive, we must only take account of associations between categories which belong to different items in any analysis of association between teacher behavior, student behavior, context, etc. Our problem is not unlike that of avoiding the contaminating influence of method variance in factor analysis of multitrait-multimethod correlation matrices, and our solution is equally applicable to that situation; i.e., simply do not fit the mono-method submatrices.

McDonald has suggested methods for the common factor analysis of multicategory data (1969) as well as methods for conducting factor analysis with residual covariance matrices of prescribed structure (1970) which can be applied to the data at hand. His work grew out of a long search for general exploratory methods for the analysis of multicategory data. Guttman (1941) initially proposed multivariate analysis of contingencies; Lazarsfeld (1950) dealt with general latent class and latent structure models; Burt (1953) showed that results equivalent to those of Guttman could be obtained simply through principal components analysis of covariances among item categories coded as binary vectors.

McDonald (1969) proposed a complex weighting scheme to bridge the gap between Burt's principal components and the more desirable common factors. The latter can be made very simple, however, by use of a flexible weighted least-squares method of factor analysis (Yates, 1971) which handles multi-category data simply by not fitting those covariances among binary-coded item categories which are biased due to mutual dependencies. Thus, instead of just fitting the one-way marginals in a contingency table (on the hypothesis of no interaction, as in the χ^2 test of independence) or the alternative of fitting both one-way marginals and certain two-way faces (as with a log-linear model), our approach compromises by fitting one-way marginals (means), while approximating, to any desired degree of tolerance (a function of dimensionalty), all of the two-way faces of an n-way contingency table. A multidimensional-multiplicative model (factor analysis) is used to estimate the elements in a joint proportion matrix among all pairs of categories from different multicategory items in order to accomplish the required fit.

Methods of Analyzing Classroom Interaction

Since the multicategory data in this study were partially aggregated (within behavioral episodes, the chosen units of analysis), they could not be used to generate a conventional joint proportion matrix. We therefore applied common factor analysis with residual covariance matrices of prescribed structure (ignoring relationships between categories belonging to the same item) to the matrix of product-moment correlations among all 68 binary-coded item categories listed in Table 3.2. Since we were now dealing with continuous data aggregated within episodes, we could make direct use of the counts of rapidly recurring events, as well as of a measure of the duration of those interactions which occupied more than one minute.

The availability of data on many individual behavioral episodes made it possible to conduct factor analyses of a large number of binary variables separately for the day-school and night-school samples, even though only six teachers were studied in each sample. Several different methods were used to determine how many common factors should be extracted from each correlation matrix. If anything, we probably erred on the side of extracting too many factors because of our desire to avoid missing any possible basis for distinguishing between teachers. Table 3.3 presents eigenvalues of the original day and night-school correlation matrices using highest off-diagonal correlations as the initial communality estimates.

TABLE 3.3

Variance of the First 12 Principal Axis Factors in a Preliminary Analysis of Correlation Matrices Derived from Observational Data

	Day School		Night School	
	Eigenvalue*	Cumulative % Trace	Eigenvalue	Cumulative % Trace
1	3.67	21	3.60	19
2	2.53	36	2.71	33
3	1.77	46	2.17	45
4	1.54	55	2.03	56
5.	1.49	64	1.88	66
6	1.33	72	1.44	73
7	1.26	79	1.30	80
8	1.09	85	1.15	86
9	.86	90	.81	91
10	.66	94	.72	94
11	.62	98	.58	98
12	.58	101	.54	100

As can be seen from Table 3.3, there is a suggestive break in the distribution of eigenvalues after eight factors for the night-school sample, but we retained nine in both samples to maintain comparability. After weighted least-squares (minimum residuals) fitting, the nine-factor root-mean-squared residual correlation between binary-coded categories from independent items had been reduced to .027 for the day-school sample and .032 for the night-school sample.

The nine minimum residuals (MINRES) factors for each sample were transformed to oblique simple structure using the direct GEOMIN hyperplane search method (Yates, 1974). The resulting primary factor pattern matrices and primary factor intercorrelation matrices are presented in Appendix E. Because the two samples used provide an opportunity to compare factors we will discuss the results from that perspective.

Factors of Classroom Interaction

One point to note about the factor patterns presented in Appendix E, Tables 1 and 3, is the fact that virtually all of the factors obtained are truly factors of classroom _interaction_; i.e., categories of behavior from distinct and independent sections of Table 3.2 load on each factor. This is as we would like it and is due to the fact that correlations among mutually exclusive and exhaustive behavior categories were ignored in the MINRES fitting phase of the analysis. One minor disappointing feature of the outcome is that one factor in each analysis represents a major contrast between two different modes of classroom interaction. This result may say something valid about the organization of classroom behavior, but it complicates matters from the point of view of discussion and interpretation.

Rather than discussing these factors in a rigid but superficial way with respect to the relative amount of variance accounted for, size of loadings, etc., we prefer to try to convey to the reader the outcome of our own attempt to infer how classroom behavior seems to be organized. After all, the point of undertaking these analyses is not to generate numbers that we are then compelled to take seriously, but to help us form a conceptual framework based upon empirical results through which we can come to understand and discuss classroom interaction. Thus the reader who is not excited by perusing tables of numbers and who trusts our judgment need not worry about either the numbers in Tables 1 through 4 of Appendix E or about the intricacies of factor analysis.

Comparison of Day School and Night School in
Patterns of Initiating Classroom Interaction

Although there are many similarities between the patterns of classroom interaction identified by factoring day-school and night-school data, it is equally informative to note some of the major differences. It seems that a major source of discrepancy between these samples (or analyses) is related to the sequencing of classroom behavior.

Teacher-model-student practice factors. We found "teacher model"-"student practice" factors in both samples. In the day-school analysis separate "model-practice" factors break out for the initiation (V) and follow-up (II) phases of interaction.[2] But in the night-school analysis one factor (I) involves both phases of interaction, while the other (V) is

[2] The Roman numerals in parentheses refer to the factor columns in the respective tables. Note that the factors are printed in arbitrary order but are sequence-numbered in terms of their relative sizes.

for follow-up. The distinction between phases of the "model-practice" interaction pattern in the day-school analysis is further accentuated by some indication that initiation of the sequence leads to a request to repeat the response rather than to automatic continuation of the "model-practice" sequence.

Differences such as these could be largely influenced by the particular teachers involved in day and night-school analyses, but, if so, they are due to differences in the structuring of interaction from moment-to-moment (episode-to-episode) and not to differences in the mean amount of "model-practice" invoked. Likewise, we know that the factors which emerged from our analyses are not due to any large extent to differences between teachers within either sample, because essentially the same factors were found even after partialling out (removing) teacher differences.

Teacher direct-student read factor. Like "model-practice", another major pattern of classroom interaction which seems to display differential sequential organization in different samples is "teacher direct" - "student read". In the night school, again (coincidentally?), we find that initiation of this sequence (IX) leads to its continued expression or repetition--but the teacher may "ask the student to repeat" in continuation of the reading sequence. This suggests rather sustained, continued reading sequences under the teacher's direction, with repetition as necessary. In the day school, on the other hand, factor (IX) suggests a pattern of initial "teacher direct - student read" behavior but, if anything, a low tendency for further teacher direction or explanation; i.e., we see a short reading episode which is not continued. With "teacher direct-student read" we find a contrast between classroom interaction patterns; since something similar happens for both day and night schools, however, perhaps we should take it seriously.

Teacher direct-student read or ask question factor. In addition to what was said immediately above about the "direct-read" sequence in each sample, there seems to be another sequence (VII day, VII night) of "teacher direct" - "student read and/or ask question." This behavior pattern seems to be in contrast to one of "teacher ask question" - "student answer." There is the further complication of a class vs. individual distinction in the night school. There "teacher ask question" applies to the class-as-a-whole, whereas individual students are "directed to read." One way of interpreting what is going on here is to regard this factor as a contrast between two situations which can alternate within the same classroom during any given session of observation; we will pursue this interpretation later in this chapter. In the night-school sample there is some slight suggestion that students may be reading mimeo materials but are being asked to answer questions about pictures. In the day school there is some indication that the teacher's behavior after a student's attempt to read or after asking a question is an explanation; this, in turn, may be followed by more reading and question-asking on the part of the student--a rather tutorial setup.

Teacher question-student answer factor. In each sample there is also another, distinct, "teacher question"--"student answer" factor (III day, IV night) that is not contrasted with student question-asking, just as we already saw distinct "direct-read" factors in each sample. In both samples it is clear that this somewhat more pure "question-answer" paradigm is directed at individual students rather than to the class at large. In both samples there is some indication that the instructor might subsequently repeat the question.

There are a few other subtleties regarding the "teacher-question"--"student answer" factor, especially in the day school. A pattern of some positive feedback is suggested for both samples, and interestingly, a disinclination of the instructor to be writing on the board while asking questions of an individual. This latter pattern probably relates to the obvious utility of the blackboard for communicating to the entire class rather than to individuals. In the day school this writing on the board (vs. questioning an individual) seems to take time and to go along with modeling or explanation--with subsequent student practice, conversation and other behaviors.

Let us now briefly name and codify the four factors introduced above, before moving on to some of the major dimensions of classroom interaction which seem to follow these initiating events: (1) "teacher model"-"student practice"; (2) "teacher direct"-"student read"; (3) "teacher direct"-"individual student read or ask question" vs. "teacher asks question"-"student in class answers"; (4) "teacher questions individual"-"individual student answers".

Patterns of Classroom Interaction Which are Sustained Once Initiated

The "model-practice" pattern seems to lead to subsequent cycling in the night school, just as for "direct-read" there, so we have already had a glimpse of some of the subsequent or follow-up behavior patterns. There are also several other factors of classroom interaction which seem to continue once started and therefore cut across the "initiating" vs. "subsequent" dichotomy which we have set up.

Other factor. One factor which emerged in both analyses (IV day, II night) is simply characterized by the "other" category of both initial and subsequent behavior on the part of both student and teacher. It is further characterized by the fact that the observer made a special notation about the content of the behavior on the observation form and the behavior tended to be of long duration.[3] In the day school we see some indication that drill was not involved.

Student-student feedback factor. Another factor which showed up in both samples is largely student initiated. "Student-student feedback" is the behavior involved here, both initiating and subsequently continued, and it is associated with the observer comment that the feedback occurred in Spanish (VI day, VI night). In the night school we see that this feedback might be associated with "teacher explanation" as well as repeating verbatim what the student has said. It is possible that in this pattern of interaction both the teacher and other students are trying to help an individual get something straight.

Free factor. A factor which emerged only in the day-school sample (VIII) is related to the "free" strategy of instruction (as opposed to question and answer). In the observation Lexicon for Appendix C we see that student initiation

[3]These notations were so diverse that any further classification would have added nothing to the analysis. Examples are given in the Lexicon under "Comments - Other." (See Appendix C.)

of a "free response" stems from the teacher's having indicated only the structure and not the idea or vocabulary to be used (there is in this pattern a very weak suggestion of teacher modeling to initiate, with writing on the board as follow-up).

Comparison of Day School and Night School in Patterns of Follow-Up Classroom Interaction

As for the clearly "subsequent" patterns of interaction detected, we see two major patterns connected with "corrective feedback."

Corrective feedback-model-practice factor. The first factor mentioned in this discussion was "model-practice" and we stated there that it "breaks up" into an initiating and subsequent factor in the day school, but appears as a cycle of continued interaction in the night school. We can now reveal, however, that something more subtle than a simple splitting up or segmenting of behaviors is going on, since a "corrective feedback"--"teacher model"--"student practice" pattern of subsequent interaction was isolated in both samples (day II, night V). Furthermore, the presence of slight but consistent loadings suggests that this pattern of corrective feedback and modeling might well be associated with a prior attempt on the part of the student to read something under the direction of the teacher.

Further, in the night school there is some indication that other students may play a role in modeling the behavior required. In the day school there is some suggestion that the teacher might ask the student to repeat the initial response; observer comments also indicate use of a backward buildup exercise.

Corrective feedback-prompt-answer factor. Another, quite distinct pattern of follow-up interaction involving corrective feedback was recovered in both samples (I day, III night). In this case there is some reason to believe that the feedback is encouraging (especially from the day-school data) and the sequence goes: "corrective feedback"-"teacher prompt or ask to repeat"-"student answer." Since prompting or cueing the student on how to modify his or her response to make it correct predominates ~~here, we~~ will call this factor "CF-prompt-answer."

However, there are other possible teacher reactions in addition to the ones mentioned above; from the night-school data we see that the teacher might further direct the student; in the day school we see the possibility of questioning the student. In either case, the observer comment indicates a strong tendency toward a nonverbal teacher action, especially in the day school. Here also we see the possibility that other students ~~will~~ get involved in modeling, prompting, or student-student feedback (day school). In the night school we see some indication that this particular pattern of interaction may occur while the student is writing at the board.

Teacher question-student answer factor. This description completes our discussion of the classroom interaction factor patterns except for one extra "subsequent" behavior found only in the night-school sample: "teacher question"-"student answer" (VIII). This factor is relatively highly correlated (.26) with the "CF-prompt-answer" factor just discussed, but "teacher question" is included on that factor for the day school. Retaining one too many factors in the night school could account for such splitting apart, especially since this factor has high correlations with several others. One of these other correlated factors (r = .20) is the initiating "teacher question"-"individual student/answer" factor (IV), so again there may be a tendency for cyclical, repeated patterns of interaction in the night school.

It is noteworthy at this point that the highest correlation found between factors (.37) occurred for the day school between "CF-prompt-answer" (I) and "teacher question individual" - "individual student answer" (III). Here, again, we see an association between questioning and prompting. Were one to take these correlations among factors seriously, it can be seen that they would lead to a higher order model of what is going on in the classroom. Unfortunately, correlations among factors can be rather unstable so we hesitate to interpret them here. They have nevertheless been included in Appendix E, Tables 2 and 4 for the sake of completeness; the interested reader should interpret them cautiously.

Classroom Differences in Interaction Patterns

Once the factor analyses of behavioral observations were completed, we were in a position to investigate differences in classroom behavior patterns--the ultimate aim being, of course, to identify interesting contrasts among classrooms which might account for differences in student achievement. We were interested in between-classroom variation in behavior for its own sake, however, and therefore sought a technique which would permit us to optimally discriminate between classrooms on the basis of behavior observed therein.

We settled upon the individual classroom as the focus of study at this stage because our interest centered upon teacher differences as well as upon differences in the ways in which any given teacher might approach students of various ability levels. For purposes of assessing variation within classrooms (i.e., the interaction of a particular instructor with a more-or-less homogeneous group of students) we chose individual sessions of observation as the basic units of analysis. That is, we contrasted overall mean differences between classrooms on each pattern of interaction (factor) with the amount of day-to-day variation observed within classrooms for that pattern of interaction. This was done simultaneously for scores on all nine factors, separately for the day-school and night-school samples.

Specifically, for the technically inclined, factor scores were computed using the complete regression method for each individual episode of classroom interaction. Mean scores were then obtained for each session of observation (approximately 20 episodes per session) and these records were entered into a canonical discriminant function analysis in order to find successive orthogonal linear combinations of the factor scores which maximally discriminated among classrooms relative to day-to-day variation.

Univariate Analyses of Classroom
<u>Differences in Interaction Patterns</u>

As a preliminary to discussion of the discriminant function analyses, let us take a brief look at univariate analyses of variance between classrooms using session means as the basic units of analysis. Of course, there is a separate analysis of variance for scores on each of the nine classroom interaction factors discussed earlier in each sample. These analyses are summarized in Table 3.4 in terms of F-ratios and probabilities for evaluating the null hypothesis of no classroom differences.

TABLE 3.4

Univariate Analysis of Variance Tests for Classroom Differences on Factor Scores

Factor Label	Day School F-Ratio	Night School F-Ratio
"Model-Practice"	4.05**	
"Model-Practice-Model-Practice"		.83
"Direct-Read"	4.32**	
"Direct-Read"		.63
"QA-Individual" vs. "Direct-Read/AQ"	2.33**	
"QA-Individual" vs. "Direct-Read/AQ"		1.24
"QA-Individual"	17.40**	
"QA-Individual"		4.81**
"Student Feedback"	7.70**	
"Student Feedback"		2.10
"Other"	1.73	
"Other"		1.20
"Free Response"	7.88**	
"CF-Model-Practice"	11.50**	
"CF-Model-Practice"		1.60
"CF-Prompt-Answer"	11.98**	
"CF-Prompt-Answer"		4.84**
"Question-Answer" (successive)		3.40*
Between Classrooms Degrees of Freedom	13	5
Within Classrooms Degrees of Freedom	367	44

*p < .05

**p < .01

QA: Teacher Question-Student Answer

AQ: Student Asks Question

CF: Corrective Feedback

We will not undertake a detailed discussion of the univariate analysis of variance results presented in Table 3.4 until after we have completed discussion of the multivariate analysis (canonical discriminant function analysis). The latter will provide an overall test of classroom differences which takes into account any dependencies among various classroom interaction factors. The results in Table 3.4 are presented here mainly to justify our emphasis upon analyzing the day-school data in what follows; and to give the reader a familiar frame of reference prior to undertaking discussion of the multivariate analysis.

From the summary results in Table 3.4 it can be seen that many classroom differences were detected in the day-school sample but, presumably due in part to the smaller number of observation sessions per classroom, few differences were detected in the night-school sample. There were from 15 to 33 observation sessions per classroom in the day school compared to from 9 to 14 such sessions per classroom in the night school.

Session-to-Session Covariation of Classroom Interaction Patterns

Because of the greater precision of day-school classroom comparisons, we will focus there in the following discussion with only a brief discussion later of the night-school results. In considering the univariate analysis of variance results given in Table 3.4, it must be remembered that the various classroom interaction factors are not uncorrelated; therefore, there is apt to be some overlap or redundancy in the F-test results reported. In order to get a feeling for this association between classroom interaction factors as they vary from day to day (observation session to observation session) we can examine the pooled within-classrooms correlation matrix among factor score session means presented in Table 3.5.

TABLE 3.5

Pooled Within Day School Classrooms Correlations Among Factor Score Session Means

	QA IND	GF-P-A	CF-M-P	MOD PRAC	OTHER	DIR READ	QA vs. DIR	SFBK	FREE
QA IND	1.00								
CF-P-A	.45	1.00							
CF-M-P	.05	[illegible]	1.00						
MOD PRAC	-.15	.0[illegible]	.30	1.00					
OTHER	-.18	.02	.27	.56	1.00				
DIR READ	-.24	-.18	-.08	.21	-.02	1.00			
QA vs. DIR	-.01	.10	.03	-.07	-.15	.06	1.00		
SFBK	.02	-.01	-.02	-.03	-.09	-.26	-.16	1.00	
FREE	-.05	-.03	-.14	-.21	-.25	-.25	.00	.13	1.00

Mnemonic label	Interaction pattern
QA IND	"teacher question - individual" - "student answer"
CF-P-A	"corrective feedback" - "teacher prompt" - "student answer"
CF-M-P	"corrective feedback" - "teacher model" - "student practice"
MOD PRAC	"teacher model" - "student practice"
OTHER	"other"
DIR READ	"teacher direct" - "student read"
QA vs. DIR	"teacher ask question" - "student answer" vs. "teacher direct" - "student read and/or ask question"
SFBK	"student - student feedback"
FREE	"free response"

For purposes of discussion we have ordered the rows and columns of the pooled within-classroom correlation matrix in Table 3.5 so as to best reveal the clustering of classroom interaction patterns. Note that most of the positive intercorrelations are near the main diagonal and that several clusters of associated interaction patterns have been blocked off there. Note in Table 3.5 a tendency for the "corrective feedback-prompt-answer" interaction pattern to occur on those days when the "question-answer" pattern is applied to individual students. These two factors thus go together to form a "macro" pattern which links individual episodes of classroom interaction. Note also that the associations presented in Table 3.5 are due only to day-to-day fluctuation in activity patterns within each classroom--they are not influenced by differences in the overall mean level of classroom or teacher behavior, since the latter differences have been removed (partialled out) in the pooling process. We will turn to between-classroom differences once the within-classroom organization of interaction patterns is clarified.

A second cluster which emerges has as its focus the "teacher model"--"student practice" interaction pattern. The inclusion of "corrective feedback"-"teacher model"-"student practice" as well as "other" in this cluster indicates day-to-day covariation in these repetitive, rehearsal-based sequences of interaction. This cluster is linked to "teacher direct"-"student read" through the latter's correlation with "teacher model"-"student practice" (.21). This day-to-day association of "direct-read" with "model-practice" contrasts with the tendency of the former not to be seen on days when a "question-answer" (-.24) strategy is in effect.

It is interesting to note that the bipolar factor which contrasts teacher questioning with student question-asking (under teacher directions to read and ask questions, QA vs. DIR) is relatively independent of the other interaction patterns as they vary from day to day within classrooms. Finally, we see that the more flexible and open patterns of classroom interaction ("student-student feedback," "free response") show a slight tendency to prevail on the same days (.13) but are inconsistent with the "direct-read" (-.25)/"model-practice" (-.21)/"other" (-.25) syndrome.

Multivariate Analysis of Classroom Differences in Interaction Patterns

Moving now to a consideration of overall classroom differences averaged across days of observation we must consider the outcome of canonical discriminant function analysis. This analysis is designed to identify those linear combinations of observed scores (in this case, classroom interaction factor scores) which maximally discriminate between groups relative to within-group variation. As far as this study is concerned, the role of canonical discriminant function analysis is to identify ways in which classroom interaction patterns can be combined to yield stable overall contrasts among the classrooms studied; i.e., to find how the classrooms differ most clearly from one another on the average while displaying minimal day-to-day fluctuations.

Table 3.6 summarizes the canonical discriminant function analysis results for the day-school behavioral observation data. Optimal discrimination was sought between 14 classrooms on the basis of nine factor scores, using observation session means as the basic units of analysis.

An inspection of the results of the canonical discriminant function analysis among day-school classrooms summarized in Table 3.6 reveals that there are two major dimensions of variation between classrooms and two or

TABLE 3.6

Day-School Classroom Observation Data
Canonical Discriminant Function Analysis

Discriminant Function	Root	Cumulative % Trace
1	1.2177	55.6
2	.4970	78.3
3	.1735	86.2
4	.1371	92.5
5	.0812	96.2
6	.0506	98.5
7	.0282	99.8
8	.0056	100.0

Wilk's Λ criterion = .192 for F (117, 2696) = 5.70

Test for residual after removing:	χ^2	df
First discriminant function	315.1	96
First two discriminant functions	166.0	77
First three discriminant functions	106.9	60
First four discriminant functions	59.4	45
First five discriminant functions	30.6	32
First six discriminant functions	12.4	21
First seven discriminant functions	2.1	12

Discriminant Function Weights Scaled for Unit
Within-Groups Variance on Original Scores

Discriminant Function

Variable	I	II	III	IV	V
QA-IND	-.250	-.076	-.126	-.225	.022
CF-P-A	-.158	-.127	.043	.140	.220
CF-M-P	.262	-.105	.000	-.231	.175
MOD PRAC	.044	-.252	-.177	.197	.078
OTHER	-.148	.293	.205	-.177	.108
DIR READ	.000	.043	.185	-.054	.260
QA vs. DIR	-.005	.001	.026	.241	.062
SFBK	-.086	-.234	.324	-.009	-.053
FREE	-.154	.267	.041	-.031	.248

three more minor dimensions. We do not, however, find interpretation of the discriminant function weights as they are given in Table 3.6 either straight-forward or enlightening. In order to gain more insight into the meaning of these results, therefore, we have elected to consider the between-groups correlations between the original scores and the discriminant axes. These correlations are readily obtained and have the advantage that they can be treated much like factor loadings of observed variables on orthogonal factor axes--the canonical discriminant functions (Cliff and Krus, 1976).

Because the first two canonical discriminant functions account for the majority (78 percent) of all differences between day-school classrooms, in terms of stable patterns of teacher-student interaction, we have chosen to plot the classroom group centroids[4] on these axes in Figure 3.2. In the figure each classroom is identified by a letter-number combination. The initial letters range from A through F and identify the six day-school instructors. The numbers range from 1 to 3 and refer to proficiency level of the class being taught as measured by the John Test. The lower case letters identify different classes at any given level where they are taught by the same teacher.

In addition to classroom group centroids, we have drawn vectors in Figure 3.2 to represent the way the nine original factors of momentary classroom interaction project into the discriminant space. Note that the important features of these vectors are their directions and their relative lengths--their absolute lengths are arbitrary and have been scaled merely for convenience of plotting. By visualizing the direct perpendicular projection of classroom

[4]The origin of Figure 3.2 is located at the center of gravity (centroid) of all the classes.

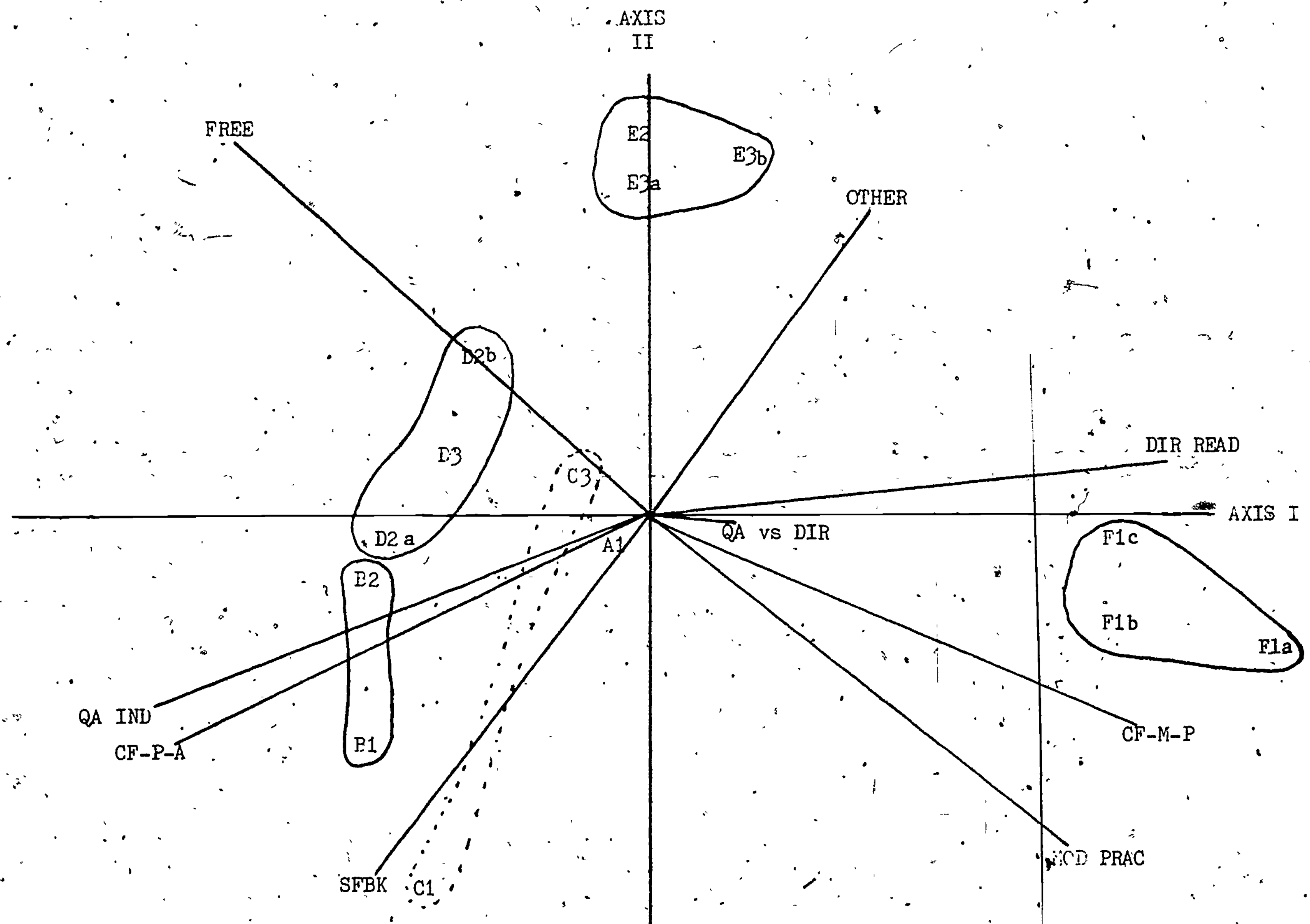

Figure 3.2

Plot of Day-School Classroom Centroids and Classroom Interaction Factors in the Space Defined by the Two Largest Canonical Discrimiant Axes

- 85 -

97

98

centroids onto these vectors (extended through the origin if necessary) one can get a feeling for the relative ordering among classrooms in terms of the behavior under consideration. By noticing the relative collinearity or perpendicularity of these vectors one can get a feeling for how classroom behaviors associate or disassociate in the profiles which distinguish classrooms from one another.

As for the classroom centroids plotted in Figure 3.2, the most outstanding feature is the obvious clustering of classrooms taught by the same teacher. This phenomenon is particularly striking in view of the fact that no information about who taught what class entered into any of the analyses leading to these results. The evidence is thus incontrovertible that teachers have consistent and distinct "styles" of interaction with students--styles which do not in general vary markedly even when teaching classes of quite different initial ability level.

Only one teacher, C, appears to employ widely different strategies when teaching students of differing ability levels; however, it must be pointed out that no other teacher had the opportunity to teach classes containing students of such widely different ability levels.

Because composite variables in canonical form are notoriously difficult and dangerous to interpret we will not attempt to label the axes in Figure 3.2 We can note, however, that the first (horizontal) axis is a contrast between about the same clusters of variables that we identified in the pooled within-classrooms correlation matrix in Table 3.5. That is, a general pattern of DIR READ, CF-M-P, MOD FRAC, and OTHER is contrasted with QA IND and CF-P-A.

The between-groups correlations of the original factor score variables with the canonical axes can serve as a basis for transforming (rotating) the canonical variates into a more theoretically informative and interpretable position. For this purpose we chose to use only the first four discriminant axes since they account for most (93 percent) of the variation between groups and little significant variation remains ($\chi^2 = 59.4$; 45 df) after the fourth axis is considered. Table 3.7 presents the DIRECT GEOMIN transformed between-classrooms canonical-variate factor pattern matrix, the factor correlation matrix, and the factor structure matrix.

Interpretation of Transformed Discriminant Function Axes

Upon transformation of the first four between-classrooms canonical variates we came up with two major contrasts between teacher-student interaction patterns (still accounting for the majority of the between-groups variance) and two smaller variates each relating essentially to only one or two patterns of momentary interaction. Because the transformed canonical variates are fairly highly correlated it is clear that there is a second order general factor in operation here. It is therefore necessary to take into account both the transformed loadings (weights) and correlations (projections) in order to interpret these factors. The existence of a second order general factor suggests that all of the transformed canonical variates are getting at different aspects of the same general distinction between a highly structured, directive, and controlled classroom climate and a more relaxed, free, and flexible climate (roughly the horizontal axis in Figure 3.2).

Interpretation of Discriminant Function Axes

In an attempt to shed light upon the nature of the four transformed canonical discriminant function axes presented in Table 3.7, we scored every episode of classroom interaction in the day-school sample in each of these four ways. That is, we obtained the appropriately weighted linear combination of classroom interaction factor scores for each episode of interaction. We

TABLE 3.7

Loadings of Classroom Interaction Scores
On Transformed Canonical Variates

	I	II	III	IV
QA IND	.92	.01	.06	.18
CF-P-A	.77	.06	.28	-.06
CF-M-P	-.16	-.89	-.08	.29
MOD PRAC	.01	-.93	.07	-.10
OTHER	-.86	.36	.02	.43
DIR READ	-.77	-.24	.03	-.14
QA VS DIR	-.08	.04	-.01	-.84
SFBK	.00	.02	.97	.03
FREE	-.05	1.01	-.09	-.05

Correlations Among Transformed Canonical Variates

	I	II	III	IV
I	1.00			
II	.50	1.00		
III	.50	.13	1.00	
IV	.11	.23	.14	1.00

Correlations of Classroom Interaction Scores
With Transformed Canonical Variates

	I	II	III	IV
QA IND	.98	.52	.54	.28
CF-P-A	.94	.47	.67	.08
CF-M-P	-.62	-.92	-.23	.06
MOD PRAC	-.43	-.94	-.06	-.30
OTHER	-.63	.02	-.30	.42
DIR READ	-.89	-.66	-.40	-.27
QA VS DIR	-.16	-.19	-.17	-.84
SFBK	.48	.11	.98	.16
FREE	.41	.96	.01	.16

then averaged these scores separately across all positive instances of each category of every item in the original classroom observation system (Table 3.2). The resulting category means are symbolically depicted in Figures 3.3 through 3.6, for the respective transformed discriminant function axes. There are ten columns in each figure, corresponding to the ten major sections[5] of Table 3.2; the vertical position of each category label corresponds to its mean score (weight) on the contrast in question over all instances of the behavioral category in question. Each figure thus depicts a contrast between classroom interaction patterns employed by teachers in the day-school sample. The method used here for weighting categories is reminiscent of the method of reciprocal averages as well as of multidimensional analysis of contingencies (Guttman, 1941).

It must be kept in mind that the four contrasts presented in Table 3.7 and Figures 3.3 - 3.6 were arrived at because they discriminate quite sharply between behavior patterns observed in different day-school classrooms in this particular study. We have already seen that the classroom differences detected are largely a function of teacher differences, however, so it is quite likely that other contrasts in teaching behavior would become salient given a different sample of teachers.

<u>First axis</u>. The coefficients presented in Table 3.7 suggest that the first axis is a specific distinction between the "question-answer-corrective feedback-prompt-answer" paradigm of classroom interaction and the "direct-read" and/or "other" paradigm.[6] When we get back down to the level of

---

[5] The "time" entry in the first column refers to those episodes of interaction which endured for more than one minute.

[6] If the reader will refer back to Figure 3.2 it will be seen that this contrast can be approximated by tilting the horizontal axis about 30 degrees counter-clockwise and then reflecting the direction of scoring.

FIGURE 3.3

Classroom Observation Category Means on Transformed Canonical Discriminant Axis I

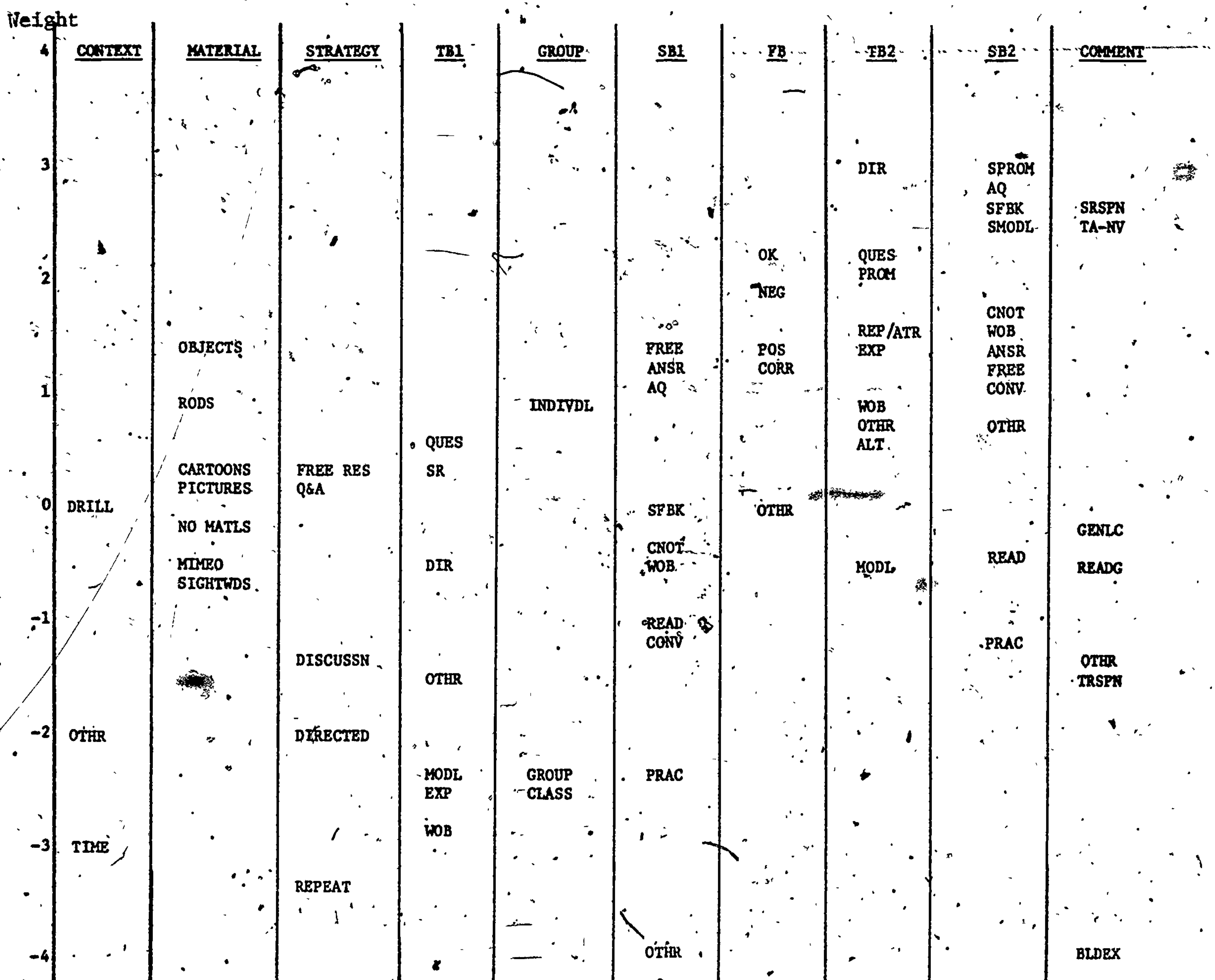

FIGURE 3.4

Classroom Observation Category Means on Transformed Canonical Discriminant Axis I

Weight	CONTEXT	MATERIAL	STRATEGY	TB1	GROUP	SB1	PB	TB2	SB2	COMMENTS
4										
3			FREE RES							
						FREE		OTHR	OTHR	
							OTHR			
2										
										OTHR
									FREE	TA-NV
									SFBK	
	OTHR									
1				OTHR				WOB	SPROM	
			REPEAT	SR		OTHR		PROM	CONV	
		OBJECTS				AQ	OK			
	TIME	PICTURES						ALT		
		RODS			INDIVDL		POS	QUES	SMODL	
				QUES		ANSR/CNOT	NEG		ANSR/WOB	
		CARTOONS								
0	DRILL					SFBK	CORR	REP/EXP	AQ	
			DISCUSSN						CNOT	SRSPN
		NO MATLS	Q&A					DIR		GENLC
				DIR						
		MIMEO								READG
						CONV			READ	TRSPN
			DIRECTED							
						WOB		ATR		
		SIGHTWDS			GROUP					
1						READ				
				EXP						
				WOB	CLASS					
2										
								MODL		
						PRAC				
									PRAC	
				MODL						
3										
										BLDEX
- 4										

FIGURE 3.5

Classroom Observation Category Means on Transformed Canonical Discriminant Axis I

Weight	CONTEXT	MATERIAL	STRATEGY	TB1	GROUP	SB1	FB	TB2	SB2	COMMENTS
4						SFBK (↑4.2)			SE (↑7.0)	SRSPN (↑13.0)
3		OTHR								
									SPROM	
2							NEG		SMODL	GENLC
									WOB	
								DIR/QUES	CONV.	TA-NV
								PROM		
							OK		CNOT	
1									AQ	
		OBJECTS				ANSR/CNOT	CORR	ALT		READG
						CONV		ATR	ANSR	
		SIGHTWDS		DIR			POS	REP		
		RODS		OTHR	INDIVDL			EXP	FREE	
		CARTOONS		QUES		READ		WOB	READ	
		MIMEO	Q&A/FREE RES							
0	DRILL					WOB		MODL		
		PICTURES				AQ				
					GROUP					
		NO MATLS				FREE				
	OTHR								PRAC	
			DIRECTED							
				SR						
			REPEAT			OTHR				OTHR
	TIME									TRSPN
			DISCUSSN							
1				EXP	CLASS					
							OTHR			
				WOB						
				MODL		PRAC			OTHR	
								OTHR		
2										
										BLDEX
3										
4										

108

FIGURE 3.6

Classroom Observation Category Means on Transformed Canonical Discriminant Axis I

Weight	CONTEXT	MATERIAL	STRATEGY	TB1	GROUP	SB1	FB	TB2	SB2	COMMENTS
4										
								OTHR		
3										BLDEX
									OTHR	
				OTHR		READ			READ	SRSPN
2										
									AQ	OTHR
						AQ				
				DIR				DIR		
								MODL		
						FREE	NEG	EXP	WOB SFBK	
1	OTHR	SIGHTWDS					OTHR		FREE/PRAC/CNOT	
						OTHR		REP	SMODL	
			REPEAT		INDIVDL	SFBK		QUES	SPROM	
		CARTOONS	FREE RES				CORR/OK		CONV	READG
	TIME	RODS		SR			POS	WOB		
0		MIMEO				WOB		PROM		GENLC
	DRILL	PICTURES	Q&A			ANSR		ATR	ANSR	TRSPN
		OBJECTS						ALT		
		NO MATLS				CNOT				
			DIRECTED DISCUSSN	QUES		CONV				TA-NV
1										
					GROUP					
					CLASS					
2				EXP						
				WOB						
						PRAC				
3										
4				MODL						

- 93 -

contrasting individual aspects of classroom interaction in Figure 3.3, however, we see that a good deal of interchangeability characterizes the follow-up aspect of the positive pole of this contrast. We see that this "question-answer" pattern directed to individual students and followed by "corrective feedback" may lead to quite a variety of teacher behaviors in addition to "prompt". "Student-student feedback" likewise figures prominently as a follow-up pattern, as does "student question-asking". "Questioning" on the part of the teacher is likely to involve "objects" or "rods" and a "free response" strategy may just as well prevail as a "question and answer" strategy.

From Figure 3.3 we can also see that the "direct-read", "other" pole of this first contrast between classroom interaction patterns is characterized by many subtleties at the level of individual episodes of interaction. The likelihood of teacher "explanation", "writing on the board", and "modeling" is noteworthy, given the indication that a time-consuming strategy of "repetition", "directed dialogue", or "discussion" is in effect.

Second axis. From Table 3.7 and Figure 3.4 the reader can see that the second transformed discriminant function axis picks up a contrast between the "free response" mode of interaction and the "model-practice-corrective feedback-model-practice" paradigm.[7] As for all four transformed axes, the positive pole applies more to individualized instruction while the negative pole refers to class or group-oriented instruction. It is notewrothy that the "free response" strategy leads to "other" categories of feedback and follow-up behavior.

[7] Referring back to Figure 3.2 the second contrast can be visualized by tilting the horizontal axis about 45 degrees clockwise and reflecting the direction of scoring.

Third axis. The third transformed canonical variate is not a strong bipolar contrast, unlike the others, but relates quite simply to the presence of "student-student feedback" as a component of classroom interaction. Along with this student feedback is some indication of corrective feedback on the part of the teacher and other follow-up behaviors similar to those seen for the first axis.

Fourth axis. Finally, the fourth transformed canonical variate is interesting in that it relates strongly and negatively to the original bipolar contrast found when we factor analyzed classroom interaction patterns as they varied from episode to episode: "teacher question"-"student answer" vs. "teacher direct"-"student read and/or ask question". We can now reiterate our earlier conjecture that the "direct-read/ask question" pole of this factor is, indeed, an interaction pattern characteristic of relatively free, unstructured classrooms, whereas the "question-answer" pole characterizes more highly structured classrooms. This argument is borne out to some extent by the positive signs of the correlations (albeit low) between this transformed canonical variate and the others, taken together with the fact that the original QA vs. DIR score has a negative weight in Table 3.7. From Figure 3.6 we see that "other" goes along with "direct" as the teacher behavior characteristic of the positive pole of the contrast in question. Likewise, it can be seen that sightwords or cartoons are the likely stimuli which students are directed to read and ask questions about. On the other pole of this contrast between teaching strategies we now see emphasis upon the "model-practice" pattern of initiating interaction with the class.

What the transformed canonical discriminant function axes seem to reveal is a certain amount of complexity or multiplicity of determination on the part of several patterns of manifest classroom interaction. Thus, the observed interaction pattern "teacher direct"-"student read" can either be part of a highly structured, "programmed" classroom interaction climate (the negative pole of the first axis) or part of a more open-ended, unstructured setting for interaction (the positive pole of the fourth axis). Looking at the opposing poles of these same axes, respectively, we can infer that "teacher asks question" can likewise be part of an encouraging, prompting, follow-up pattern or it can form part of a more demanding pattern of drill.

Of course, we have been attempting to identify the transformed discriminant axes with hypothetical constructs which might underlie and help account for variation in observed patterns of classroom interaction. In so doing we must go somewhat beyond the empirical results to engage in conjecture and speculation; the reader is thus entitled to their own interpretation of the transformed discriminant axes or may abstain from theoretical speculation. In this connection we can note that the particular orientation arrived at through transformation is completely arbitrary in the sense that it has no effect whatever upon the total amount of between-classroom variance accounted for. It is only when we begin to attribute fractions of the total between-classrooms variance to one axis or another that the orientation of transformed axes must be fixed by some means (e.g., canonical form or simple structure).

RELIABILITY OF CLASSROOM INTERACTION PATTERN CONTRASTS

The succession of data reduction procedures applied to classroom behavioral observations has ultimately led to only 56 essential numbers, the classroom centroid coordinates on each of four transformed discriminant function axes. These coordinates define four contrasts between day-school classrooms which can now be used in an attempt to predict student achievement over the school year from classroom interaction experiences. How reliable are these indices of classroom interaction since unreliability limits predictive validity? The issue of reliability also provides us an opportunity to return, as promised, to the univariate analysis of variance results presented in Table 3.4, since we will discuss reliability in terms of variance components.

We stated earlier that the nine factors of momentary classroom interaction detected by analyzing the total day-school sample of some 7,000 episodes were recovered in substantially the same form even after partialling out all teacher and classroom effects. This result led us to conclude that the factors were not due to any large extent to differences between teachers. The analyses of variance summarized in Table 3.4 indicate, however, that large differences exist between day-school teachers on almost every factor. We can reconcile these results by considering components of variance in the total sample of classroom episodes.

Components of Variance Analysis of Classroom Interaction Factors

The sample comprises a completely nested four-level analysis of variance design having approximately 20 episodes nested within each session, approximately 25 sessions per classroom, and about two classrooms per teacher.

From Table 3.8 it can be seen that from 53 to 90 percent of the total variance of each classroom interaction factor can be attributed to within-session variation (see line labeled Episode in Table 3.8); i.e., to variation from interaction episode to interaction episode during the same session (day) of observation. The factors are thus primarily factors of within-session variation.

In Table 3.8 we have indicated the nested analysis of variance F-test probability levels associated with each component of the total sample variance, along with the associated degrees of freedom. From these results, which are more complete than the simple tests of classroom differences relative to session-to-session variation given in Table 3.4, many aspects of variation in classroom interaction patterns can be clarified. It is clear, for instance, that only a very small and generally nonsignificant contribution to variation in interaction patterns can be attributed to differential treatment of separate classrooms by the same teacher (see line labeled Classroom in Table 3.8)--an outcome already suggested by the compact clustering of classroom centroids for most teachers in Figure 3.2 of the multivariate analysis.

In this connection it is noteworthy, however, that "student-student feedback" (SFBK) and "teacher question-student answer" vs. "teacher direct-student read/ask question" (QA VS DIR) are the only two classroom interaction factors for which very small and insignificant teacher differences are indicated in Table 3.8. Recall that these two factors are heavily weighted in the last two transformed canonical discriminant axes (Table 3.7, Figures 3.5 and 3.6); we will shortly see that the latter two axes are somewhat sensitive to differential handling of separate classrooms by the same teacher. Notice that the classroom interaction factors have been listed and partitioned in

TABLE 3.8

Components of Unit Total Sample Variance of Day-School Classroom Interaction Factors Due to Teacher, Classroom, Session, and Episode Differences

Source	QA IND	CF-P-A	OTHER	DIR READ	FREE	MOD PRAC	CF-M-P	SFBK	QA VS DIR	df
Teacher	16.0**	9.4**	1.8*	5.9*	11.6**	4.6**	4.6*	.5	.5	5
Classroom	0.0	.2	0.0	0.0	1.6*	0.0	1.5**	1.8**	1.0	8
Session	15.2**	12.7**	30.1**	24.4**	33.9**	11.0**	10.8**	7.5**	33.8**	368
Episode	68.8	77.6	68.1	69.7	52.8	84.4	83.1	90.2	64.7	6704
Discriminant Axis	+I	+I	-I	-I	+II	-II	-II	+III	-IV	

*p < .05

**p < .01

Table 3.8 in a way which reveals their connection to the transformed canonical variates in Table 3.7--the signed Roman numeral at the bottom of each column is to remind the reader of the major weight received by each classroom interaction factor in defining the transformed discriminant function axes.

We can now reconcile the discovery of highly significant teacher differences on factors which are primarily indicators of episode-to-episode variation within the same session in any given classroom. The components of variance in Table 3.8 indicate that we gain little practical information about what might take place in any given episode of classroom interaction through knowledge of which classroom is being observed (from .5 to 16 percent of variance can be accounted for, line labeled Classroom in Table 3.8). In order to account for the variance among episodes in a theoretically useful way we are thus left to the postulation of hypothetical determinants (classroom interaction factors) which apply regardless of which teacher is involved.

The relatively large components of variance for Sessions suggest that a number of these factors fluctuate substantially from day-to-day.[8] Especially noteworthy are the large components of variance for interaction patterns involving "free response" (33.9 percent), "other" (30.1 percent) and teacher "direction" (24.4 percent to 33.8 percent) behaviors. These results suggest that any given teacher might be inclined to devote certain sessions to these "specialized" activities but refrain from doing so in other sessions. Note, however, that considerable session-to-session variance characterizes many of the classroom interaction factors. This result indicates that each session of

[8]Some of the session-to-session fluctuation could be attributed to observer biases, since various observers collected data on different sessions. This suspicion will be dispelled shortly, however, when we consider the inter-observer agreement approach to reliability.

observation provides a rather narrow view of what is going on in the classroom in general and may explain why stable teacher differences could not be well established with the limited night-school observation schedule.

In summary, differences among teachers account for from .5 to 16 percent of the total variance in the observational data (see line labeled Teacher in Table 3.8). The differences among the teachers are statistically significant on all but two factors (SFBK and QA VS DIR). In ordinary language the teachers teach differently, and they differ most on the factor "question-answer--individual student" (QA IND) and "free response" (FREE). There is not much evidence that the particular classroom being handled by a teacher has any influence upon the patterns of interaction which take place therein, as we have seen earlier in connection with Figure 3.2. The major component of overall variation in classroom interaction patterns can be attributed to alternation in interaction patterns from episode-to-episode, which is not surprising, but session-to-session changes also play an important role. Notice that session differences are statistically significant on all nine factors.

Fortunately, predicting what might take place in each individual classroom episode or even in any given session of observation was not our task in this study, so we can be content to account for such variation by reference to hypothetical determinants or "factors" of momentary classroom interaction. Our aim was to detect stable classroom or teacher differences and it is to assess our attainment of that goal that a reassessment of the components of variance in Table 3.8 is required. That significant teacher differences were obtained despite the fact that much session-to-session and episode-to-episode variation occurred is not hard to understand because so many sessions and episodes were recorded for each teacher.

When the components of variance in Table 3.8 are weighted to take into account the fact that approximately 1000 episodes of interaction were observed per day-school teacher, about 500 per classroom, and about 20 per session, we see that a very large fraction of the observed variation among teacher means on each classroom interaction factor is reliable (i.e., can be attributed to true differences between teachers). Since substantially the same issue is addressed by the F-tests among teachers (given in Table 3.4), it is clear why so many significant results were found in the day-school sample.

Components of Variance Analysis of Transformed Discriminant Function Scores

The multivariate analysis of variance (canonical discriminant function analysis) in Table 3.6 revealed only a few significant contrasts among day-school classrooms. From the first four of these canonical discriminant functions we derived the four transformed axes discussed earlier in this chapter. Since these are the classroom interaction contrasts which we wish to employ to predict student achievement, it is in order to determine how sensitive teacher scores on these contrasts are to various possible underlying sources of variation: teacher differences, classroom differences, session differences, episode differences, and observer differences as well. The issue of inter-observer differences will be taken up shortly.

First, however, let us look at nested analysis of variance results (components of variance) for the transformed discriminant axes, just as was done for the nine original factors of classroom interaction in Table 3.8. The upper section of Table 3.9 gives components of unit total sample variance for the transformed canonical discriminant function axes comparable to the entries in Table 3.8. From these entries we can see that there is substantial episode-to-episode variation in all four of the classroom interaction contrasts (from 58 to 86 percent of the total variance). Note, however, that the maximization of between-classrooms variance relative to session-to-session variation (defining the canonical orientation of the original discriminant function axes) has carried over from the analysis presented in Table 3.8 in the form of large components of total sample variance attributable to teachers (Table 3.9, upper section).

Teacher differences account for approximately one-fourth of the total observed variance on each of the first two transformed axes (I and II). What this means, in a practical sense, is that we could reduce our uncertainty about which pole of either contrast might manifest itself in any given episode of classroom interaction by a substantial amount merely by knowing which teacher is in charge. If the teacher is high on the first contrast (teacher B in Figure 3.2), we would do well to predict "question-answer-corrective feedback-prompt-answer" interaction episodes. On the other hand, if the teacher is low on the first contrast (teacher F or perhaps E), then we would do well to predict "direct read" and "other" episodes.

Considering the second contrast we would proceed in a similar fashion, predicting the "free response" mode of individualized instruction for the positive pole (teachers D, B, and E), and group "model-practice-corrective feedback-model-practice" chains of episodes for the negative pole (teacher F).

Even the last two transformed canonical variates ("student-student feedback" and "direct read and/or ask question vs. question-answer,"

TABLE 3:9

Components of Unit Total Sample Variance of Transformed Discriminant Axes Due to Teacher, Classroom, Session, and Episode Differences

Source	I	II	III	IV
Teacher	22.9**	25.7**	4.5*	4.0*
Classroom	.8*	.3	2.1**	1.0**
Session	13.4**	15.8*	7.1**	15.1**
Episode	63.0	58.2	86.2	79.9

Relative Contributions of True Score Differences Between Teachers, Classrooms, Sessions, and Episodes to Observed Variance Among Teacher Means on Transformed Discriminant Axes

Source	I	II	III	IV
Teacher	97.2%	98.2%	79.8%	83.4%
Classroom	1.5%	.5%	16.5%	9.2%
Session	1.1%	1.1%	2.3%	5.9%
Episode	.2%	.2%	1.4%	1.5%

*$p < .05$

**$p < .01$

respectively) show larger teacher and classroom components of variance in Table 3.9 than do the original factors of classroom interaction from which they are largely derived. However, it is clear that the latter are not major contrasts in teaching styles so we could not use knowledge of average teacher performance on these variates to predict much about individual episodes of classroom interaction. Let us turn, therefore, to a consideration of how reliably the teacher means on these contrasts can be estimated per se.

The lower section of Table 3.9 gives a breakdown of the weighted contribution of each factor in the nested observation design to observed variation in teacher means on each transformed discriminant function axis. Whereas the first two contrasts are almost perfectly reliable indicators of teacher differences (97 to 98 percent of the observed variation in teacher means can be attributed to true differences among teachers), the last two contrasts are somewhat sensitive to differential interaction in various classrooms taught by the same teacher.

It is clear that observation of more different classrooms would be required before we could get a very reliable indication of how inclined any given teacher is to allow "student-student feedback" (III) to take place. From an inspection of Figure 3.3, moreover, it can be seen that the original classroom interaction factor of "SFBK" is oriented in a direction which is sensitive to variation in the way in which teachers C, D, and B handle classes which vary in level; more "student-student feedback" tends to occur in lower level classes. Note the high episode-to-episode variation in "student-student feedback" (III, opposite Episode), however, as well as the low session-to-session variation (III, opposite Session)--this difference is diagnostic of the generally low frequency with which this form of interaction occurs (see Table 3.2).

Both more classrooms per teacher and more observation sessions per classroom would be required to pin down precisely any given teacher's tendency to "direct" individual students while they "read and/or ask questions" vs. their tendency to engage the entire class in "practice" sessions with "modeling," "writing on the board," "explanation," and "questioning" (IV in Table 3.9). Apparently individual teachers are somewhat flexible about alternating between these strategies from class-to-class as well as from session-to-session, which helps to explain why a bipolar factor emerges in the analysis even when individual classroom interaction episodes are analyzed.

The components of variance discussed above reveal that we are dealing with highly reliable indices of teacher variation; the reliabilities of teacher means on transformed discriminant axes range from .80 to .98 in Table 3.9. But a more important feature of this investigation is the evidence it provides that those classroom interaction contrasts which show differential treatment of different classrooms by the same teacher are diagnostic of variation in the ability levels of the classes involved. Whereas our initial impression of Figure 3.2 holds true--teachers do have distinct and consistent styles of interaction with students as indicated by transformed axes I and II--there is also evidence that certain more limited aspects of classroom interaction vary from classroom-to-classroom taught by the same teacher. That the latter variation might be in response to student characteristics is of great interest and will be taken up later.

Inter-Observer Reliability

A final question about reliability remains before we can proceed to our main task of relating student achievement to teacher performance: do different observers record the same aspects of variation in classroom interaction patterns? In order to study the issue of possible observer bias, data was collected in some of the early observation sessions by pairs of independent observers observing the same sessions. Since three observers were used in the study, it was arranged to have concurrent observations made by each of the three pairs in eight different classrooms.

We are not in a position to make a strict comparison of inter-observer agreement in the usual sense, since observers worked independently and there is no way to know which of their data-sheet entries should correspond. (The relevance of this correspondence is frequently ignored in estimating inter-observer reliabilities.) However, we can compare data on the basis of individual sessions. In Table 3.10 we have summarized the results of this comparison for each of the three pairs of observers, using as the basic units of analysis session means on the four trarsformed classroom interaction contrasts. It is appropriate to look at possible observer influences on these scores because we have used these scores to predict differences in student achievement.

Note that we are discussing observer assessment of session-to-session variation in these interaction patterns, but we have already seen from Table 3.9 that enough sessions of observation were obtained in the day-school sample to yield highly reliable teacher means despite any session-to-session variation. Since each classroom was visited by all observers, the influence of any possible observer biases on classroom mean scores is indicated by session-to-session variation, which as we have seen, accounts for less of the variance than teacher differences (lower section of Table 3.9).

From Table 3.10 it can be seen that session-to-session correlations among transformed discriminant function scores derived from the data collected concurrently by different observers range from .95 to .99 for

TABLE 3.10

Means, Standard Deviations, and Correlations for Pairs of Observers on Transformed Discriminant Function Scores

Variate		Observer Pairs 1	2	1	3	2	3
I	$\bar{X}$*	.93	.70	.13	.32	.33	.52
	s	2.36	1.89	1.62	1.47	1.45	1.58
	r	.98		.99		.99	
II	$\bar{X}$	.65	.54	.00	.04	.13	.32
	s	1.01	1.13	1.39	1.73	1.17	1.59
	r	.95		.96		.98	
III	$\bar{X}$	.47	.08	.08	.59	.26	.47
	s	1.42	.96	1.82	1.82	1.54	1.03
	r	.66		.81		.37	
IV	$\bar{X}$	.87	.55	-.13	-.09	-.11	.23
	s	2.44	2.29	1.50	1.65	1.83	1.63
	r	.97		.97		.97	

*$\bar{X}$ = mean, s = standard deviation, r = product moment correlation; eight observations.

all but the third axis. For the latter, the correlations suggest that certain observers may have had difficulty either recognizing or recording "student-student feedback in Spanish." Since the lowest correlations involve Observer 2, it is possible that this one individual is the problem;[9] however, none of these correlations are as high as for the other axes.

One problem is that this variable (as we mentioned earlier) occurs relatively infrequently--particularly in the more advanced classes. For this reason each such occurrence is highly weighted (as seen in Figure 3.5) and the failure to record any given instance could have a sizeable influence upon the outcome for that session.

When correlations comparable to those given in Table 3.10 are studied for all nine of the original classroom interaction factors, a similar conclusion is reached; only for "student-student feedback" is there a serious problem with inter-observer correlation. These results are briefly summarized in Table 3.11.

Finally, it can be pointed out that inter-observer correlations for different scores were much lower than for the same score, indicating that the observers were indeed discriminating between different classroom interaction patterns.

NIGHT-SCHOOL TEACHER DIFFERENCES ON CLASSROOM INTERACTION FACTORS

As we saw in Table 3.4, fewer differences were detected among night-school classrooms (only one classroom per teacher) than among day-school classrooms. This result could be due to a real lack of differentiation of teaching styles in the night-school, or to the low number of sessions of observervations per teacher.

[9] Notice from Table 3.1 that observer 2 collected very little data in the study.

TABLE 3.11

Interobserver Correlations for Nine
Classroom Interaction Factors (8 Observations)

Variate	Observer Pairs 1 2	1 3	2 3
CF-P-A	.97	.99	.99
DIR READ	.97	.93	.86
MOD PRAC	.96	.98	.99
OTHER	.79	.99	.96
QA vs. DIR	.98	.94	.96
QA-IND	.99	.98	.98
CF-M-P	.82	.96	.52
SFBK	-.13	.42	-.09
FREE	.94	.96	.97

Our analyses of the classroom interaction factors for the night school paralleled what was done for the day school. We will only present a brief discussion of the night-school results in what follows, however, because of the limited data available. As for the day school, it is in order first to consider day-to-day associations among classroom interaction factors (pooled within-classrooms), then to move on to classroom (teacher) differences.

Table 3.12 gives the (pooled) correlations among night-school classroom interaction factors as they vary across sessions taught by any given teacher. These correlations are comparable to those given in Table 3.5, and we have again ordered the variables so as to reveal clustering.

The reader will notice that session-to-session covariation of classroom interaction factors is quite different for the night-school and day-school samples, a contrast with the high degree of similarity found when we compared structural features of the individual classroom interaction factors for the two samples. It suggests that the same factors of classroom interaction are present in both samples at the level of individual episodes, but that variation in overall classroom organization is different in these samples. The night-school correlations show a tendency for drill ("direct-read"; "model-practice-model-practice") to vary in conjunction with prompting ("corrective feedback-prompt-answer"; "student-student feedback") across sessions. By and large, these associations are not as easy to interpret as for the day school.

The first two canonical discriminant axes account for 73 percent of the variation between the six night-school classrooms (relative to pooled within-group variation), as can be seen from Table 3.13.

TABLE 3.12

Pooled Within Night-School Classrooms Correlations Among Factor Score Session Means

	DIR READ	MOD PRAC	SFBK	CF-P-A	OTHER	CF-M-P	QA_2	QA IND	QA vs. DIR
DIR READ	1.00								
MOD PRAC	.69	1.00							
SFBK	.33	.32	1.00						
CF-P-A	.41	.23	.28	1.00					
OTHER	-.03	-.16	-.09	.02	1.00				
CF-M-P	-.46	-.15	.06	-.18	-.02	1.00			
QA_2	-.13	-.33	-.27	.31	.10	-.08	1.00		
QA IND	-.11	.14	-.38	-.39	-.30	.07	-.04	1.00	
QA vs. DIR	.12	.04	-.35	.08	-.13	-.28	.17	.24	1.00

Mnemonic label	Interaction pattern
DIR READ	"teacher direct" - "student read"
MOD PRAC	"teacher model" - "student practice" - "teacher model" - "student practice"
SFBK	"student - student feedback"
CF-P-A	"corrective feedback" - "teacher prompt" - "student answer"
OTHER	"other"
CF-M-P	"corrective feedback" - "teacher model" - "student practice"
QA_2	"teacher question" - "student answer" (follow-up)
QA IND	"teacher question - individual" - "student answer"
QA vs. DIR	"teacher ask question" - "student answer" vs. "teacher direct" - "student read and/or ask question"

TABLE 3.13

Night-School Classroom Observation Data
Canonical Discriminant Function Analysis

Discriminant Function	Root	Cumulative % Trace
1	1.4645	47.2
2	.7973	73.0
3	.5945	92.1
4	.1796	97.9
5	.0662	100.0

Wilk's Λ = .112565 for F (45, 164) = 2.30

Test for residual after removing:	χ^2	df
0. First discriminant function	92.8	45
First two discriminant functions	54.5	32
First three discriminant functions	29.6	21
First four discriminant functions	9.7	12
First five discriminant functions	2.7	5

Discriminant Function Weights Scaled for Unit
Within-Groups Variance on Original Scores

Discriminant Function Axes

Variable	I	II	III
DIR READ	.029	-.217	.026
MOD PRAC	.027	.210	-.034
SFBK	.095	-.012	.226
CF-P-A	-.219	.191	-.008
OTHER	-.060	.075	.152
CF-M-P	-.051	-.054	-.040
QA_2	.175	.149	.087
QA IND	-.134	-.095	.269
DIR vs. QA	-.018	-.115	-.020

Just as for the day-school data, we have plotted classroom centroids on the first two canonical discriminant function axes in Figure 3.7. Night-school teachers have been identified in the figure with the letters G through L and we have also included vectors representing the original nine classroom interaction factors. Interpretation of this figure can proceed in the same manner as for Figure 3.2. Again, we find the night-school results to be less intuitively compelling than the day school results.

RELATION OF TEACHER BACKGROUND CHARACTERISTICS TO TEACHING STYLES

As mentioned at the conclusion of Chapter Two, considerable information was obtained about each teacher in both the night and day-school samples through a questionnaire administered at the outset of the study. It is of interest to see whether a teacher's performance in the classroom might be related to this prior information about his or her education, experience, and preference among teaching styles.

Only 12 teachers were studied, background characteristics were often badly skewed (e.g., number of years of teaching in Table 2.7), and information was available for more background variables than teachers. We therefore decided to use rather crude data reduction techniques in order to break these data down to a minimal set of contrasts among teachers. To this end, we rank-ordered the 12 teachers on each of the first 17 variables in Table 2.7, then we computed a matrix of squared distances among all pairs of teachers by summing their squared rank-differences across all 17 variables. We then factored the matrix of scalar products found by taking - 1/2 the double centered matrix of squared euclidean distances among teachers (Torgerson, 1958).

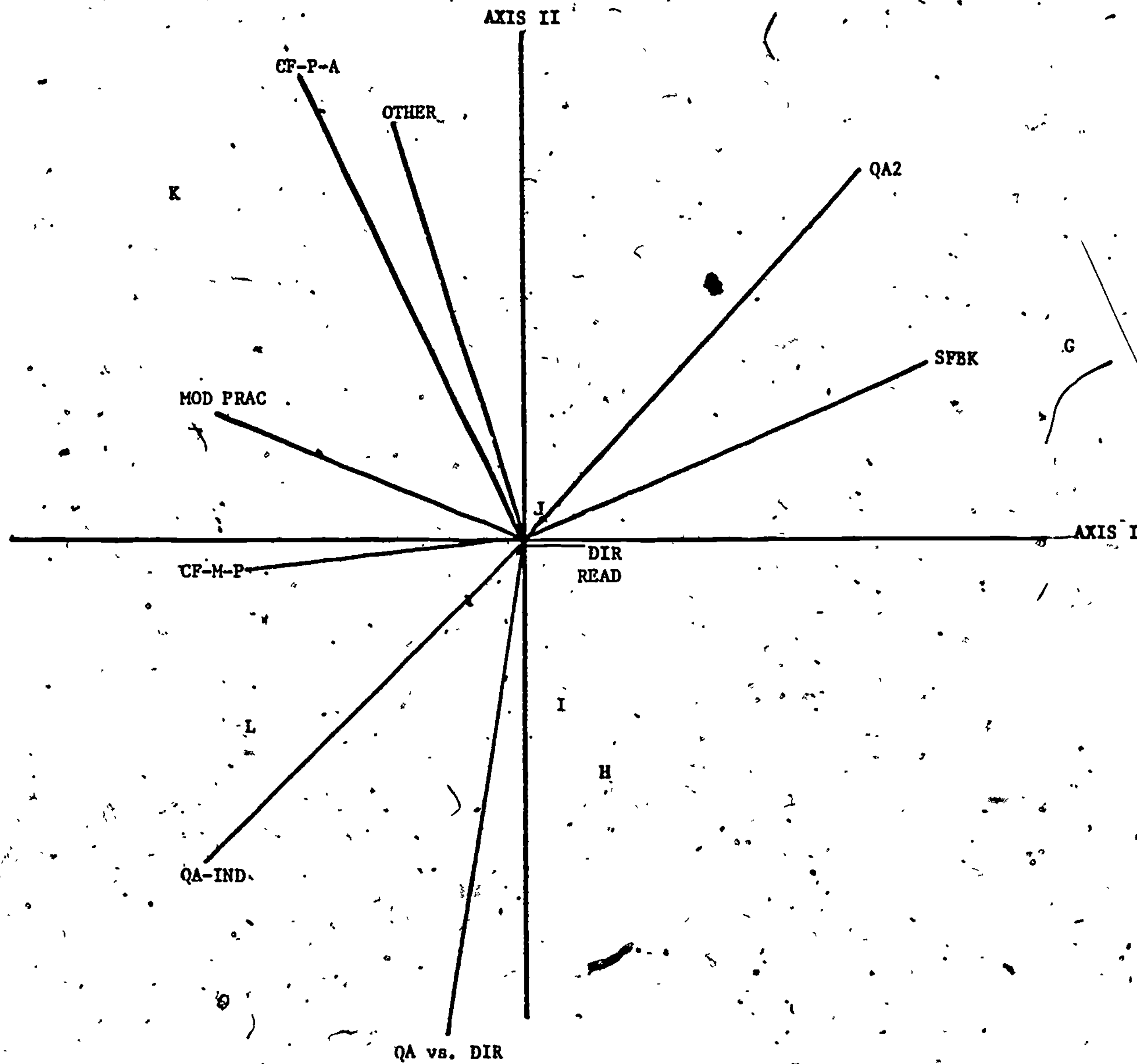

FIGURE 3.7

Plot of Night-School Classroom Centroids and Classroom Interaction Factors in the Space Defined by the Two Largest Canonical Discriminant Axes

The first two eigenvalues of the scalar products matrix among teachers were very prominent, suggesting a dimensionality of two. Teacher scores on the first two principal axes were correlated with the rank-ordered variables and the resulting matrix was rotated to orthogonal simple structure. The outcome of this crude approach to metric multidimensional scaling is presented in Figure 3.8. In this figure teachers are represented by the same letters used in earlier figures (A-F for day school, G-L for night school) and their background characteristics are represented as vectors.

Despite the approximate nature of the analyses under discussion, Figure 3.8 reveals interesting information about the teachers included in this study. The first thing to note is that the horizontal axis is aligned with the variable "percent audio-lingual" and it marks a rather strong contrast between day-school and night-school teachers. Only one day-school teacher, J, is placed toward the "silent way" (left) pole of the horizontal axis. As for the vertical axis, it is clearly an indication of educational level which differentiates teachers within both the day and night-school samples. It is noteworthy that "years of teaching" is closely aligned with this vertical axis but that amount of experience teaching English as a second language to adults ("ESL day adult") is aligned with the "silent way" pole of the horizontal axis. In other words, the more experience a teacher has had teaching English as a second language to adults, the more likely he or she is to use or approximate "the silent way." The more experience a teacher has had in teaching in the elementary or secondary schools, the more likely they are to use the audio-lingual method or an approximation to it.

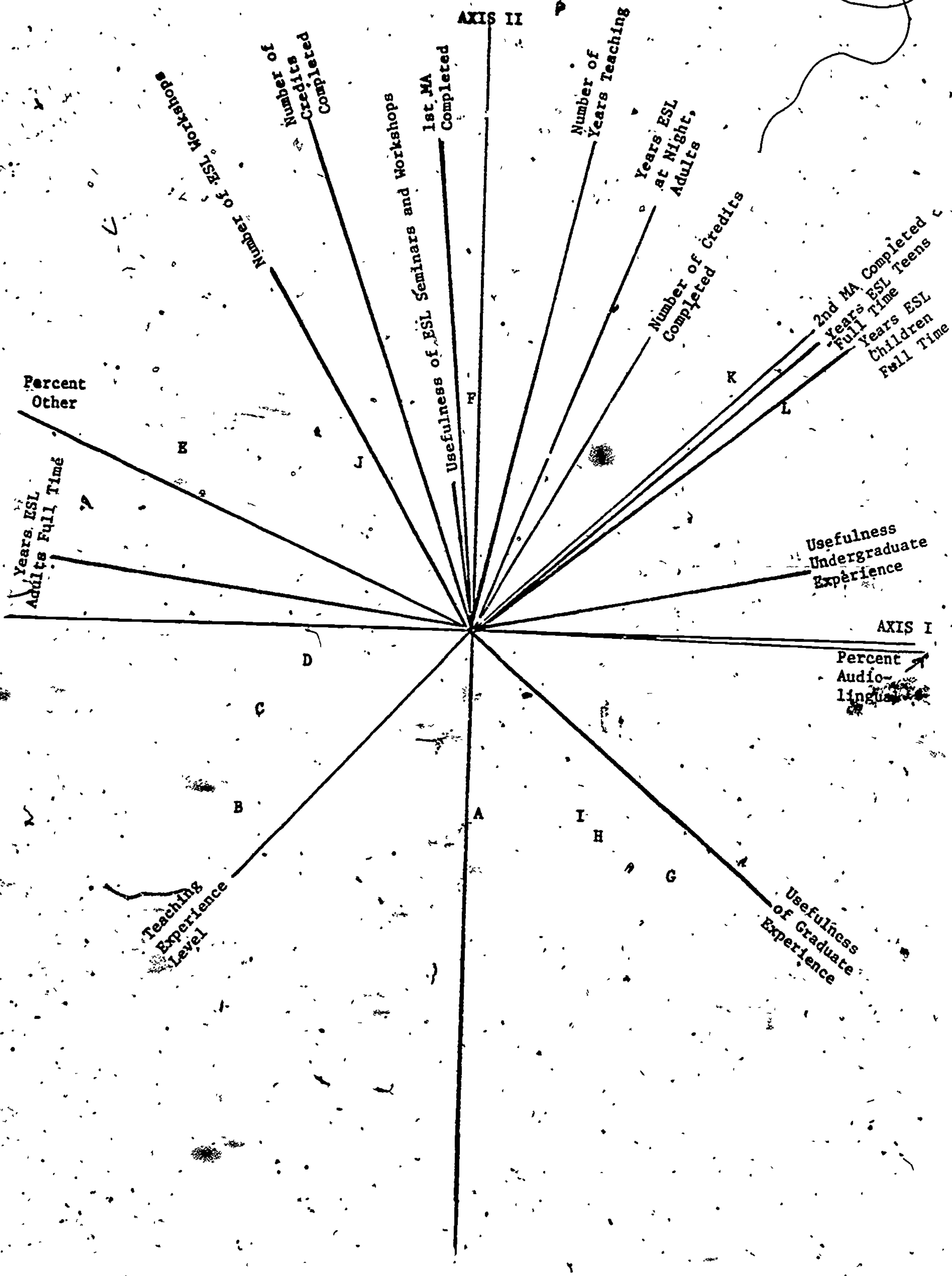

Figure 3.8

Multidimensional Scaling of ESL Teachers on the Basis of Similarity in Their Background Characteristics; Day School Teachers (A-F), Night School Teachers (G-L).

It seems that the lower the level (from adult to child) at which an individual has had experience teaching, the more likely they are to be educated, experienced, and a practitioner of the audiolingual method (e.g., teachers K and L).

Comparing Figure 3.8 to Figure 3.2, there appears to be some association between classroom interaction patterns (teacher performance, if you will) and teacher background characteristics. Notice that the day-school teachers fall into roughly comparable circular patterns (from A through F) in Figures 3.8 and 3.2. As usual for the night school, however, the comparison of Figures 3.8 and 3.7 proves confusing.

An objective way to compare the subjectively similar placement of day-school teachers with respect to their teaching behaviors (Figure 3.2) and their background characteristics (Figure 3.8) has recently been brought to light by Schultz and Hubert (1976). It is possible to compute a quantitative measure of agreement between the two sets of day-school inter-teacher distances represented by Figures 3.2 and 3.8; a kind of coefficient of proportionality called Γ in the literature or quadratic assignment where it was developed. Once this coefficient is available it is possible to assess its probability on the basis of random reorderings (e.g., mislabelings) of the teachers in either figure. If the correct labeling gives an improbably high index of agreement between the two ways of getting at similarity among teachers, then we have objective evidence that the configurations conform to one another. Based upon the distances between teacher centroids in Figure 3.2 (combining classrooms taught by the same teacher) and in Figure 3.8, we get $\Gamma = 566.5$. Since the mean Γ coefficient from all possible permutations of teacher labels in one figure while holding the other fixed is $\Gamma = 496.2$, with a standard deviation of 36.3, the degree of agreement shown with the correct labeling seems unlikely to have occurred by chance ($p < .21$ from Cantelli bound which is conservative).

REFERENCES

Burt, C. Scale analysis and factor analysis: comments on Dr. Guttman's paper. British Journal of Statistical Psychology, 1953, 6, 5-24.

Cliff, N. & Krus, D. J. Interpretation of canonical analysis: rotated vs. unrotated solutions. Psychometrika, 41, 1976, 35-42.

Guttman, L. The quantification of a class of attributes. In P. Horst, et. al, The Prediction of Personal Adjustment. New York: Social Science Research Council, 1941, 321-347.

Lazarsfeld, P. F. The logical and mathematical foundations of latent structure analysis. In S. Stouffer et. al., Measurement and Prediction. Princeton: Princeton University Press, 1950.

McDonald, R. P. The common factor analysis of multicategory data. British Journal of Mathematical and Statistical Psychology, 1969, 22, 1965-1975.

McDonald, R. P. Three common factor models for groups of variables. Psychometrika, 35, 1970, 11-128.

Schultz, J. & Hubert, L. A nonparametric test for the correspondence between two proximity matrices. Journal of Educational Statistics, 1976, 1, pp. 59-67.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.

Yates, A. A weighted least squares approach to maximum likelihood, alpha, and traditional common factor analysis. Psychometric Society Meeting, St. Louis, 1971.

Yates, A. An Oblique Analytic Transformation Technique for Primary Factor Pattern Simplification which Permits Factoral Complexity in Exploratory Analyses. Psychometric Society Meeting, Palo Alto, 1974.

CHAPTER FOUR

THE RELATION OF TEACHING PERFORMANCE TO STUDENT LEARNING

STRATEGY FOR ANALYSIS OF STUDENT ACHIEVEMENT

The major problem to be solved in the analysis of the student achievement data was to determine how we could account for variation in posttest scores. There are three domains of variables which are potential predictors of posttest scores: background characteristics of the students, students' initial level of proficiency as measured by the pretests, and classroom experiences.

Variables describing the students' backgrounds are worth considering as potential predictors of final achievement because these measures may be indirect indicators of aptitude for learning, of academic skill, or of prior achievement of proficiency in English. Obviously, the students' initial proficiency (the second domain mentioned above) may be associated with their final achievement status. Pre- and post-instruction scores on the same test are usually highly correlated with each other because the experience acquired in the interval which separates these measures does not greatly alter the relative ordering of students with respect to their abilities; the latter have, of course, been built up over an entire lifetime of experiences, for which the background measures are indicators or proxies. Nevertheless, the relationship between pre- and posttest scores can be altered somewhat through the influence of intervening events, including classroom interaction experiences. This is our third domain of predictor variables and is of most interest in this study because it is the only domain over which some degree of control can be exercised.

Other relevant experiences which might intervene between pre- and posttesting (such as use of English at home, on the job, and in the community) unfortunately could not be objectively measured in this study. The possibility of controlling these extracurricular sources of experience is slight, in any case, but it would be useful from a theoretical point of view to take them into account. The best we can do at this point, however, is to bear in mind that certain "background" characteristics might serve as proxies for sustained extra-curricular experiences (e.g., occupational level for the need to speak English in the workplace; length of time in the United States for assimilation into an ethnic community).

Given these three domains of predictor variables, which are organized in an obvious temporal sequence (background experiences ⟶ pretest performance ⟶ classroom instruction), the analytic problem is to find how the information can be most parsimoniously combined to predict final achievement. Multiple linear regression can be applied to this task, since the squared multiple correlation between a set of predictors and a criterion variable (e.g., posttest score) indicates what proportion of the observed variance in the criterion can be accounted for by prior knowledge about the given set of predictors.

In this case we have three successive sets of predictors which can be taken into account in a logical sequence in order to attribute as much influence as possible to those experiences which have priority in time before entertaining more complex (and more recent) determination of final achievement. Background characteristics must be taken into account first because these variables are direct measures of or proxies for educational attainment, competence to cope with the processes of schooling, motivation and aptitude. If such variables account for most of the variance in final scores, then classroom experiences can have little differential influence on final status; this is not to say that large pre- to posttest gains could not have occurred, only that they are not likely to have changed the relative ordering of students in terms of achievement.

Next we must consider how much initial status in terms of pretest scores adds to the prediction of outcomes beyond what we have been able to learn from a knowledge of background characteristics alone. A related issue here is how well initial status *per se* can be predicted from background characteristics. This analysis tells us how adequate and useful our information about background experiences is in the first place. Had we in our possession *all* of the background information truly relevant to achievement, it is clear that we would not have to measure initial proficiency at all. Enough background material would be available to *predict* pretest scores accurately within the limits of their own unreliabilities. Of course, this ideal state of affairs cannot generally be approached in practice because of limited sampling of background characteristics, measurement errors, nonlinearities, etc.

The final step in the regression analysis is to assess the unique contribution of classroom experiences in the prediction of final achievement, above and beyond any predictive utility of initial status and background experiences. We also must determine whether different forms of classroom interaction appear to have different effects upon achievement. The purpose of relating classroom interaction to achievement is to find out how much this information adds to the prediction of achievement and to identify any components of classroom interaction which can be hypothesized to have an impact upon specific forms of achievement.

In sum, the strategy is first to see how well student achievement can be predicted by the background characteristics of the students, then to determine how much the prediction of posttest achievement can be improved by adding information about pretest performance on all of the fall achievement measures. The reader will see that both of these sets of variables do predict final student achievement, as might be expected. The essence of the analysis, however, is to find out how much of the variance in final student performance can be predicted from classroom experiences once prior experience and achievement are taken into account.

Notice that the classroom experience variables differ fundamentally from all of the other variables in the analysis in that they are not individual measures, but apply to all individuals in each classroom equally. From the point of view of this study, these variables are the "independent" variables whose possible effect upon final achievement we are most interested in learning. These variables take the form of contrasts among classrooms in terms of observed patterns of teacher-student interaction, the four transformed canonical variates developed and discussed in Chapter Three. Every effort was made to arrive at a small set of independent variables which are reliable and not too highly intercorrelated, in line with the requirements of the multiple regression model. On the other hand, the background and pretest variables are more error prone; but they only play the role of covariates to adjust for preexisting differences among classrooms in the final analysis. Moreover, these data are available for every individual, so stable estimates of the required regression parameters can be obtained even with fallible measures.

Predicting Fall and Spring Student Achievement from Student Background Characteristics

It is conceivable that student achievement can be accurately predicted from knowledge only of certain characteristics associated with students' backgrounds. These characteristics are proxies for previous educational achievement, for sociocultural conditions which promote educational attainment, and for motivational factors that contribute to greater educational achievement. It is possible that such factors account almost totally for differences in achievement because the students in effect become their own teachers and are not highly dependent upon the specific teaching performances of the teachers to whom they are exposed.

In Table 2.5 we saw many significant zero-order correlations between student characteristics and student achievement scores, both in the fall and in the spring. The background and achievement domains are clearly related at the level of individual pairs of variables, so it is in order to see how background information can be optimally combined (weighted) to predict individual achievement scores and to determine how effective this prediction can be when simultaneous use is made of all available background information.

Each achievement test score was predicted from the background characteristics using multiple linear regression. The results of these analyses for the day-school sample are summarized in Table 4.1. In that table the labels across the tops of the columns designate the fall and spring achievement scores; note that logarithmic transformations of the oral proficiency scores were predicted.[1] The rows of the table list particular background characteristics used in the

[1] All of the Oral Proficiency Test scores were subjected to a logarithmic transformation before being entered in the regression analyses. The logarithmic transformation was found to be necessary because the original Proficiency Test score variances within different classrooms were not homogeneous; i.e., the variances were highly correlated with classroom means. What this suggests is that, rather than responding in an additive fashion to any factors which might differentiate classrooms, oral proficiency is in a sense multiplied by such effects. Not only does the logarithmic transformation stabilize variances, it increases the linear predictability of proficiency scores from background, pretest, and classroom interaction variables.

regression analyses.[2] The number in any cell of this matrix is the standardized regression weight for a particular background variable (row) as a predictor of a given achievement score (column). These entries complement the zero-order correlations already given in Table 2.5. The former show the direct contribution of variation in each background characteristic to achievement, whereas the entries in Table 2.5 do not take into account the fact that the background characteristics are correlated among themselves (i.e., they are redundant).

The line in Table 4.1 labeled R^2 gives the squared multiple correlation of each achievement test score with all of the background variables, taken simultaneously. These numbers indicate the proportion of the variance in each achievement score predicted by the complete set of background characteristic variables; for example, the background characteristic variables account for 26 percent (.26) of the variance in the fall Decoding 1 scores; for 13 percent (.13) in the fall Decoding 2 scores; for 44 percent (.44) of the variance in the John Test scores, and so on.

The F-ratios presented in Table 4.1 (degrees of freedom in parentheses) provide a statistical test of the null hypothesis of no predictability of each dependent variable (Achievement Test score) from the set of independent variables (background variables). In other words, is the mean score on each test the best available estimator of how any given student will perform, or can we learn more about that performance by taking his or her background characteristics into account? The probability levels associated with these statistics are coded with asterisks.

[2]The background characteristic "high school diploma" was not included in these regression analyses because it was found to be highly collinear with "number of years of education."

TABLE 4.1

Standardized Regression Weights, Squared Multiple Correlations and Statistical Tests for Background Characteristics as Predictors of Pretest and Posttest Scores; Day-School Sample

	Pretests					Posttests					
	D_1-F	D_2-F	L-F	J-F	M-F	D_1-S	D_2-S	L-S	Log_{10} P-CT	Log_{10} P-CP	Log_{10} P-ST
Sex	.03	-.01	-.02	.03	-.01	-.01	-.09	-.14	.08	.11	.06
Age	-.32**	-.22*	-.31***	-.16	.03	-.26*	-.06	-.29**	-.17	-.20*	-.20*
Time in U. S.	-.12	-.08	.27**	.22*	-.04	-.14	-.07	.10	.04	.05	.04
Former Job	.22*	.10	.21*	.13	.20*	.27*	.15	.13	.23*	.28**	.26*
Job in U. S.	.00	-.02	.00	.00	.08	.19*	.12	.13	-.01	.02	.01
C-Origin	-.03	-.17	-.01	-.11	-.16	-.05	.07	-.03	-.02	-.07	-.04
Years Education	.05	-.02	.14	-.01	.01	.07	.09	.09	.03	.02	.03
Former Eng.	.11	.22*	.22*	.27**	.11	-.05	.04	.09	.07	.07	.10
Eng. in U. S.	.18*	.11	.38***	.52***	.37***	.11	-.05	.24*	.24*	.34***	.27**
R^2	.26	.13	.47	.44	.29	.25	.08	.25	.20	.33	.24
F (9, 71)	2.76**	1.14	7.13***	6.25***	3.23**	2.66*	.73	2.58*	1.94	3.82***	2.51*

*p < .05
**p < .01
***p < .001

Labels Code:

D_1: Decoding, Part 1
D_2: Decoding, Part 2
L : Literacy Test
J : John Test
M : Morano Test
P : Oral Proficiency Test

CT: Correctness Score
CP: Comprehension Score
ST: Structure Score
F : Fall
S : Spring

A prediction equation could be written with the regression weights in any given column of Table 4.1 as the coefficients of the background variables. The actual standard scores of a particular student for each of these variables would be entered in this equation to predict his or her standardized score on the corresponding achievement test. It can be seen that variables with larger weights have more influence on the predicted outcome. The following example will illustrate this concept. We will use the numbers in the column labeled P-CT, the correctness score on the Oral Proficiency Test which was taken in the spring. The prediction equation for this score after logarithmic transformation is:

$$(1) \quad Z_{\log_{10}(P\text{-}CT)} = .08\,Z_S - .17\,Z_A + .04\,Z_T + .23\,Z_{FJ} - .01\,Z_J - .02\,Z_{CO} + .03\,Z_{YE} + .07\,Z_{FE} + .24\,Z_E$$

As will be seen shortly, two of the above weights are significantly different from zero in the statistical sense. The three which are underlined in the P-CT column of Table 4.1 are large enough to give some feeling for which background factors might affect correctness scores (P-CT). They are age, which has a negative weight (- .17), status of job in former country (.23) and amount of English taken in the U. S. (.24). In other words, students who are younger, had better jobs in their former country, and have taken more course work to learn English after arriving in this country will have higher predicted correctness scores.

There are two other features of the information in Table 4.1 which should be noted. Those regression coefficients which are significantly different from zero in a statistical sense are marked with asterisks. In addition to significance in a statistical sense, however, it is important to consider how much each predictor variable contributes to the total predictability signified by R^2, without regard to sample size.[3] The regression weights for those variables which contribute at least .01 to the magnitude of the squared multiple correlation, R^2, have therefore been underlined in Table 4.1 (i.e., these variables account for at least one percent of the total variance in the Achievement Test score).

Implications of the Regressions of Achievement Test Scores on Background Characteristics

The R^2 for prediction equation (1) is rather low (.20). This equation is not a very accurate predictor of Oral Proficiency correctness scores. We may still speculate, however, that students with three background characteristics (Youth, Good Former Job, Eng. in U. S.) are likely to have higher correctness scores as long as we remember that the prediction equation will be highly in error in any given instance. By way of contrast, consider the comprehension scores (P-CP). The prediction equation in this case is likely to be a little more accurate (R^2 = .33 for comprehension vs. .20 for correctness). In that equation amount of English studied in the U. S. has the largest weight. Three other characteristics are worth noting: age, which again has a negative weight (- .20), status of job in former country (.28), and amount of English in the U. S. (.34).

The similarity in the patterns of regression weights for all three columns relating to the Oral Proficiency Test is striking. From Table 2.5

---

[3] That some weights do not differ significantly from zero in a statistical sense does not mean that we can accept the null hypothesis and conclude that background characteristics do not relate to scores; all that is implied is that either our sample size is too small or our variables are too error prone, skewed, etc. to reject the hypothesis of no relationship on the basis of this one study.

we can see that these three alternative scoring systems for the same set of items are very highly correlated (before transformation), as might be expected. The scoring for correctness is most stringent, whereas that for comprehension is least stringent--it is not unlikely that the lower predictability of the more stringent scores is due in part to a restriction in their ranges because few students got many items strictly correct. The logarithmic transformation helped to overcome skewing of proficiency test score distributions due to the stringency of scoring and increased their predictability. Consequently, the three alternative scoring methods yielded results which are even more highly equivalent after transformation than is indicated by the correlations given in Table 2.5.

Background characteristics predict comprehension scores best (R^2 = .33), structure scores less well (R^2 = .24) and correctness scores most inaccurately. A reasonable hypothesis from these data, however, is that those students will be more proficient in spoken English by the end of the year who are younger, had a higher status job in their former country, and have taken more English courses or programs since coming to the United States.

It should be remembered that predictive background characteristics do not necessarily "cause" higher or lower scores. Age, for example, does not necessarily make a person less proficient. But age is associated with several other characteristics which, taken together, may give us some idea about why age is a negative predictor of proficiency. Older students are more likely to have a low status job in this country, although they may have had a higher than average status job before arriving here. They are more likely to have had course work in English in the previous country than in the U. S., and are more likely to have arrived from Cuba and the Carribean than from Western Europe.

Amount of English studied in this country is more likely to be a direct influence on achieving proficiency. But, again, students who have had more work in English here are also younger, have arrived more recently, and have a history of higher level employment both here and abroad. They probably have both more need to speak English and more opportunity to do so. Thus a combination of previously acquired proficiency and opportunity to speak English is likely to facilitate the progress of these students during their time at the Adult Learning Center.

Differential Prediction of Fall and Spring Achievement from Student Background Characteristics

Of the five squared multiple correlations for predicting fall tests from background variables, all but one are statistically significant at the .01 level. Note also that the multiple correlations are generally lower in magnitude and of lower statistical significance for the spring tests. These results suggest that the role played by background experiences in achievement is substantially diminished over the course of the year, as new experience with English as a second language is acquired through classroom instruction and extracurricular activities. That students gain substantially in their average test scores has already been seen in Table 2.3.

That background characteristics do help predict performance in the fall (R^2 = .44 for the John Test; R^2 = .47 for the Literacy Test) but that their predictive utility is diminished in the spring (R^2 = .33 for Proficiency-Comprehension; R^2 = .25 for Literacy) is the most important piece of information to note from Table 4.1. Only the comprehension score on the spring Proficiency Test retains a highly significant relationship to background characteristics. This fall-spring distinction suggests that something beyond the students' background experiences prior to instruction accounts for their final spring achievement. The next question is whether or not we can identify any component of classroom interaction which accounts for the change in achievement.

There are a number of interesting relations apparent in Table 4.1 beyond those mentioned already. Consider the variable, amount of English studied in the United States; one might expect that the amount of English any student has been exposed to or has learned will be related to his or her measured achievement in the fall. This prediction is borne out as can be seen by reading across in the line labeled "Eng in U. S." Note that the Literacy Test, the John Test, and the Morano Test scores in the fall are all predicted by this variable. This variable also predicts the Literacy Test scores and the scores on the Oral Proficiency Test in the spring, as noted earlier. It accounts for not less than one percent of the total variance on all of the tests except for Decoding 2 in the spring. Obviously, the amount of English learned or studied while in the United States is an important predictor both of initial status and of most aspects of final achievement. In fact, one might be suspicious of the validity or reliability of any achievement measure which is not so predicted.

Consider now the amount of English studied in the student's former country. One might suspect that English training in the country of origin is useful, but may or may not be as qualitatively good as training acquired in the United States. But this variable also predicts some test scores, as can be seen by reading across from "Former Eng" in Table 4.1. The regression coefficients for the Decoding 2, Literacy, and John Tests in the fall are all significantly different from zero. Notice, however, that the amount of English studied in the country of origin does not predict spring scores.

The difference in patterns of prediction for English studied in the U. S. vs. English studied in the former country is intriguing. It suggests that those individuals who have studied their English in the former country are more responsive to language experiences acquired during the interval between pre- and posttesting than are those individuals who have studied their English in the U. S. That outcome is not surprising, however, when it is found that amount of English studied in the previous country is our best available indicator of the occupational level acquired in the United States ($r = .29$). In this connection, note from Table 4.1 that "Job in the U. S." is one "background" characteristic which apparently has a higher relationship to certain posttest scores than to any pretest score. Perhaps those individuals who have attained high-level employment in the U. S. have done so partially by virtue of their prior English-language training in the previous country. This prior training also goes along with a higher overall level of education and employment in the former country, so it is not surprising that these individuals are highly motivated to make use of their current English-language experience, both in the classroom and on the job.

Note that the regression weights for "age" are generally negative in sign. In this case we see weights which are significantly different from zero in predicting both fall and spring achievement scores, just as for former employment level and English studied in the United States.

It is obvious from this analysis that a number of background characteristics are significant predictors of fall achievement and a few of them are of importance for predicting spring scores. Most of the multiple correlations are significant when all of the predictors are used in the regression analysis, regardless of whether fall or spring achievement is the criterion.

Predicting Posttest Scores Using Pretest Scores in Addition to Background Characteristics

The next step in the analysis was to predict the posttest scores using information about the pretest scores in addition to information about background characteristics. The results of these analyses for the day-school sample are presented in Table 4.2. This table is read in much the same way as Table 4.1. Across the top of the columns are listed the posttest scores, beginning with the two Decoding scores, then the Literacy score, and the last three columns contain logarithmic transformations of the three scores on the Oral Proficiency Test. The predicted scores used in this analysis were all scores taken from tests administered in the spring.

Down the left-hand column of the middle section of Table 4.2 are listed the labels for the pretest scores; included are the two parts of the Decoding Test, the Literacy Test, the John Test, and the Morano Test. The first line of this table gives the squared multiple correlations resulting from prediction of posttest scores from background characteristics; these numbers are repeated from the right-hand portion of Table 4.1. The entries in the line labeled, "R^2 with Addition of Pretests," can be compared to the corresponding entries in the first line. Note under D_1-S that R^2 when only background characteristics are used is .25, but R^2 increases to .60 when the pretest information

TABLE 4.2

Standardized Regression Weights, Squared Multiple Correlations and Statistical Tests for Posttest Scores Predicted from Pretest Scores Adjusted for Background Characteristics; Day-School Sample

	D_1-S	D_2-S	L-S	P-CT	P-CP	P-ST
R^2 from Background	.25	.08	.25	.20	.33	.24
D_1-F	.64***	.24	.16	.27**	.29***	.25**
D_2-F	.08	.15	.10	-.01	.09	.01
L-F	-.07	.07	.47*	.09	.02	.07
J-F	.00	-.14	.39*	.49***	.60***	.52***
M-R	.07	.03	-.04	.09	-.06	.11
R^2 with Addition of Pretests	.60	.19	.61	.72	.82	.76
F (5, 66) Test of Information Increase	11.17***	1.72	12.23***	24.83***	35.86***	27.85***

*$p < .05$

**$p < .01$

***$p < .001$

is added. In other words, if only background characteristics are used to predict posttest scores, then 25 percent of the variance in the Decoding 1 posttest is accounted for. But an additional 35 percent is accounted for if information from the pretest scores is added. Note, for example, that the R^2 for predicting transformed Oral Proficiency comprehension scores from background characteristics is only .33, but with the addition of pretest information it becomes .82. A similar pattern is apparent for the other two scores of the Oral Proficiency Test. Thus, the students' initial proficiency plus some information about their background characteristics accounts for a substantial proportion of the variance in the comprehension scores. In summary, to understand how much the prediction is improved by adding pretest data to background information, simply compare the R^2 in the first line with the R^2 in the third portion of the table.

It should be noted that most of the squared multiple correlations for background characteristics plus pretest scores are substantial (.60 to .82). Moreover, the increase in all of these R^2's except for part two of the Decoding Test are highly significant when compared to prediction from background characteristics alone.

Except for part two of the Decoding Test, all of the squared multiple correlations of posttest scores with background and pretest variables are at or above .60. From Table 2.3 we can see that the posttest reliabilities for all except Decoding 2 range from .88 to .94 (Decoding 2 has lower internal consistency, yielding a reliability of .77). If we take these reliabilities to mean that roughly 90 percent of the variance in most posttest scores is reliable, then it can be said that over 60/90 or two-thirds of that variance can be accounted for by the background and pretest variables which we have measured. This leaves up to one-third of the reliable posttest variance which might be accounted for by classroom experiences. However, we must keep in mind that we do not have separate measures of classroom experience for each of the students in any classroom--only contrasts between classrooms. This means that we cannot expect to increase the accuracy of our prediction to the limit of the reliability of posttest measures on the basis of classroom experiences. Even if classroom experiences are highly influential, their effect cannot be detected unless there is sufficient homogenity among students within classrooms with respect to these experiences. We initially planned to use attendance as a means of assessing individual differences in classroom experiences, but the data on attendance did not prove helpful because of little documented variation in attendance.[4]

The consequence of all this is that our concern in assessing the impact of classroom experiences upon final achievement scores must focus primarily upon the results of significance tests (comparing between vs. within classroom variation) rather than upon any absolute increases obtained in R^2.

Before moving to the assessment of the impact of classroom interaction contrasts upon final achievement beyond that predicted from background and pretest variables, it is in order to notice several features of Table 4.2. As far as achievement is concerned, the only predictor of Decoding 1 in the spring is the corresponding pretest score. Spring Literacy scores, however, are seen to relate both to the Literacy Test and to the John Test in the

[4]Students who fail to attend regularly are not continued in the program.

fall. The spring Proficiency measures, which are highly intercorrelated as we mentioned earlier, are largely a function of John Test performance in the fall but also appear to reflect Decoding 1 performance in the fall.

Posttest Scores Predicted from Classroom Interaction, with Adjustment for Pretest Scores and Background Characteristics

Table 4.3 presents information from the regression analyses for the day-school sample in which posttest scores were predicted from classroom interaction contrasts, pretest scores, and background characteristics. This table is organized in the same way as Table 4.2.

The first line of Table 4.3 gives R^2 from the regression of posttest scores on background characteristics and pretest scores, as seen already in Table 4.2. Next are four lines with Roman numerals I through IV; within each line can be found the standardized regression weight of the corresponding transformed canonical variate (see Chapter Three) in the prediction equation for the achievement posttest indicated at the head of the column.

The sixth line in Table 4.3 gives R^2 once the four transformed canonical variate classroom interaction contrasts have been included in each prediction equation. Again, the lines containing the initial and final R^2 values can be compared. From this comparison we can see that the addition of classroom interaction contrasts adds little in an absolute sense to the accuracy of prediction of posttest scores (increases in R^2 vary from .04 to .09).

Despite the relatively small absolute increases in multiple prediction equation accuracy obtained by supplementing background and pretest information with classroom interaction information, it can be seen from the F-test results in Table 4.3 that the increases in predictability obtained cannot be attributed to chance alone.

TABLE 4.3

Standardized Regression Weights, Squared Multiple Correlations, and Statistical Significance Tests for Posttest Scores Predicted from Classroom Interaction Contrasts, Adjusted for Student Background Characteristics and Pretest Scores; Day-School Sample

	D_1-S	D_2-S	L-S	$\log_{10}$ P-CT	$\log_{10}$ P-CP	$\log_{10}$ P-ST
R^2 for Background and Pretests	.60	.19	.61	.72	.82	.76
I	-.27	-.29	-.28	.25	.37$^+$	.40$^+$
II	-.22	-.23	-.34*	.28$^+$	.34*	.21
III	.40$^+$	.27	.42$^+$	-.18	-.36$^+$	-.19
IV	.32**	.25*	.27*	.10	.02	.09
R^2 with Addition of Classroom Interaction Contrasts	.65	.28	.66	.76	.85	.81
F (4, 62) Test of Information Increase	2.26$^+$	2.03$^+$	2.35$^+$	2.74*	2.73*	3.91**

$^+p < .10$

$^*p < .05$

$^{**}p < .01$

As we pointed out earlier, the classroom interaction contrasts apply equally to all individuals within each classroom, so great homogeneity of experiences would be required in order for these contrasts to yield a large absolute increase in the accuracy of prediction of posttest achievement. The F-test results relate, however, to relative increases in accuracy of prediction; i.e., given that we have already reduced uncertainty about posttest performance quite considerably from knowledge of background characteristics and pretest performance. These relative increases in the accuracy of prediction are substantial enough, even given our small sample size, to support further exploration of hypotheses to the effect that the classroom interaction experiences measured in this study contribute to final achievement above and beyond all the student background characteristics and initial abilities which we measured.

Interpretation of the standardized regression coefficients in the columns of Table 4.3 requires reference to the earlier interpretations given to the four transformed canonical variates (see Chapter Three). Recall that the first two classroom interaction contrasts reflect broad stylistic differences among teachers. The negative pole of each contrast relates to rather highly structured class or group-oriented activities. The prevalence of "teacher model-student practice" interaction patterns on the negative poles suggests implementation of the audio-lingual method of instruction and seems to be related to teacher background characteristics (see end of Chapter Three).[5]

[5]Possible relationships of the classroom interaction contrasts to teacher background characteristics were inferred from a comparison of Figures 3.2 and 3.8, coupled with a canonical correlation analysis which is not reported here.

The presence of the "Free" and "Other" patterns of classroom interaction on the positive pole of the second transformed axis leads us to think of this variate as a contrast between the silent way and audio-lingual methods. The flexible, individualized, supportive, and encouraging aspects of the positive pole of the first axis suggests implementation of a rather eclectic version of classroom interaction. This eclecticism seems to be characteristic of those teachers who have worked mainly with adults and have little formal training; perhaps they treat the students more as peers than do other teachers and thus rely less upon preconceived notions about "systems" of instruction. The third axis relates most directly to the occurrence of "student-student feedback" during instruction. The fourth axis can be ignored for the moment.

The most striking aspect of the pattern of standardized regression coefficients presented in Table 4.3 is that it appears that those features of classroom interaction which are associated with higher than would otherwise be expected posttest scores on the Oral Proficiency Test (i.e., axes I and II) are associated with lower than would otherwise be expected posttest scores on the Literacy and Decoding Tests (compare the weights in the left-hand columns of Table 4.3 opposite I and II with those in the right-hand columns opposite I and II). If we were to take these coefficients as the basis for formulating hypotheses about the effectiveness of individualized instruction vs. group instruction, the silent way vs. the audio-lingual method, an open and supportive vs. a highly organized and directive classroom climate, etc. (however one sees fit to interpret what is in common to the first two transformed canonical variates), then it would be in order to expect the former (individualized, supportive, silent way) to lead to increased oral proficiency while expecting the latter (grouped, directive, audio-lingual) to lead to increased literacy/decoding skills. Moreover, there is the implication that what might lead to higher than would otherwise be expected final achievement in the oral proficiency domain might lead to lower than would otherwise be expected final achievement in the literacy/decoding domain.

It appears, therefore, that any hypotheses based upon the achievement test regression weights on axes I and II in Table 4.3 must involve a trade-off in terms of the achievement goals to be stressed--what appears to be incremental for oral proficiency appears to be detrimental for literacy/decoding, and vice versa. The same can be said for axis III, "student-student feedback," since its occurrence in a classroom is predictive of lower than might otherwise be expected oral proficiency (comprehension mainly, as seems quite plausible upon due reflection), but higher than might otherwise be expected literacy/decoding.

As for that ubiquitous bipolar contrast in classroom interaction patterns (whether at the level of individual episodes, sessions, classrooms or teachers), "teacher direct-student read and/or ask question" vs. "teacher question-student answer," we again see it playing an important role. In this case there is no apparent trade-off, however. On the contrary, it is suggested quite clearly that those classrooms in which students are directed to read are also those classrooms in which we can predict that the students will achieve higher than would otherwise be expected on the Literacy and Decoding posttests. The contrast seems to bear no real relationship to oral proficiency, suggesting that nothing is to be lost and literacy/decoding skills may well be gained through classroom implementation of the "teacher direct-student read and/or ask questions" paradigm (line labeled IV in Table 4.3).

THE ANALYSIS OF DIFFERENCES IN STUDENT ACHIEVEMENT BY CLASSROOMS

Notice that we have been rather liberal about discussing the possible implications of the regression weights in Table 4.3, as far as stating hypotheses about the effects of teaching styles is concerned. Furthermore, we have relied mainly upon the signs of the weights without regard to their magnitudes or their departures from zero in a statistical sense. This approach is perfectly acceptable, as long as we indulge only in the generation of hypotheses, and not in the acceptance or rejection of such.

It is not possible to confirm causal inferences in an exploratory survey of naturally existing populations (e.g., classrooms), even where statistical significance of relationships is rigidly adhered to as a basis for entertaining hypotheses. We can only rely upon statistical procedures to help clarify apparent relationships and to arrive at plausible models of a quasi-causal nature which might later be subjected to experimental refutation.

One basis for evaluating a mathematical model is the degree to which it "saves the appearances"; i.e., the extent to which it yields a closer fit to the manifest data than other models which are comparable in complexity (e.g., in the number of parameters to be estimated from the data, such as regression coefficients). A closely related issue is the question of how much improvement in the fit of the model to observed data can be achieved through relaxation of restrictions on parameters in the model which have been fixed a priori, and thus are not estimated from the data. This is the actual basis of the strategy which we have employed so far in comparing regression analyses; i.e., we proceeded from the simplest model, in which only the grand mean was used as an estimator of posttest performance, through successively more complicated models. In going beyond estimating mean posttest performance alone, we first allowed the regression coefficients for background characteristics to depart from their a priori values of zero, then the coefficients for pretests, and, finally, the coefficients for the four classroom interaction contrasts.

In each case of successive relaxation in a priori restrictions upon regression coefficients, the F-test of increase in information indicated whether or not enough gain in accuracy of prediction (fit of the mathematical model to observed data) was achieved to justify the associated complication in the model.

Predicting Posttest Scores Using All Possible Contrasts Among Classrooms With Adjustment for Student Background Characteristics and Pretest Scores

We can now carry the model relaxation process one step further than was finally done with the data as presented in Table 4.3, by allowing parameter estimates for any possible differences between classrooms or teachers. Thus we no longer restrict ourselves to the four classroom contrasts derived from observations of teacher and student interactions over the course of the study. If the expanded (relaxed) model results in much improvement in fit to the observed data, then we must conclude that what was most reliably recorded in the classroom observation phase of the study, taken together with pretest and background characteristics, is not adequate to account for variation in posttest performance. What emerged from the factor analyses of classroom interaction data in Chapter Three seems to relate more to variation in the structure of teacher and student interactions than to variation in the content of those interactions. Variation in this content is one possible source of classroom differences in achievement which we may not have assessed adequately.

In the first line of Table 4.4 we have presented again the squared multiple correlations of posttest variables with background and pretest scores. These entries are the same as those seen on the first line of Table 4.3. The second line of Table 4.4 contains comparable R^2 values. These R^2's were obtained by entering a complete set of 13 contrasts (one for each degree of freedom) among classrooms into the prediction equation. The F-tests of information-increase associated with these R^2 values are all statistically significant, indicating that there are real differences in student posttest achievement among the various classrooms in the study, even after adjusting classroom means for preexisting differences in student background characteristics and pretest achievement (assuming homogeneity of within-classroom regression planes).

We have already seen indirect evidence that the original four classroom interaction contrasts account for some of the differences in adjusted classroom means on the posttest achievement measures (Table 4.2). Now we are in a position, however, to partition adjusted posttest variation among classrooms into one component which can be attributed to the four classroom interaction contrasts and another component which must be attributed to differences among classrooms which we failed to measure reliably in the observation phase of the study.

The last four lines in Table 4.4 are addressed to this issue. They contain information about the separate contributions of our original four

TABLE 4.4

Regression Analysis Results when Possible Classroom Effects are Entered, Compared to Results when only Four Classroom Interaction Contrasts are Entered; Day-School Sample

	D_1-S	D_2-S	L-S	$\log_{10}$ P-CT	$\log_{10}$ P-CP	$\log_{10}$ P-ST
R^2 from Background and Pretests	.60	.19	.61	.72	.82	.76
R^2 from Background, Pretests and full set of Classroom Contrasts	.73	.48	.92	.82	.88	.86
Overall F (13, 53) test of Adjusted Classroom Means	1.95*	2.21*	14.96***	2.19*	1.95*	2.99**
I	- 9.34	-15.63+	- 1.22	.22	.35	.44
II	4.92	-27.90+	- .30	.70	.73+	.22
III	5.09	26.77*	- 2.93	-.12	-.48	.12
IV	13.94**	9.67+	16.94***	.24	.00	.17
% Reduction in Residual Sum of Squares per df, Canonical Classroom Contrasts	10.0	8.3	2.6	5.7	6.5	6.9
F (4, 53) Contribution of Canonical Contrasts	2.53*	2.38+	5.10**	1.62	1.64	2.70*
% Reduction in Residual Sum of Squares per df, Remaining Classroom Contrasts	6.7	7.4	9.9	8.6	8.2	8.0
F (9, 53) Contribution of Remaining Contrasts	1.70	2.15*	17.99***	1.80+	1.51	2.26*

+ $p < .10$
* $p < .05$
** $p < .01$
*** $p < .001$

classroom interaction contrasts (transformed canonical variates), as well as the nine possible remaining classroom contrasts, to variation in adjusted classroom means on the respective posttests.

Information in this section of the table has been presented in two different ways in order to help the reader get a feeling for the relative sizes of the effects involved as well as their statistical significance. The "mean percentage reduction in residual sum of squares per degree of freedom" value for each collection of contrasts gives some feeling for the size of effects involved; for example, $4 \times 10\% + 9 \times 6.7\% \cong 100\%$ of the reduction in sum of squares for the D_1-S posttest. The associated F-tests indicate whether the additional parameters estimated can be justified on the basis of improved fit of the model to the data.

From the last four lines of Table 4.4 we can learn a good deal about the possible bases of differences among classrooms on posttest performance, beyond what is accounted for by the background characteristics and pretest performance of the students involved. Only for the Decoding 1 posttest is it safe to conclude that the four classroom interaction contrasts derived from canonical discriminant function analysis of the classroom interaction data are sufficient to account for all classroom differences in adjusted posttest performance. The regression weights in the mid-section of Table 4.4 reveal that the relationship of Decoding 1 performance to classroom interaction stems primarily from contrast IV, "teacher direct-student read and/or ask question" vs. "teacher question-student answer," as was suggested by the results in Table 4.3.[6]

Notice in studying the regression coefficients of Decoding 1 on classroom interaction contrasts in Table 4.4 the fact that the second contrast (II) has a positive weight. This result contradicts the tentative results given in Table 4.3, but the coefficients in question do not differ significantly from zero in either case. In only one other case does inclusion in the model of additional contrasts among classrooms change the direction of implied effects from those given in Table 4.3; but again, the coefficients involved are non-significant (regression of the Proficiency Structure score on Contrast III, "student-student feedback").

From the last two lines of Table 4.4 we can see that the Literacy, Proficiency Structure, Decoding 2, and, perhaps, Proficiency Correctness

---

[6] The regression coefficients given in Table 4.4 are in raw score form and apply to transformed canonical classroom interaction contrasts which have been normalized in the metric of the classroom cell counts. Seven of the remaining nine contrasts referred to in the table were derived by taking, in successive order, those seven original classroom interaction factors which had maximum residual between classrooms variation <u>after</u> partialling out any effects which could be attributed to the four original transformed canonical variates. The reason for this step was to obtain, in succession, and to as large an extent as possible, stable classroom differences. The final two classroom contrasts were actually teacher contrasts since they were the projections of the day-school teachers (identical projection for different classrooms taught by the same teacher) onto the axes of the multidimensional scaling solution presented in Figure 3.2. The nine new contrasts were transformed to be orthogonal to the original four classroom interaction contrasts as well as to be mutually orthogonal in the metric of the classroom cell counts (in succession, using Gram-Schmidt orthogonalization). Since the choice of these additional contrasts is arbitrary from the point of view of overall fit, several other alternative bases were explored in hopes of gaining insight into the nature of the classroom variation involved. Since nothing much came of these explorations in the way of explaining the relationship of classroom interaction to student performance, the results will not be presented here. Suffice it to say that there was some evidence for reliable differences in classroom interaction factors beyond that measured by the first four canonical variates, but more evidence that teacher background characteristics might contribute to student achievement (experience in adult day-school ESL for proficiency; education for literacy).

posttests all show classroom-to-classroom variation in adjusted means which must be attributed to something beyond the reliably measured aspects of classroom interaction embodied in the four transformed canonical variates. This fact provides justification for looking at mean performance data for the individual classrooms in order to seek an explanation for their differences.

Simple Comparison of Pretest and Posttest Performance in Day-School Classrooms

In order to study the mean performance of individual classrooms we reverted to simple analysis of covariance procedures in which each pretest was the sole covariate for its respective posttest. An exception: in the case of the Oral Proficiency Test, only the correctness score was investigated and both the Decoding 1 and the John Test were considered, individually, as covariates. For the sake of completeness, the Decoding 1 Tests were also studied with simple analysis of covariance, even though there is no evidence in Table 4.4 for classroom differences beyond the four classroom interaction contrasts.

Figures 4.1 through 4.5 are plots of the 14 day-school classroom means on pretest scores (abscissa) vs. posttest scores (ordinate) for the Decoding 1, Decoding 2, Literacy, and Proficiency Tests (the proficiency means are in $\log_{10}$ units and are plotted against John Test scores in Figure 4.4 and fall Decoding 1 scores in Figure 4.5 because the Proficiency Test had not been administered as a pretest). Individual classrooms are coded in the figures by the same letter and number system used in earlier figures.

The line drawn through the points in each figure is the pooled-within-classrooms regression line, and may be taken as the point of reference for

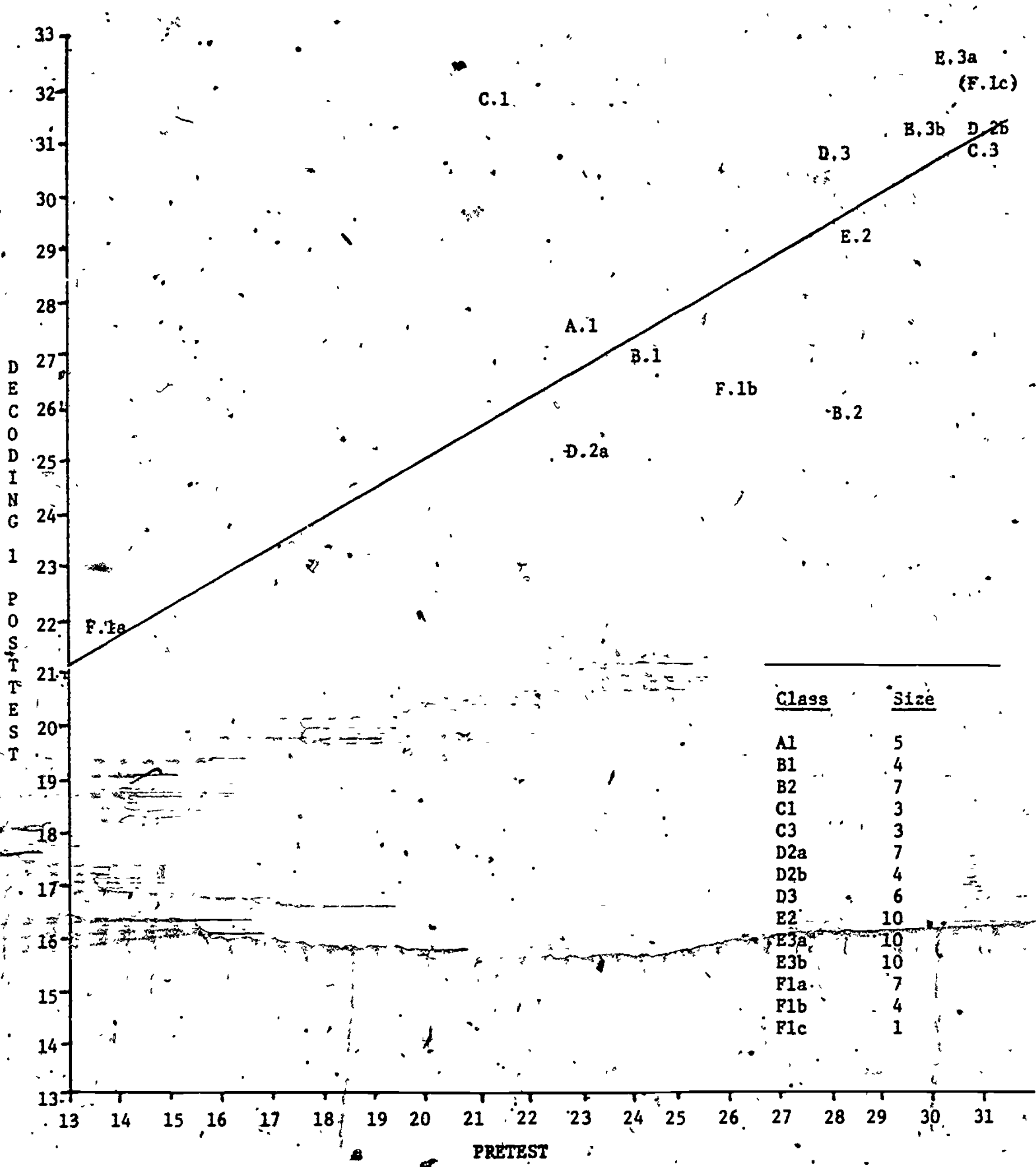

FIGURE 4.1

Decoding 1 Classroom Means for Fall and Spring Administrations, Day-School Sample (Pooled Within-Classrooms r = .69; Total Sample r = .73)

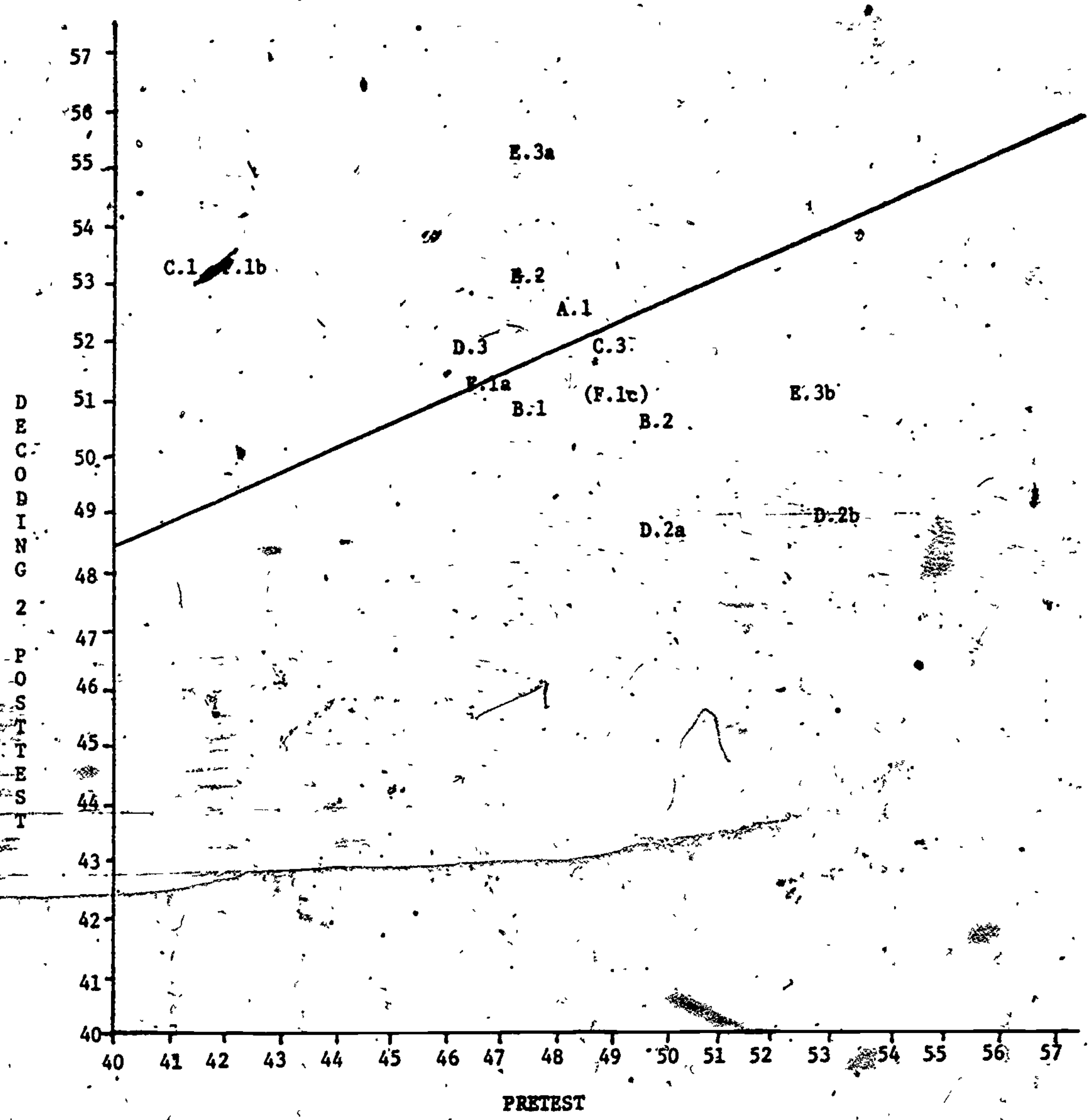

FIGURE 4.2

Decoding 2 Classroom Means for Fall and Spring Administrations, Day-School Sample (Pooled Within-Classrooms r = .48; Total Sample r = .27)

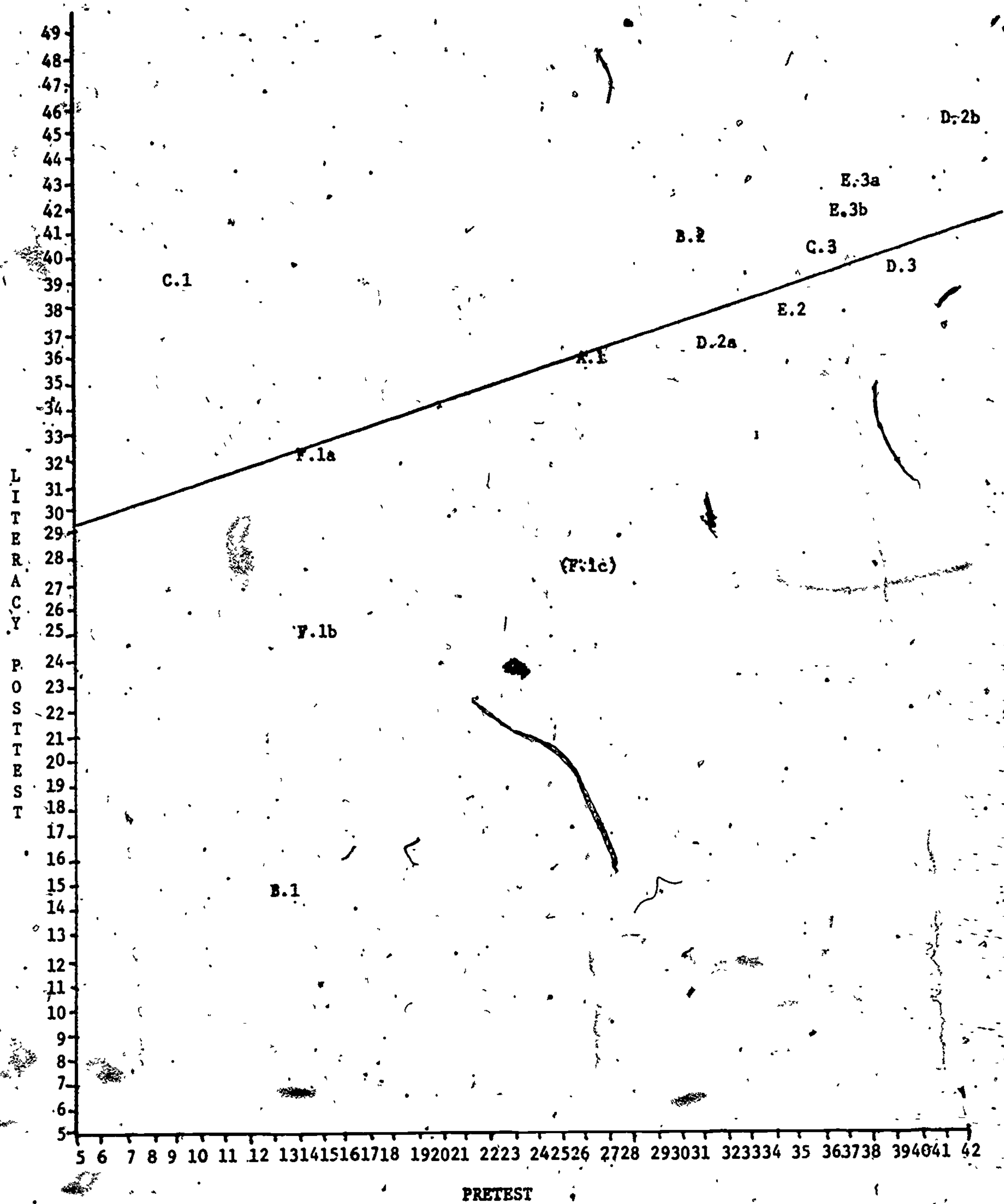

FIGURE 4.3

Literacy Classroom Means for Fall and Spring Administrations, Day-School Sample (Pooled Within-Classrooms r = .59; Total Sample r = .71)

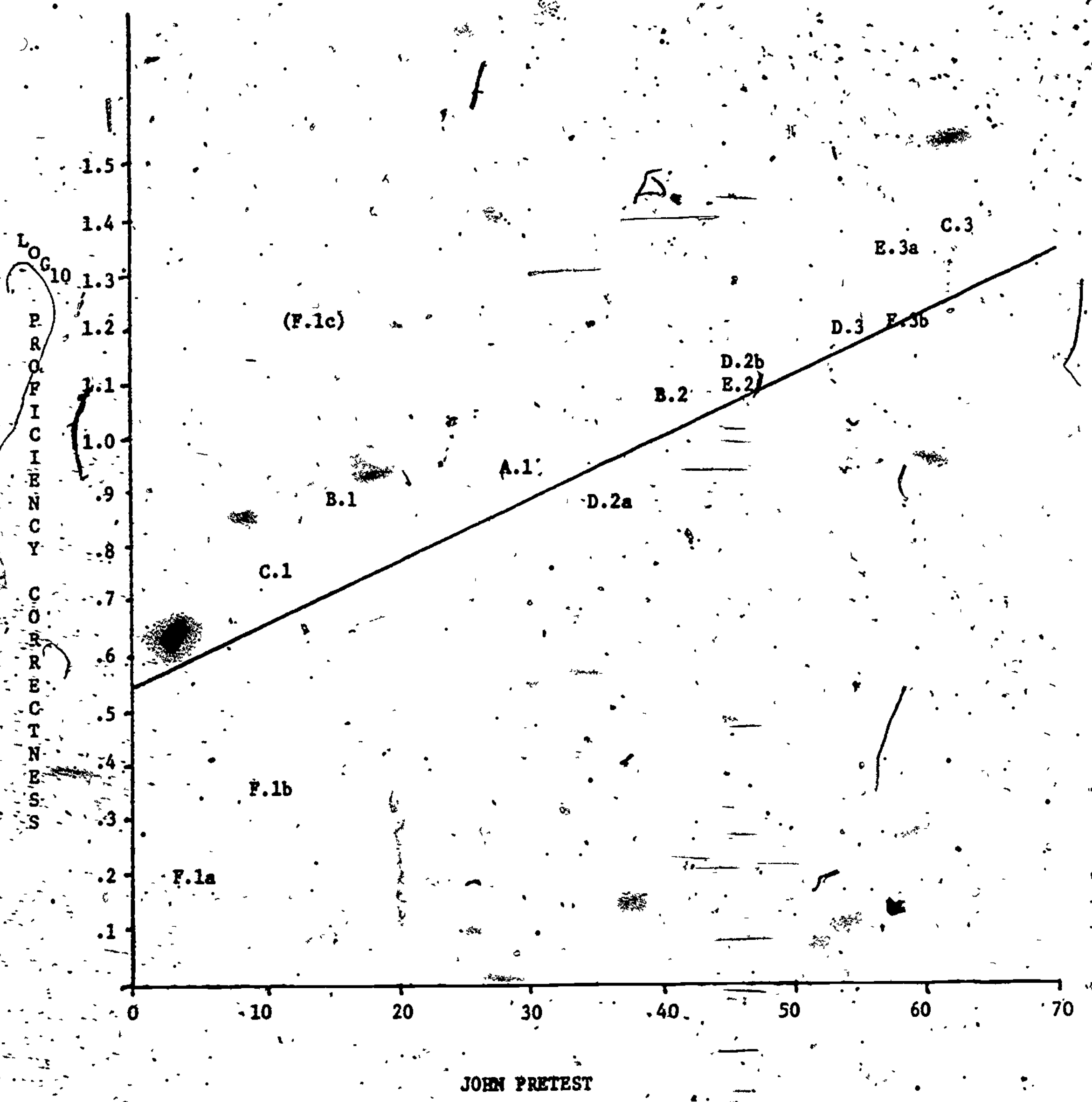

FIGURE 4.4

Log_{10} Spring Proficiency Correctness Classroom Mean Scores x Fall John Test Classroom Mean Scores, Day-School Sample (Pooled Within-Groups r = .28, Total Sample r = .77)

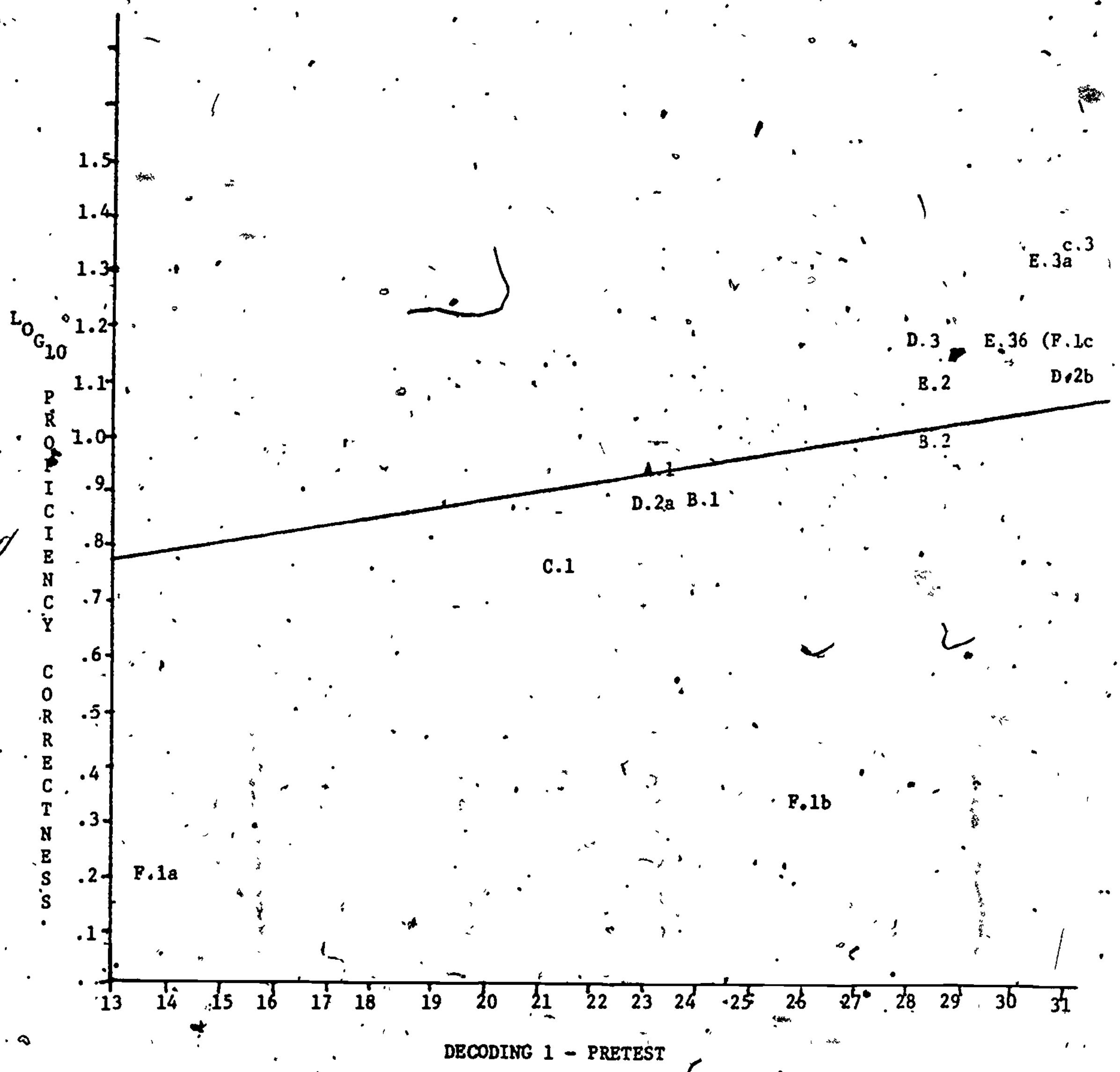

FIGURE 4.5

Log_{10} Spring Proficiency Correctness Classroom Mean Scores x Fall Decoding 1 Classroom Mean Scores, Day-School Sample (Pooled Within-Groups r = .44, Total Sample r = .67) determining the adjusted gains or losses of individual classrooms if the within-classrooms regression coefficients are assured to be homogeneous enough to justify pooling. Whether or not pooling is justified is hard to ascertain with the limited data available because class sizes range only from 1 to 10, as indicated in the inset in Figure 4.1. Since some of the within-classrooms regression coefficients were negative for all analyses except Decoding 1, there is reason to suspect that the assumption of homogeneity of within-groups regressions is untenable.

Differences in Achievement of Literacy

Referring back to Table 4.4, we see that the Literacy Test shows evidence of much variation among adjusted classroom posttest means which cannot be accounted for by reference solely to the four reliable classroom interaction contrasts. From Figure 4.3 we see that this variation is probably due to the exceptionally good posttest performance of the level 1 students in classroom C1. These students begin the year at the very lowest level of performance on the Literacy Test, but, by the end of the year, they are as literate as students in several level 2 and level 3 classrooms. This performance contrasts markedly with that of students in classroom B1, where spring Literacy Test performance is essentially unchanged from that in the fall. Obviously, teacher C is increasing the English language literacy of level 1 students more than other teachers at that level. A similar conclusion holds for Decoding 1 in Figure 4.1.

Differences in Achievement of Decoding Skills

The Decoding 2 test has been revealed to be a bit anomalous in the foregoing analyses in that performance on the test is not highly predictable from a knowledge of student background characteristics, pretest performance, or classroom interaction experiences. From Table 4.4 we see, moreover, that there are real differences among classrooms in terms of adjusted Decoding 2 posttest means which cannot be accounted for by reliable differences in observed classroom interaction patterns. Figure 4.2 helps clarify what might be going on with Decoding 2 scores, which in part measure how well students can pronounce or understand the pronunciation of English words. It seems that certain classrooms in which students have the poorest relative grasp of phoneme-grapheme correspondences in the fall are the very classrooms in which the students have achieved a relatively superior grasp of these correspondences by the end of the school year; and vice versa. One obvious way in which we can interpret these results is to hypothesize that teachers in some sense overreact to their students' initial abilities in English pronunciation: if these abilities are initially minimal, then improvement is sought; if these abilities are initially superior, then other aspects of performance are emphasized.

Differences in Achievement of Oral Proficiency

As for the Oral Proficiency Test structure scores, little is revealed by inspection of Figures 4.4 and 4.5 except that the achievement of level 1 students in classes F1a and F1b is lower than might be expected. Notice, however, that neither the John Test nor the Decoding 1 Test is an optimum "pretest" for oral proficiency; unfortunately, the Oral Proficiency Test itself was not developed until after pretesting was completed.

The tendency for level 1 classes to lie below the pooled-within-groups regression lines in most of the figures (except 4.2), while level 3 classes tend to lie above the same line, indicates the need to include more than one covariate in each regression analysis in order to account adequately for preexisting differences among classrooms. This was our reasoning in the decision to use all available background and pretest scores as covariates in the regression analyses reported earlier. Although the relationships of these predictors to posttest performance were thoroughly explored and discussed in connection with Tables 4.1 and 4.2, we had not yet taken into account the grouped structure of the data (students within classrooms) as was done with the analyses presented in Table 4.4.

Once all 13 possible classroom interaction contrasts are included in the analysis, moreover, there can be changes in the regression weights which must then be attached to background and pretest variables for optimal prediction of posttest performance. These final regression weights characterize a model in which the grouped structure of the data is taken properly into account. They are of interest because they should generalize to predict variation among students within a broader range of classrooms than the limited selection included in this study. In other words, once all the effects of classroom differences are eliminated, we can generalize about variation among students within any classroom which might be similar to those studied here (provided we accept the assumption of homogeneity of regression planes).

Pooled Within-Classrooms Regressions of Posttest Achievement Scores on Student Background Characteristics and Pretest Scores

Table 4.5 presents standardized regression coefficients of spring posttest scores on student background and pretest variables with the effects of classroom

TABLE 4.5

Standardized Pooled Within-Classroom Regression Coefficients; Day-School Sample

	D_1-S	D_2-S	L-S	$\log_{10}$ P-CT	$\log_{10}$ P-CP	$\log_{10}$ P-ST
Sex (F)	-.12	-.13	-.17*	.01	.12	-.01
Age	-.11	.10	-.30**	.10	.08	.04
Time in U. S.	-.05	.04	-.11	-.21+	-.26*	-.24+
Former Job	.11	.09	.13+	-.02	.13	.04
Job in U. S.	.12	.10	.06	-.03	.05	-.01
C-Origin	-.09	.18+	-.05	.08	.00	.12
Years Education	.01	.14	.05	-.03	-.05	.02
Former Eng.	-.16+	-.06	-.12	-.27*	-.32**	-.24*
Eng. in U. S.	.05	.04	.16+	-.34*	-.23*	-.32**
D_1-F	.58***	.07	.24*	.20	.32*	.14
D_2-F	.18+	.39**	.00	.05	.16	.09
L-F	.04	.23+	.38**	.26*	.16	.26*
J-F	-.06	.09	.03	.27*	.29*	.36**
M-F	-.09	-.03	.01	.00	-.13	.04

+1.30

*1.67

**2.40

***3.26

differences eliminated (day-school sample). This table is a companion to Table 4.4 since the results were obtained in the same overall analysis. We have held off discussion of these coefficients until now because the coefficients presented are analogous to the pooled-within-classrooms regressions (correlation, actually, since they are standardized) discussed in connection with Figures 4.1 through 4.5.

When the standardized pooled-within-classrooms regression coefficients in Table 4.5 are compared to corresponding entries for the total sample in Table 4.1 (posttest section) or in Table 4.2, several noteworthy discrepancies can be observed. It is not surprising that the large role played by initial John Test scores in the prediction of total-sample variation in Literacy and Proficiency posttest scores is substantially diminished once classroom differences are partialled out. A major distinction between classrooms is the proficiency level of the students, as measured by John Test scores, so within-classrooms variation in John Test scores is minor relative to between-classrooms variation. Notice that the residual within-classroom variation on John Test scores only retains predictive utility for spring Oral Proficiency Test scores (Table 4.5).

We have already noted the fact that between-classroom regressions tend to contradict within-classroom regressions for the Decoding 2 Test, as can be seen in Figure 4.2. A comparison of Decoding 2 standardized regression coefficients on pretest scores in Tables 4.2 and 4.5 will reveal the outcome of this contradiction: a low weight on the Decoding 2 pretest in the total sample even though that score turns out to have substantial predictive ability within classrooms.

Notice from the lower section of Table 4.5 that most, if not all, of the posttests are related to more than one pretest, when predicting any student's final achievement relative to that of other persons in the same classroom. This suggests that the use of only one pretest as the covariate for each posttest (i.e., the same test administered in the fall) is inadequate; it also suggest theoretical models which might account for why some students achieve more than others within the same classroom. For instance, it looks as though the more literate students within any given classroom may get a head start on learning phoneme-grapheme correspondences (Decoding 2). Moving to the upper section of Table 4.5, we can perhaps infer that a western European background as well as a prolonged education may also facilitate the acquisition of Decoding 2 skills, as does being a member of the male sex.

A similar approach to theorizing about within-classroom variation in achievement can be indulged in on the basis of the coefficients in any column of Table 4.5. It could be argued that the acquisition of literacy might well be facilitated by prior word meaning (Decoding 1), having prior English training in the United States as opposed to abroad, a higher level of prior employment, youth, and masculinity (although we have no direct evidence to bear on this, the last two effects may well be due to the tendency for housewives to be isolated from English-language stimulation in the home, and for older individuals to be assimilated into an ethnic community).

Achievement on the Decoding 1 posttest, relative to peers in the same classroom, seems to be primarily a function of a student's initial command of the vocabulary skills required by that test. It is noteworthy, however, that prior and current levels of employment may be positive predictors while the amount of English studied in the previous country is a negative predictor; it is tempting to speculate that proper familiarity with word meaning must be acquired in a context of active use.

As for the prediction of within-classroom variation in achievement on the Oral Proficiency Test scores, we see from Table 4.5 that all three scoring systems relate positively to prior John Test performance. It is noteworthy, however, that achievement of proficiency under the more stringent scoring systems seems to be facilitated by initial literacy. Perhaps those students who have the ability to read better than their peers within the same classroom can gain more experience with the structural features of the English language through the course of the year. Achievement on the Oral Proficiency Comprehension score, on the other hand, seems to be facilitated by a student's initial knowledge of vocabulary, as measured by the Decoding 1 test. Perhaps an initial repertoir of root words and syllables can facilitate the acquisition of sufficient vocabulary for comprehension; notice that it seems to have little impact upon the achievement of structural accuracy, however.

The pooled within-classrooms regressions of Oral Proficiency scores on student background characteristics seen in Table 4.5 are intriguing. It is apparent that those students within a given classroom who have relatively less prior training in English, whether here or abroad, are the very students who, in terms of final achievement, profit the most from instruction. A similar phenomenon seems to occur with respect to the variable "length of time in the United States." Recent arrivals who have had little prior opportunity to live in an English-language milieu make substantial progress in achieving proficiency by the end of the school year.

BETWEEN CLASSROOMS RELATIONSHIPS AMONG STUDENT BACKGROUND CHARACTERISTICS, ACHIEVEMENT, AND TEACHER/STUDENT INTERACTION VARIABLES

A comparison of the total-sample regressions of posttest scores on student background characteristics in Table 4.1 with the corresponding pooled within-classrooms values in Table 4.5 reveals substantial change in the predictive role played by age as well as history of employment in the country of origin. This suggests that classrooms are differentiated not only by level of proficiency, as measured by initial John Test scores, but by certain student background characteristics as well. In order to explore this possibility we turned to univariate analysis of variance and canonical discriminant function analysis, just as was done in Chapter Three to study classroom differences in student/teacher interaction patterns.

Table 4.6 summarizes the outcome of univariate analyses of variance among day-school classrooms on student background characteristics. Notice that only the amount of English studied in the United States varies widely among classrooms (not surprisingly, since this is the best predictor of John Test scores recorded in Table 4.1). There is some indication, however, that previous employment and educational level also differ from classroom-to-classroom, so it is worthwhile to study the outcome of canonical discriminant function analysis in hopes of clarifying these relationships.

Not too surprisingly, one highly significant discriminant axis emerged from the analysis, and it accounted for 59 percent of the between-classroom variation. Details of the analysis need not be presented here; suffice it to say that amount of English studied in the U. S., English studied in the previous country, length of time in the U. S., years of education, youth, and previous occupational level were weighted in the first discriminant function in the

TABLE 4.6

Univariate Analysis of Variance Tests for
Classroom Differences in Background Characteristics

Variable	Day-School F-Ratio
Sex (F)	1.34
Age	.66
Time in U. S.	.77
Former Job	1.72^{+}
Job in U. S.	.54
C-Origin (NC)	.73
Years Education	1.79^{+}
H. S. Diploma	.63
Former Eng.	1.47
Eng. in U. S.	4.32^{***}
Between Classrooms Degrees of Freedom	13
Within Classroom Degrees of Freedom	67

$^{+}p < .10$

$^{***}p < .001$

order listed. By taking into account the second canonical axis, the percentage of variance accounted for could be brought up to 73 percent, and the resulting solution is plotted in Figure 4.6. The second axis is of marginal statistical significance; it is a weighted linear combination of sex, English studied in the previous country, and national origin, in that order.

As for other figures relating to discriminant function analysis, we have represented the classroom centroids in Figure 4.6 with unique letter-number codes, while the student background characteristics have been represented by vectors. Inspection of the figure reveals wide variation among day-school classrooms in terms of student background characteristics. It can also be seen that much of the variation among classrooms is related to level differences as determined by John Test scores.

As far as the background characteristic vectors plotted in Figure 4.6 go, there seem to be some rather suggestive clusters: youth and amount of English studied in the U. S.; educational and occupational level in the previous country; amount of English studied in the previous country and occupational level in this country; sex (female), origin (Cuban), and length of time in the United States. All in all, these results suggest the possibility that students have been assigned to more-or-less homogeneous classrooms on the basis of background characteristics in addition to their John Test scores.

It is enlightening to compare Figure 4.6 to the earlier figures in this chapter, in which pre- and posttest achievement are plotted without taking any notice of student background characteristics within the classrooms involved. This comparison makes the exceptional Literacy gains of students in classroom C1 seem all the more impressive, for instance (Figure 4.3).

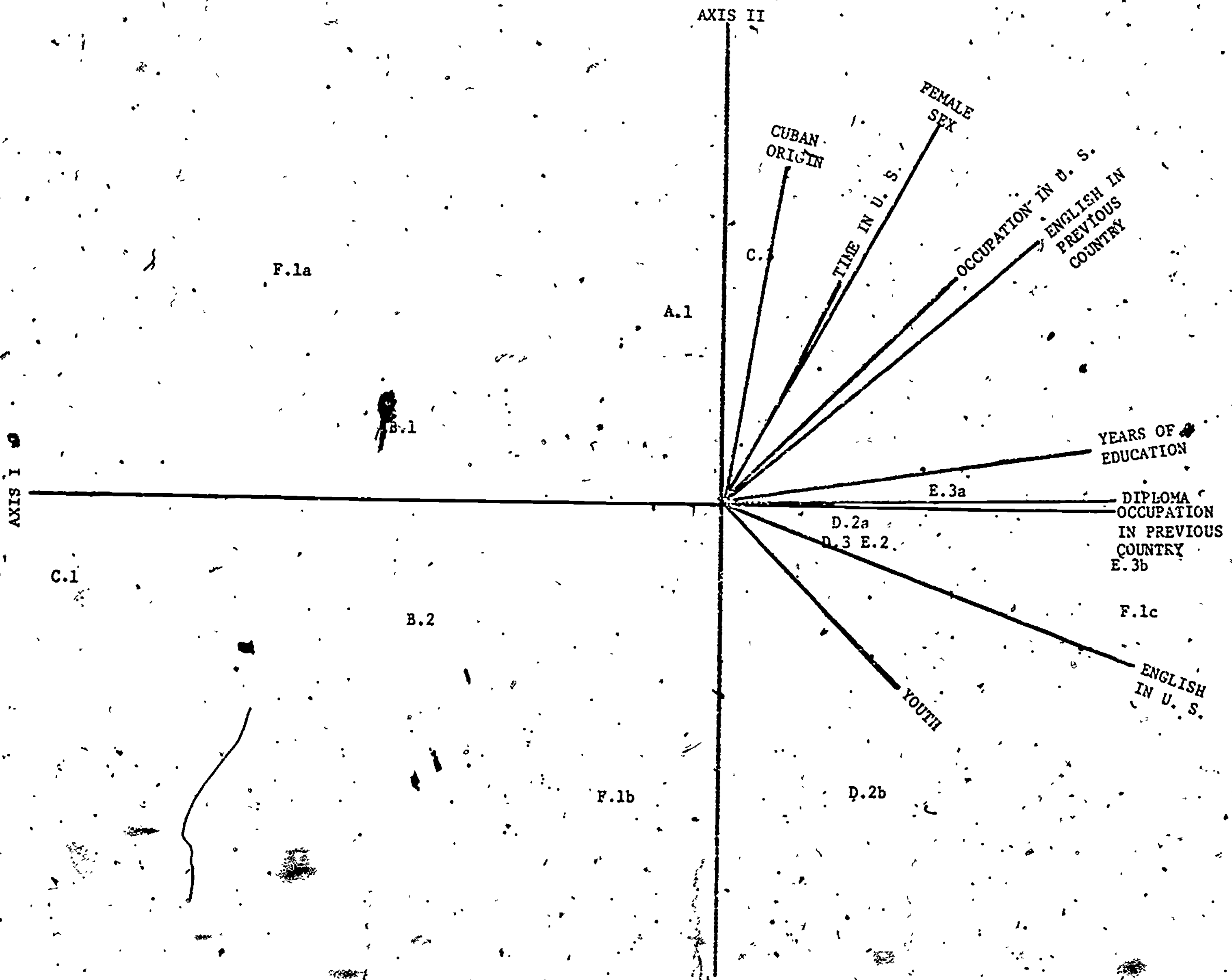

FIGURE 4.6

Plot of Day-School Classroom Centroids and Student Background Characteristics in the Space Depicted by the Two Largest Canonical Discriminant Axes

When Figure 4.6 is compared to Figure 3.2--a similar analysis of classroom differences, but in terms of teacher and student interaction patterns--several fascinating points of agreement emerge. Notice that the two classrooms taught by teacher C are widely discrepant with regard to student background characteristics (Figure 4.6), and that these two classrooms are handled in widely different ways by teacher C (Figure 3.2). These data suggest that the manner in which a particular class is taught may reflect a special adaptation on the part of the teacher to the background characteristics and ability level of the students involved. Although a bit more speculative than our reasoning with regard to teacher C's behavior, notice that the three level 1 classrooms taught by teacher F are spread widely with regard to background characteristics (Figure 4.6) and that their respective locations in Figure 3.2 suggest the application of a more directive teaching strategy in the less well prepared classrooms. As for teachers B and D, there seems to be a possible association between the amount of student-student feedback in Spanish which occurs in a classroom (Figure 3.2) and its composition in terms of country of origin, sex, and length of time in the U. S. (vertical direction in Figure 4.6); a similar inference can be drawn for teacher C.

In summary, we have accumulated a good deal of evidence suggestive of relationships between classroom interaction variables and student achievement variables. The regression analyses presented in earlier sections of this chapter were designed to reveal how classroom interaction experiences might come to affect student performance. Now we see, however, that student performance and background characteristics may well affect classroom interaction; i.e., a model of reciprocal causation is probably most appropriate for this data.

In order to carry out a more systematic exploration of possible relationships between classroom composition and the teacher/student interaction which takes place therein, we turned to the technique of canonical correlation. Since the primary product of the canonical discriminant function analysis carried out in Chapter Three was a set of classroom centroid scores on discriminant axes (as plotted in Figure 3.2), and a corresponding set of classroom centroid scores is now available with respect to student background characteristics (as plotted in Figure 4.6), we can apply canonical correlation analysis to search for systematic relationships between these two sets of classroom coordinates. In fact, canonical correlation analysis seeks to find those orthogonal linear combinations of the scores in each set which are, in successive order, maximally related to one another.

The potential benefit of applying canonical correlation analysis to the classroom centroid coordinates derived separately through canonical discriminant function analysis of teacher/student interaction and student background characteristics (or student performance variables) stems from the fact that we are dealing with highly reliable indices of between-classroom variation in each case. Discriminant function analysis serves to maximize between-classrooms variations relative to within-classrooms variation. It has already allowed us to assess classroom differences relative to day-to-day variation in teacher/student interaction (Figure 3.2) as well as classroom differences relative to student-to-student variation in background characteristics (Figure 4.6); these analyses were done separately, however.

Whether dealing with teacher/student interaction or student characteristics, the canonical discriminant function analyses yield reliable indices of variation at the classroom level; the possibility of relationships among these different domains of variation at the classroom level can thus be explored through the use of canonical correlation analysis of classroom centroid projections onto the discriminant function axes.

Since our interest in variation between classrooms includes not only teacher/student interaction and student background characteristics, but student pretest and posttest performance as well, the latter variables were included in a new canonical discriminant function analysis comparable to that conducted in connection with Figure 4.6. It is not surprising that multivariate analyses of variance on pretest and posttest achievement measures indicated highly significant differences among the 14 classrooms in the day-school sample for everything except Decoding 2 scores. These level of achievement differences were reflected in the composition of the canonical discriminant axes derived from student background, pretest, and posttest variables combined.

The first two canonical axes accounted for 84 percent of between groups variation, relative to within groups variation, and weighted primarily the fall John Test and Spring Literacy Test, respectively. In all, four significant canonical discriminant function axes were obtained, accounting for 92 percent of the relative variance among classrooms. We will not report the analysis in more detail here, however, since it served only to provide classroom centroid coordinates for input to canonical correlation analysis.

As input to the canonical correlation analysis of classroom means on canonical discriminant function axes, we used the four composite scores for student characteristics just mentioned as well as scores on the first five canonical discriminant function axes derived from classroom interaction data (Table 3.6). Five classroom interaction canonical scores were used instead of only the four emphasized in other sections of this report because of evidence we found while conducting regression analyses (reported earlier in this chapter) that reliable classroom interaction differences existed beyond the original four transformed classroom interaction contrasts. Notice that the fifth canonical axis reported in Table 8.6 is marginally significant, in line with this finding.

The canonical correlations resulting from mutual prediction, on a classroom-by-classroom basis, of student background and achievement data from classroom interaction data (or *vice versa*) were .96, .74, .56, and .34. Even allowing for a sample size of only 14, Wilk's Λ criterion yielded a significant χ^2 value of 33.7 with 20 degrees of freedom, confirming our suspicion that what goes on in classrooms with respect to stable patterns of teacher/student interaction is highly related to student background and performance characteristics.

The Structure of Between Classroom Variation in the Adult Learning Center

The canonical correlations reported above give evidence of rather strong relationships between student background/performance characteristics and teacher/student interaction factors at the classroom level, when dealing with scores on canonical discriminant function axes. It is in order to ask whether we might be able to relate these relationships back down to the level of individual background characteristics, test scores, and classroom interaction factors. The answer is yes, although we must keep in mind that the results are dependent upon only 14 classrooms studied in the Adult Learning Center and may not generalize widely. Only where a similar system of assignment of students to classrooms, a similar range in teaching styles, and a similar population of students can be assumed to exist will these results generalize to other ESL systems.

Table 4.7 contains the final outcome of a series of projection and transformation operations. In brief, the two sets of canonical correlation weights derived from relating student background/performance classroom centroids to teacher/student interaction classroom centroids were used to transform between-classrooms projections of individual student background characteristics, test scores, and classroom interaction factors so as to relate to a common canonical orientation of axes. Then the resulting combined matrix of projections of original variables onto common axes (subject to some uncertainty due to less-than-perfect canonical correlations) was transformed in an attempt to achieve a positive manifold simple structure, using a version of DIRECT GEOMIN rotation (Table 4.7).

The relationships between student background/performance variables and teacher/student interaction variables for the 14 day-school classrooms represented by all four canonical correlations (.96, .74, .56, .34) were included in what led up to the results presented in Table 4.7, even though the last two canonical correlations are quite low in value. This was done because we found that rotation of all four common axes for positive manifold simple structure in their between-classrooms relationships to the original variables did not yield a marked change in position from the original canonical orientation; therefore, any uncertainty in the relationship between student background/performance variables and teacher/student interaction factors due to imperfect canonical correlations is largely confined to the later "factors" listed in Table 4.7.

TABLE 4.7

Between-Classrooms Factors of Student Background Characteristics, Pre- and Posttest Achievement, and Teacher/Student Interaction; Adult Learning Center

Factor Pattern Coefficients

	I	II	III	IV
Background				
Sex (F)	.00	.03	.74	-.21
Age	-.36	-.15	.04	-.16
Time in U. S.	-.04	.31	.33	-.75
Former Job	.38	.05	.55	.00
Job in U. S.	-.21	.36	.79	-.02
C-Origin (NC)	.04	.05	-.34	.45
Years Education	.22	.06	.56	-.25
H. S. Diploma	.38	-.17	.47	-.21
Former Eng.	.14	.24	.59	-.11
Eng. in U. S.	.78	.04	.14	.21
Pretest				
D_1-F	.63	.10	.23	.48
D_2-F	.59	.35	.05	-.61
L-F	.66	.50	.15	.05
J-F	.52	.47	.35	.21
M-F	.35	.35	.55	.32
Interaction				
QA-IND	-.09	.38	-.23	.66
CF-P-A	-.04	.37	-.51	.64
SFBK	-.73	.15	-.15	.59
FREE	.57	.43	.04	.47
OTHER	-.03	.50	.23	.64
DIR READ	.02	-.18	-.11	-.81
MOD PRAC	.28	-.22	-.45	-.68
CF-M-P	.42	-.05	-.08	-.87
QA vs. DIR	.59	-.52	-.55	-.02
Posttest				
D_1-S	.10	.15	.64	.49
D_2-S	-.41	-.04	.51	.71
L-S	.10	.94	.01	-.04
R-CT	.29	.35	.51	.40

Factor Intercorrelations

	I	II	III	IV
I	1.00			
II	.20	1.00		
III	.34	.21	1.00	
IV	.01	.20	-.08	1.00

Interpretation of Between-Classrooms "Factors" of Student Background Characteristics, Pre- and Posttest Achievement, and Teacher/Student Interaction

The "factors" given in Table 4.7 can be interpreted in much the same way as those resulting from conventional factor analysis as long as it is kept in mind that we are discussing hypothetical determinants of manifest variation among the 14 classrooms at the Adult Learning Center. Although only 14 classrooms were studied, we know that the classroom centroid scores which served as the basis for this analysis were highly reliable to start with (i.e., they showed maximum variation among classrooms relative to variation within classrooms--whether day-to-day for teacher/student interaction or student-to-student for background characteristics and achievement).

Notice that the background, pretest, interaction, and posttest variables listed in Table 4.7 have been organized and grouped in the temporal sequence in which they occur in practice. We are not attempting to infer how these domains of variables might influence one another, however--that was the role of the regression analyses reported earlier. Rather; we are attempting here to discover what factors might determine how students from various backgrounds and of different final as well as initial ability levels might come to be exposed to various specific forms of student/teacher interaction within the 14 classrooms of the Adult Learning Center.

In attempting to infer the bases for classroom-to-classroom variation in the Adult Learning Center through inspection of the coefficients in Table 4.7, the first thing we can note is that most of the variables have like-signed loadings across all four factors and that the factors show generally positive intercorrelations. These features, coupled with the fact that the John and Morano pretests as well as the Oral Proficiency correctness posttest score are positively loaded on <u>all four</u> factors, suggests strongly that all of the factors are getting at different aspects of the same general distinction among classrooms within the Adult Learning Center: level of English language proficiency.

Each factor can therefore be regarded as a specific bipolar contrast between high and low level classrooms. If we interpret the loadings as they are presented, we will thus arrive at descriptions of hypothetical ways in which classrooms might come to contain students of superior proficiency in English (within the Adult Learning Center); reflection of the signs of the loadings can therefore indicate how Adult Learning Center classrooms might come to contain students of lower proficiency in English.

Looking at the positive pole of the first factor, we see the likely tendency for superior Adult Learning Center classrooms to contain students who have a history of higher than average English Study in the U. S., Level of Former Job, and Educational Level; while having lower than average Age and Level of Job in the U. S. Moving to the loadings of pretest performance on this factor, we see a tendency toward superior performance on all initial achievement measures. It is not surprising that these indicators of competence are getting at the same aspect of variation among classrooms as prior English study, level of former job, and education. The negative loading for level of employment in the U. S. is interesting, however, and may betray the fact that students who have the time to engage in extensive day-school English-language training in this country tend either to be unemployed or underemployed, compared to their former job and educational level (hence the motivation to improve their English).

The loadings of student/teacher interaction variables on the first between classroom factor imply above average prevalence of the "free response" mode of interaction, as might seem appropriate for students of higher than average initial ability. Note the tendency for "student-student feedback" either not to occur or not to be allowed, along with the "model-practice-corrective feedback-model-practice" paradigm. There is a tendency for students within classrooms with such a high level of prior training and initial ability not to be expected to read and/or ask questions under direction of the teacher; on the contrary, a (group?) question-answer interaction pattern goes along with "free response" format of instruction.

As for the final achievement implied by this first factor of between-classrooms variation, somewhat above average Oral Proficiency correctness is indicated along with below average Decoding 2 performance. Here we see again the phenomenon noted in connection with Figure 4.2; the implication is that students in higher-level classrooms do not get exposed to instruction which deals with pronunciation in terms of phoneme-grapheme correspondence--they consistently end up being surpassed in these skills by students in the initially lower-level classes. The latter classrooms define the negative pole of the first factor; notice that students there have higher than average levels of employment in this country and are older.

At the positive pole of the second factor of between-classrooms variation there is revealed a tendency for some superior classrooms to contain students who have been in this country for a relatively long while and who are established in higher level jobs. This goes along with a history of English study in the former country and backs up our earlier conjecture that prior English language training may be instrumental in the acquisition of higher level employment in this country. Note, however, that a high overall level of education is not indicated.

All of the pretest achievement scores except Decoding 1, the vocabulary indicator, are loaded on this factor. Glancing ahead to the loadings of posttest scores, we find that we are dealing here with a factor which relates to exceptional achievement of posttest literacy, along with the above-average achievement of proficiency characteristic of all four factors.

When it comes to the nine classroom interaction factor loadings, it is not surprising that the "direct-read and/or ask questions" pole of the last interaction factor is highly associated with this between-classrooms factor relating to posttest literacy. Other characteristic classroom interaction patterns are "Other" and "Free," along with the "question-answer-corrective feedback-prompt-answer" sequence. This leads us to suspect that some of the "Other" activities which occur with high frequency within certain higher-level Adult Learning Center classrooms might involve reading materials.

Most of the student background characteristics load on the third factor of between-classrooms variation at the Adult Learning Center. We could view the positive pole as indicating a tendency for superior classrooms to contain well employed females who have been relatively well educated. Individuals in these classrooms, who may well have arrived from Cuba some time ago (relatively speaking), have had superior English language training in their country of origin, but not necessarily in the United States. Age is the only other background variable not related to this factor.

It is interesting that the Morano pretest has a large loading on the third between-classrooms factor, especially given its minor relationship to other measures of within-classrooms variation (Table 4.5). The Morano score is an indicator of grammatical skill. Looking ahead to the loadings of posttest measures on the third between-classrooms factor, we see that the oral proficiency correctness score has its highest loading here. Taken together with the high loadings for posttest decoding skills it appears that substantial ESL achievement tends to occur in Adult Learning Center classrooms characterized by the positive pole of the third factor. This could well be attributed to the high current as well as prior level of employment, former ESL training, educational level, or sex (female) of the individuals involved. Notice that only the "Other" factor of classroom interaction has a position loading on the third factor in Table 4.7, and the "teacher-direct-student read and/or ask question" pole of the last classroom interaction factor predominates. Other classroom interaction patterns are deemphasized, including "model-practice" as well as "question-answer-corrective feedback-prompt-answer." This leaves us with little basis for speculating about what <u>does</u> go on in these classrooms; perhaps the student background characteristics mentioned make instruction irrelevant to achievement.

Just as the third between-classrooms factor is strongly related to student background characteristics, the fourth factor is strongly related to all but one of the original nine classroom interaction factors. The pattern of loadings suggests the familiar contrast between classroom interaction factors seen along the horizontal axis of Figure 3.2: QA-IND, CF-P-A, SFBK, FREE VS. MOD-PRAC, CF-M-P, DIR READ, and OTHER. At earlier points in our discussion we have referred to the positive (left in Figure 3.2) pole as flexible, free, supportive, eclectic, individualized, while referring to the negative pole as structured, class or group oriented, etc.

Those classrooms in which a more supportive and individualized pattern of teacher-student interaction prevails tend to contain students who are recent arrivals from Western Europe, perhaps not too highly educated , a bit younger than average, and male. This means, of course, that the more highly structured classrooms (e.g., those of teacher F in Figure 3.2) tend to contain older females of Cuban or Carribean origin who have been in the United States for some time. As for the test performance, it is clear that the recent arrivals from Western Europe are very deficient in Decoding 2 skills (phoneme-grapheme correspondences) in the fall but become exceptionally proficient in pronunciation by the end of the school year. Again, this illuminates the tendency noted in connection with Figure 4.2 for Decoding 2 scores to increase dramatically in those classrooms in which they are initially low, and not to change if they are initially high.

What these four factors of between-classrooms variation signify in a global sense is the complexity and richness of the ESL training situation in terms of student background characteristics, initial test performance, student teacher interaction patterns, and final achievement. The Adult Learning Center must be viewed as a dynamic system in which students are channeled into classrooms which promise to provide them with an optimal learning experience. Table 4.7 indicates that there are at least four ways in which classrooms can come to be above average in the achievement of English language proficiency at the end of the period of instruction. It is clear that initial proficiency as determined by the John Test is a good across-the-board guarantee of final proficiency, not surprisingly; but, among other aspects of pretest performance, background characteristics and classroom interaction patterns must be taken into account in order to fully characterize the variety of patterns of achievement seen in different classrooms.

In discussing Table 4.7 we have mainly taken note of loadings which exceed .20 in value. Since our discussion has necessarily been brief, it would be useful for the reader to carry out a detailed inspection of the entries in Table 4.7 on their own. This analysis was arrived at through a rather unconventional approach but it shows promise as a way to relate between-classrooms variation in student characteristics to between-classroom variation in daily patterns of teacher/student interaction without the need to record that interaction on a student-by-student basis.

CHAPTER FIVE

OVERVIEW, SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

INTRODUCTION

In this chapter we present essentially the same information as in the preceding chapters but in a less technical way. We also discuss the alternative interpretations of the data, the kinds of hypotheses about teaching that ought to be considered, and some proposals for action that derive from interpreting the results of this study. The goal is to make the technical information presented in the preceding chapters available to a larger audience and in particular to the staff of the West New York Adult Learning Center and to members of the State Department of Education who may wish to present it to their colleagues who have less technical training than is required to read Chapters Three and Four.

The preceding chapters presented a wealth of factual information: about the characteristics of the students in the Adult Learning Center in West New York; a detailed description of the teaching styles of the teachers; a description of the students' achievement in acquiring proficiency in speaking English; and an analysis of the relationship between the nature of the teaching styles employed by the teachers, student characteristics and student achievement. These chapters have been presented deliberately in a technical style. They represent a set of facts and the results of analyses which a reader may examine to see if he or she would draw the same conclusions from the information that has been provided.

This chapter is more than a summary. It is a briefer description of the work of the project and the results of the research methods that were employed. It also provides interpretations of these results with a view to suggesting courses of action.

We remind the reader, as we have at frequent points in this document, that the results of this study are not generalizable to other samples of students or teachers or adult learning centers except insofar as these other students or teachers or centers have characteristics highly similar to those of the West New York Adult Learning Center. Also the methods of the study do not, strictly speaking, permit us to draw strict causal conclusions; we can only make inferences about probable causes.

The practitioner, however, is interested in knowing what might be done with the results of this study. In the latter sections of this chapter we will make suggestions about such courses of action. Our presentation there will consist in presenting the results of the study as premises in a rational analysis of practical action. In so doing, we will suggest the implications of these premises for practical action, and inevitably will present different courses of action.

Since we have given ample warning, the reader will recognize that we are not presenting the results of this study as definitive statements about effective teaching of English as a second language. With these warnings in mind, whose origins are in the logic of statistical and empirical analysis, we can suggest ways of thinking about these data that would lead to practical action.

Others may develop their own plans for practical action by following a similar kind of reasoning. Those who undertake practical action based upon the analyses that we have presented ought also to evaluate the results of these actions. If that course is followed, the conclusions drawn for action are tested, and their utility can be assessed.

The order of presentation of ideas and information in the following section is: (1) the major questions and problems to be studied; (2) the methodology of the research project; (3) a description of the students of the Adult Learning Center; (4) the analysis of teaching performance; (5) the relation of teaching performance and other variables to the students' achievement of proficiency during the course of the study; and (6) proposals for practical action based upon the conclusions of the study.

THE MAJOR QUESTIONS AND METHODS OF THE RESEARCH PROJECT

What was the purpose of this research project? The purpose was to find, if they existed, relations between teaching performances and the achievement of proficiency in speaking English. The idea was to collect information about these performances by observing teachers on a regular basis who were teaching English as a second language to the students in the Adult Learning Center. The initial status of these students on measures of proficiency was to be assessed; their proficiency at a later point in time was also to be assessed. The original methodology planned was to relate the performances of the teachers to differences in acquired proficiency.

Several methods are available for achieving this goal. One method is to correlate different kinds of teaching performances with the final achievement scores, partialling out the initial status of students. The idea behind this method is to identify which particular performances are most highly correlated with final status. The statistical methodology, that of partial correlational methods, utilizes information about initial status. Usually the students' initial scores are highly correlated with their final scores, and this correlation needs to be taken into account if one is to assess the unique relation of one or more teaching performances to final status.

When this method is used, the partial correlations between teaching performances and final status are produced. Let us assume, for example, there is a teaching performance that has been observed which we will call Modeling (in this performance the teacher provides a verbal example of the desired speaking performance), and that a statistically significant correlation of .35 is obtained between this teaching performance and a measure of proficiency in speaking English. This correlation of .35 (which is not an actual finding, but only an example) is the association between Modeling scores and the achievement score when the correlation between the initial achievement score and the final achievement score has been controlled for through statistical means, that is, has been "partialled out." The appropriate conclusion to be drawn from data of this kind can be reduced to the following statement: the correlation between Modeling and achieved proficiency in speaking English is .35 when the correlation between the students' initial score and final score has been accounted for.

This statement is the literal empirical result. What it means is that we can predict differences in achievement from differences in the degree to which the teachers use Modeling as part of their repertoire of teaching performances. The magnitude of the correlation coefficient gives us an idea of how good a predictor the teaching performance is.

A typical way of describing the utility of the predictor is in terms of the amount of variance in final scores accounted for, or the amount of variance in the final scores that can be predicted from information about the predictor variable. One method used to estimate the utility of this predictor is to square the correlation coefficient and transform the square to a percentage; for example, a correlation of .35 squared is .13, or 13 percent. An appropriate statement is that this teaching performance, independent of the initial status of students, accounts for 13 percent of the variance in final scores.

The goal of the research methodology is to account for as much variance in the outcome measures as possible. In a study of this kind the researcher uses the information that he has available on teaching performance, but has no idea in advance how many and which of the predictor variables are likely to be "good" predictors (except when similar studies have already been done). So the empirical method is in essence an inductive analysis of the predictors of student achievement.

What this means in this particular instance is that a variable such as <u>Modeling</u> is a predictor; that is, with a significant correlation coefficient of .35, this variable accounts for about 13 percent of the differences in achievement among students at the end of the research period. Some other teaching performance or some other variable may account for more or less of the variance.

The goal is to measure the partial correlation of the predictors (teacher performances) and achievement (student proficiency scores) to see if any are statistically significant and to estimate how much of the differences among the students in final measured proficiency is accounted for by each of these predictors. The result would be a listing of the predictor variables that were significant and the amount of the variance for which each accounts. This approach is typical in most educational research on teaching.

A second approach is much like the first one except that it uses the basic information in a different way. Imagine a list of teaching performances such as modeling, eliciting, practice, positive and negative reinforcement, and so forth. Ordinarily the researcher will find correlations among these various performances. They are not truly independent, and practical educators are aware of this concept as evidenced by the fact that they frequently describe teaching in terms of styles or collections of performances which are characteristic of one teacher but not necessarily of another teacher. From a statistical point of view these intercorrelations ought not to be ignored.

Multiple regression methods are available that take into account the intercorrelations among variables. When a multivariate analysis is performed, the result is a multiple R, which is a correlation coefficient. This correlation coefficient is generated by entering all of the information available, for example, the information on the various kinds of teaching performances, into an analysis which yields a <u>prediction equation</u>. A prediction equation takes generally the form described below. We will use a simple example constructed to illustrate what is involved in multiple regression analysis.

Assume that we have four teaching performances: X_1--modeling; X_2--practice; X_3--positive reinforcement; X_4--negative reinforcement. These are four descriptors of teaching performance. For each of these variables we will have scores for each of the teachers participating in the study. We correlate their scores on each of these measures with student achievement and with each other. Assume that the correlation matrix looks like this (note that these are not actual correlations from this study):

	P_1	P_2	X_1	X_2	X_3	X_4
P_1	1.00	.78	.50	.34	.42	.10
P_2		1.00	.35	.20	.45	.35
X_1			1.00	.30	.32	.15
X_2				1.00	.70	.35
X_3					1.00	.15
X_4						1.00

$r \geq .35$, $p \leq .05$

P_1: one measure of oral proficiency
P_2: second measure of oral proficiency
X_1: modeling; X_2: practice; X_3: positive reinforcement;
X_4: negative reinforcement

We have chosen to use two measures of proficiency in this example, P_1 and P_2. The teaching performance variables are those listed above. The numbers in the table are the zero-order correlations (they are not the partial correlations) between the scores on each of these variables. Look in the upper left-hand corner of the table and read across from P_1. We see that the correlation between the two measures of proficiency, P_1 and P_2, is .78. A correlation of this magnitude is substantial and it means that the two measures are highly related to each other. (We have also indicated in the table that correlations above .35 are statistically significant.) Again, reading across the first row we see that the correlation between P_1 and X_1 (modeling), and between P_1 and X_3 (positive reinforcement) are statistically significant. The correlation between P_1 and X_2 (practice) is very close to significance, but the correlation between P_1 and X_4 (negative reinforcement) is not significant.

The way these data are interpreted, if one is looking only at these correlations, is that modeling and positive reinforcement are correlated significantly with the first measure of proficiency (P_1). When we translate these correlations into the percentage of variance accounted for (by squaring the correlation coefficients), we would say that X_1 accounts for 25 percent of the variance in the scores of P_1, and X_3 accounts for 18 percent of the variance in the P_1 scores.

Similarly, if we read in the second row from P_2 we note that X_1 (modeling), X_3 (positive reinforcement), and X_4 (negative reinforcement) are all correlated significantly with P_2. Again we can square each of these correlation coefficients to determine the percentage of variance accounted for by each of these variables: X_1 (modeling) accounts for 12 percent of the variance as does X_4 (negative reinforcement); but X_3 (positive reinforcement) accounts for 20 percent of the variance.

We have two predictors of P_1—X_1 and X_3; and three predictors of P_2—X_1, X_3 and X_4. But note that P_1 and P_2 are substantially correlated ($r = .78$) and that X_1 is correlated with X_2, X_3 and X_4 even though the correlations are not statistically significant. X_4 has a significant correlation with X_2. In general most of the teacher performance variables are intercorrelated well above a zero correlation. These relations suggest certain interdependencies among the performances which are revealed in the correlation matrix.

The methods of multiple regression analysis use the data provided in a matrix such as that presented above. The calculations utilize all the correlations among the variables. The technical analysis produces two basic pieces of information, (1) a multiple R and (2) beta weights. The multiple R is a correlation coefficient and is interpreted in the same way as the kind of correlation coefficients that we have previously discussed; that is, R^2, translated into a percentage, represents the percentage of variance predicted by all the predictor variables entered into the equation. From this figure we know how useful the array of predictor variables is in accounting for the variance in the dependent variable, in this case a proficiency score. The beta weights are standard partial regression coefficients which are ratios of various combinations of the original or zero-order correlation coefficients. These beta weights are then transformed into the coefficients in the prediction equation by multiplying the beta weights by the ratio of the variances of the variables involved. What each of these coefficients represents is the weight attached to a particular variable independent of the other variables; they are partial regression coefficients. These weights vary in magnitude and may be either positive or negative in sign. A coefficient which is negative in sign means that the variable subtracts from the prediction.

Consider now the following equation:

$$(1) \quad P_2 = b_o P_1 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$$

In this equation P_2 on the left-hand side of the equation is the set of scores in proficiency which are to be predicted by the variables on the right-hand side of the equation. Notice that P_1 is included in the prediction equation. The reason for including P_1 is that it is highly correlated with P_2 and is, therefore, a predictor of P_2. The X's represent the four teaching performance variables that we have been talking about, and the coefficients of these values are represented by the letter b. These coefficients are the regression weights that we discussed above. Each of these would have a numerical value.

and may have a negative sign. (Actual values of these coefficients would be calculated from the correlation matrix; the numbers in the matrix were made up and we have not actually calculated weights from that matrix.)

For this equation there will be a multiple R. Assume that for this equation the multiple R is .64; applying the method described above, we square .64; $R^2 = .41$; .41 x 100% = 41 percent. The empirical conclusion to be drawn from these data is that the four predictor variables (the four teaching performance variables) and the other proficiency variable predict 41 percent of the variance in the P_2 scores. Since 100 percent of the variance is to be accounted for, these five variables account for less than half of the variation in the scores.

You may ask what else is related to variation in the proficiency measure? Our equation only includes four kinds of teaching performances and a measure of initial proficiency. The remaining variance may be predicted by such factors as the previous experience of the students with English, how long they have been in the United States, the kinds of communities in which they are currently living, their educational attainments, and their past and present employment. If we entered measures of these variables into the prediction equation, first having found the correlation of such variables with the performance variables and the proficiency measures, we would find out if the addition of these variables accounts for some or all of the remaining variance.

If, alternatively, we assume that R for equation (1) is = .87; then $R^2 = .76$; .76 x 100% = 76 percent. In this case the teaching performance variables and the proficiency measure score predict 76 percent of the variance in the P_2 scores. This R^2 tells us that the predictor variables are very useful in predicting the variation in the proficiency scores (P_2).

Again, by adding other predictor variables more of the variance may be accounted for. In this example there is only one-fourth of the remaining variance to be accounted for, and some of this variance will be accounted for by unreliability in the measures of the predictor variables. (We will return to this point below.)

The R^2 value tells the researcher how useful his or her array of predictor variables are for predicting the differences among students on the achievement measure. As the reader can perceive, unless the researcher has knowledge from previous research or from theory, he or she could select a set of predictor variables which have little utility for predicting the variation on the outcome measure. In developing a research study in which multiple regression analysis methods are to be used, a rational way to proceed is to decide on those factors most likely to be related to the outcome measure. Even when previous knowledge suggests that a relationship between a predictor variable and an outcome measure is likely to be found, it may turn out that in a particular sample of observed data that the predictor variable does not relate to the outcome measure in the same way that it has in previous studies.

The logical method underlying the use of multiple regression analysis in research is inductive in character. The researcher enters as many variables into a prediction equation as is feasible in terms of the practical matter of gathering data on those variables. The regression equation tells the researcher how good the particular combination of variables that he or she has selected is in predicting the variance in the outcome measure. The signs of the coefficients of each of these variables tells him or her how the variable contributes to the prediction either additively or subtractively. The numerical size of the coefficient tells the researcher how much weight each variable has in the prediction; also the statistical significance of the regression weights tells him or her how reliable a predictor that variable is.

Let us illustrate these ideas with our made-up example; assume that the calculations yield the following equation (which is the preceding equation with numbers substituted for the b's).

(2) $$P_2 = 1.20\ P_1 + .98\ X_1 + 1.05\ X_2 + 1.25\ X_3 - .75\ X_4$$

In this prediction equation note the following: (1) the largest numerical weight is associated with P_1, the next largest with X_3 (positive reinforcement), the third largest weight is associated with X_2 (negative reinforcement). Assume that all of these coefficients are statistically significant.

What this equation tells us is that the best predictor of the measure of oral proficiency, P_2, is the other measure of oral proficiency (a result which is not surprising). Among the four performance variables we notice that negative reinforcement (X_4) contributes negatively to the prediction; that is, if we were using only this predictor variable, we would predict lower scores for students that had teachers who used more negative reinforcemement. In contrast, X_3 (positive reinforcemnt) has a larger and positive weight. If we considered only this variable, we would predict higher scores for students who had teachers that used more positive reinforcement.

But, of course, the purpose of using multiple regression methods is to consider all of the variables together and not confine the analysis to the predictive power of each variable by itself. What the prediction equation represents is a set of data from which one may theorize about the relative influence of a variety of factors. The logic of this inductive analysis proceeds as follows. First one looks at the size and significance of R^2. If R^2 is significant, the first requirement, then its magnitude is next considered. If the magnitude is not large, that is, a relatively smaller proportion of variance has been accounted for, we know that our predictor variables do not account for most of the variance in the outcome measure. One reasonable conclusion is that there may be other predictor variables which when entered into the equation improve the prediction. Even if R^2 is small, one may theorize about the variables that are in the equation; but obviously one is likely to develop an incomplete theory because the variables in the theory only account for a small proportion of the variability in the phenomena to be predicted.

If, in contrast, the multiple R is relatively large, then one has more confidence that the variables in the prediction equation represent a comparatively solid basis for theorizing about the nature of the phenomena involved in the prediction. Suppose, as we did above, that R^2 is large so that the prediction is accounting for about 75 percent of the variance. It is not unreasonable to infer that we have a set of variables which constitute a reasonable basis for developing a theory of instruction.

Referring back to the equation that we have been discussing, our theorizing might go something like this. Since P_1 has the largest positive regression weight, we may infer that the initial proficiency status of the student is a relatively critical factor in determining what the students' final achievement is likely to be. This idea makes sense since new learning is always built upon or intergrated into or reorganized with previous learning. In this instance we are studying the acquisition of language; we are reasonably certain on theoretical grounds that the level of acquisition previously attained is a base upon which subsequent acquisition might be developed. The empirical data, as represented in the prediction equation, corresponds to our theoretical expectation.

Next, we noticed that X_2 and X_3 both have large (in the context of this equation) positive regression weights. These two variables are the practice--positive reinforcement variables. We would speculate or hypothesize that teachers who provide more practice in using English and positive reinforcement for appropriate language responses have created an instructional model which is likely to be associated with higher achievement scores. But we also noticed that X_1, the modeling variable, has a positive regression weight near in magnitude to that of X_2. We infer that the provision of modeling is a factor which also significantly affects predicted scores.

At this point in the inductive analysis we are reasonably sure which variables are likely to be associated with higher scores and which are not. Now we should consider the limitations of the method to prevent ourselves from drawing inferences that are too strong and that go beyond the power of the method. We are making inferences from one study. The generalizability and stability of the results are unknown. But, since in this case our inferences also make theoretical sense, the risk in applying the results does not seem great.

We might also be tempted, for example, to say that practice is the most important variable (after that of the students' initial status) because it has the largest regression weight. But we must remember that we have only this one prediction equation. It has not been cross-validated in other studies, and we know that these regression weights like other statistical phenomena vary in their magnitudes in different samples. So we cannot conclude that positive reinforcement is the most critical variable and that modeling is relatively less important. All we can say is that we have a basis for hypothesizing that three of the performance variables are positive predictors and that a fourth variable is a negative predictor. Note, however, that these variables are also the critical variables in social learning theory and reinforcement theories. Therefore, we can have reasonable confidence (from theory and our data) that we have identified variables which may be critical in the acquisition of proficiency in speaking English.

We have spent several pages going through the concepts of the methods of multiple regression because they are the methods which we have used in the research reported here. While we have stressed in the more technical chapters the limitations and qualifications of the results of the analyses, we have taken pains here to illustrate how, within those empirical limitations, one can begin to reason to a set of ideas or hypotheses or even a theory about the nature of the instruction that is likely to be effective.

Applying the Results of the Research

The obvious question for the practitioner who wishes to use the results of an empirical study of this kind is "How do I apply these results to design instructional strategies?" Should the practitioner, for example, recommend to teachers of English as a second language that they use more positive reinforcement in their teaching, or that they increase the amount of modeling they do for students? Underlying these questions is an epistemological question. May we make causal inferences from correlational data of the kind produced in a multivariate analysis of teaching performance?

A correlation between two variables does not imply that one variable is necessarily the cause of the other. All that we learn from the correlation is that the two variables are associated. When we use partial correlational methods, however, we are separating out the relative influences of several variables. When we partial out initial proficiency scores to find the correlation between a teaching performance and final proficiency scores, we estimate the relative influence of initial proficiency and teaching performance. We still cannot make a strict causal inference that the teaching performance variable "causes" the change from initial status to final status, but by a process of reasoning we can estimate the plausibility of this possible causal relation. We know that the proficiency of a student at the time instruction begins was acquired before the actual instruction. Therefore, when we find a teaching performance that is correlated with final status scores (after partialling out initial scores), we may infer that the teaching performance has added to proficiency over and above what the student initially has acquired.

This inference, however, is only one of several that may be made. It may be that characteristics of the students help them to be more responsive to instruction and this responsiveness to the teaching performances accounts for the gain in proficiency. Or, there may be a relation between the level of performance acquired prior to instruction and the effectiveness of particular methods of instruction.

The methods used in this study should help sort out the likelihood that one of these inferences is more valid than another. But strictly speaking no conclusion about the relative validity of these different hypotheses can be drawn from the data of the study itself. Only by a process of analysis and reasoning can one estimate a likelihood of one explanation being more valid than another, but the conclusion about such an interpretation is a hypothesis which needs to be tested.

The only way we can really know which predictor variables directly produce a change in proficiency is by creating an experiment in which these predictor variables are manipulated in a systematic way. The next step, after a study of this kind, would be to design an experiment in which teachers are trained to use those teaching performances which are significantly correlated with final proficiency, and in which the effect of these increases in the teaching performances on proficiency can be measured. Only if we conduct a study in which we systematically control some of the variables so that we can estimate the effects of other variables, can we draw a strict conclusion about which variables have a causal effect upon proficiency and which do not, and the relative influence of each potential causal variable.

The practitioner may not always be in a position to conduct such an experiment or to have it conducted for him or her, and therefore wants to know how they should proceed when they have received data of the kind produced in this study. One way to proceed is to make inferences about potential causal relations and to act on these inferences. If we have found, for example, that a particular pattern of teaching performances is correlated with outcome measures we can urge teachers to use these methods or train them to use them, and then gather data systematically to see how much improvement in proficiency is actually achieved.

At the end of this chapter, after we have presented the major results of this study (which are more complex than the examples we have been using) we will make some suggestions about practical action. For the more technically-minded reader, we state that we are obviously reasoning beyond the data to make inferences about potential causality which might lead to programs of action. For the practitioner we are suggesting what might be tried and ways of estimating whether, when these actions are taken, they in fact have a significant effect.

The Selection of Predictor Variables

In performing a study of the kind described here the first step is to select variables likely to be predictors of proficiency in speaking English. Since there is no formal theory that can guide us in the selection of these variables, but there is some relevant information and some generally accepted ideas, the investigator proceeds by collecting as much information as is feasible about a range of predictor variables. We are reasonably certain, for example, that the initial proficiency of the student is likely to be related to final status because this result has been found in a very large number of studies on instruction. This selection also makes theoretical sense because the acquisition of language is in part the acquisition of discrete responses which are progressively integrated into more complex patterns of responses; that is, the learner acquires new responses to some degree by using previously learned responses.

We would also expect characteristics of the students to be related to how well they learned. Unfortunately in studies of this kind it is not usually possible to gather as wide a range of information on students' characteristics as might be desirable. Frequently, as in this case, we have to settle for descriptors of characteristics which are proxies for underlying characteristics. A student's motivation to acquire new language, for example, is likely to be an important variable, but it is not easy to measure this motivation directly. Therefore, we resort to gathering information about the student which are indirect measures of motivation such as the amount of education they have attained, or whether or not they studied English in their country of origin. These measures are obviously indirect measures but a reasonable case can be made that a student who proceeds further in the educational system probably has more motivation to be educated or has attitudes towards education that stimulate them to seek more learning. The reader will recognize the relative weakness of a proxy of this kind by thinking that a student who has not progressed very far in the educational system may not have had the economic resources to do so. But we are using the indicator to find out if a variable that may be an indicator of motivation is related to student performance. If it is, further exploration can be made to pinpoint the precise nature of the underlying characteristic which may be affecting the student's responsiveness to instruction or which helps him or her to become more proficient.

In this study we have used three major categories of predictor variables. The first of these is the initial performance of students, that is, how they scored on different measures of proficiency in English at the beginning of the research period. The students enter the Center with varying degrees of ability to speak English. They, in fact, are placed in language classes on the basis of their measured performance in understanding and usage of English.

We have, therefore, included among the predictor variables those measures of proficiency that were used at the beginning of the research period or that were used when the students were admitted to and placed in classes in the Center.

Another class of variables are those which describe the previous experience of the student. Since this study uses adults as its sample of students we know that we are studying individuals who have a relatively long learning history. It is reasonable to expect that various kinds of influences on that learning history would be related to how the students acquire greater proficiency while in the Adult Learning Center.

We collected data on a variety of background factors such as the amount of education that the students had, the amount of direct experience with English, both here and in their country of origin, their age, the kind of employment they held in their country of origin and that they may be holding here. We would expect variables like these to be indicators of previous learning history. Amount of education, for example, is a proxy for such variables as ability and motivation. We know that students who proceed further in the educational system have to have achieved successfully as they progress through this system. We also assume that those who have proceeded further in the system, particularly when the age for compulsory education is lower in the country, may have more motivation to become educated (allowing for differences in economic resources to allocate to one's education). Note that we are not claiming that such variables are causes of proficiency in speaking English or causes of educational motivation or attainment. We are simply saying that these are indicators or proxies worth entering into a multiple regression analysis to see which predict acquired proficiency in English.

The third category of predictor variables are those related to how the teachers actually teach. We obtained these data by directly observing the classes in the Adult Learning Center. Our assumption was that different methods may be more or less effective in facilitating the acquisition of proficiency in speaking English.

In this study, of course, we may analyse only those teaching performances which actually were observed. It is conceivable that there are other ways of teaching English as a second language that might be more effective than the methods that we have observed. Thus, one of the major limitations of this kind of study is that we can find out only if the observed performances are more or less effective.

In selecting these major categories of predictor variables we have picked those most likely to be predictors of subsequent proficiency. There may be other variables that are more powerful that we do not presently suspect to be powerful predictors. We have chosen those which on the basis of previous research and theoretical knowledge are most likely to relate to learning of this kind.

Major Questions of This Study

The major question of this study is, "How much difference do the methods of teaching make in influencing the acquisition of proficiency in speaking English?" As we have pointed out, we cannot give a strictly causal answer to that question; the question we do answer is, "How well do teaching performance variables predict proficiency when we allow for the influence of other variables on proficiency?"

We are also interested in identifying those teaching performances which are most effective. Teaching performances are complexes of a variety of behaviors or actions which are organized in many different ways by each teacher. In the Adult Learning Center we observed teachers who frequently modeled the desired speaking performance for the students; we observed teachers who asked questions that stimulated the students to practice structures of the English language; we observed varying amounts of corrective feedback; we observed teachers who stimulated students to speak English discursively and did very little modeling of the appropriate responses. Are some of these ways of teaching more effective than others?

A third question is related to the consistency of the procedures used by the teachers. No teacher teaches in exactly the same way on every occasion. But there are two aspects to this variability of the teachers' performances. A teacher may have an overall style which varies somewhat from occasion-to-occasion. Or, a teacher may have combinations of different kinds of styles which he or she organizes differently from day-to-day. In the field of teaching English as a second language there are at least two major methodological styles, the audiolingual method and the "silent way." These two styles in their ideal form differ radically. We know that different teachers in the Center prefer one style to another, but there are also teachers at the Center who use mixtures of these styles.

A major problem in this study was to identify the characteristic styles of each teacher and to study how consistent they were within a style or how they combine styles and how they varied their teaching from occasion-to-occasion. It is possible that there is so much variability in a teacher's style that it would not be possible to identify any one teaching performance or even a collection of teaching performances which had a systematic effect on proficiency.

(The data presented in Chapter III are the results of our analyses of the teaching styles used in the Center. That chapter also contains information on the consistency or variability in the styles used by the teachers.)

Although the major goal of this study was to study the teaching performances or methods of instruction, it was also necessary to ask questions about the influence of other variables on acquired proficiency. For this reason, as we noted above, we have included in our analyses other potential predictor variables related to the characteristics of the students, both their background characteristics and their initial status as measured at the beginning of the research period. Two major questions are, therefore: "How well do descriptors of student's background characteristics predict acquired proficiency?" and "How well does the student's initial proficiency predict his or her final proficiency?"

As can be seen, the major questions of the study ask which of the various categories of variables predict best the final proficiency of the students. It will be useful to consider at this point the potential implications of various kinds of answers so that the reader can begin to anticipate how to use the data as he or she progresses through the summaries of it in this chapter.

Possibilities for Applying the Research

Consider how the results of this study may have come out and what the implications of these results might be. Suppose that teaching performances were found to be nonsignificant predictors of acquired proficiency. Would this result imply that the methods used in the Center were ineffective? Not necessarily. It may be that the methods, while they "added nothing" to the prediction of proficiency may in fact be the methods required for the characteristics of the students to have their effect on acquiring proficiency. We cannot automatically assume that because the methods being used are not good predictors of final proficiency that they were not useful in some other way.

Suppose that we find that no single style of teaching is particularly effective, that is, none of them is a good predictor of final acquired proficiency. This result seems to have more practical meaning because it suggests that strict adherence to any one method is not in and of itself likely to be sufficient to produce a substantial effect. Unfortunately, the method of the study does not permit us to draw a strict conclusion about the effectiveness of one method over another for several reasons: (1) there are very few instances of a method being used in its ideal form; (2) there are too few instances of teachers adhering to a method to make a strict comparison. At best we can say from this study that adhering more or less strictly to a method did or did not produce greater proficiency in speaking English, but this result may well be contaminated by characteristics of the teacher which interacted with the method.

The staff of the Center may examine the purer forms of a method being used by particular teachers to see if the use of the method could be improved. Or the staff can study other characteristics of the teachers which may be interacting with the use of the method.

Another major possibility is that the characteristics of the students are the best predictors of their acquired proficiency. This result leads to several different kinds of possibilities. One possible course of action is to find out if teaching methods interact with characteristics, that is, are there some teaching performances or methods which in combination with characteristics of the students are associated with greater achieved proficiency? If such were the case, one could be more prescriptive about the organization of classes and the appropriate methods to be used in the different classes.

These are different ways in which the results of the study may come out, and what the implications might be for action if such results were obtained. The reader should be aware that in this type of study we are trying to estimate which variables predict best acquired proficiency in speaking English. From the data produced by the study, it will not be possible to draw strict causal inferences that these variables in fact made the critical difference in the acquisition of proficiency. But the results will suggest some possibilities for action. These possibilities should be regarded as hypotheses about or theories of instruction that might be tested subsequently.

We now turn to summarizing the details of the study itself, and at the end of the chapter we will again engage in an analysis of the possible meanings of these results and the possibilities for practical action.

THE SAMPLE OF CLASSES STUDIED

The Adult Learning Center in West New York provides training in English for students whose native language is some other language. The students are adults, as the name of the Center indicates, ranging in age from 19 to 73. The majority of the students are immigrants from Cuba; most of the other students are from Caribbean and South American countries, so that the native language of over 90 percent of the students is Spanish.

The classes are conducted both during the day and at night. The day-school classes are held at the Adult Learning Center which is located in an office building in down town West New York. The night-school classes are conducted at Memorial High School in West New York. The class schedule follows the public school calendar.

The majority of classes in the day school met for an hour-and-a-half a day, five days a week. Of the night-school classes studied five met for a two-hour session three evenings a week; the sixth class met only two evenings a week.

All of the classes in the day school were studied, with one exception which was dropped from the study because the teacher transferred to another position. There were 14 day-school classes taught by six different teachers.

There were 22 classes available for study in the night school of which six different classes with six different teachers were selected by applying the following criteria: (1) the teachers did not also teach in the day school; (2) their classrooms had a sufficient number of "representative students"; (3) there was a range among classes in the level of student proficiency in speaking English; and (4) there was a range among teachers in previous training and experience. The representativeness of the classes was assessed by gathering data on the students, then describing each class in terms of these characteristics, and then ~~finding~~ a set of classes where the class had the modal characteristics of the students in the entire set of 22 classes.

Students provided background information on 10 characteristics: their sex, whether or not they had received a high school diploma, their age, the educational level attained, the amount of time they had been in the United States, the amount of English they had studied in their former country, the amount of English they had studied in the United States, their country of origin, their occupation in their former country, and their occupation in the United States. Table 2.1 and Table 2.2 provide the summary statistics on these characteristics of the students. (These tables are reproduced from Chapter 2, and the reader may refer to that chapter for a more detailed discussion of the students' background characteristics.)

The background information was collected in the day school on 148 students at the beginning of the study; however, a number of these students dropped out of the Center before the completion of the study (there were 81 of the original students at the completion of the study; the attrition rate was 45 percent between November and June for day-school students). Table 2.1, therefore, presents two sets of data. The data in Table 2.1 under the heading "<u>Matched</u>" is the data on students who were present during the entire period of the study, that is, from November to June. A comparison was made between the original sample of students present in the Center in November, and the sample remaining in June. No significant differences between the two samples were found.

TABLE 2.1

Day School Student Background Information--
Pretest and Matched Samples: Descriptive Statistics

	Pretest: Total Freq.	Pretest: Total Percent	Pretest: Freq. by Level 1	Pretest: Freq. by Level 2	Pretest: Freq. by Level 3	Matched: Total Freq.	Matched: Total Percent	Matched: Freq. by Level 1	Matched: Freq. by Level 2	Matched: Freq. by Level 3
Number	148		52	42	54	81		24	28	29
Sex: M	54	36	22	21	11	26	32	9	13	4
F	94	63	30	21	43	55	68	15	15	25
Diploma: Yes	63	43	17	15	31	32	40	8	10	14
No	83	56	34	27	22	49	60	16	18	15
Age: Mean	45		46	44	44	46		48	45	45
Range	19-73		21-69	19-70	24-73	19-70		21-69	19-70	24-67
Education:										
Mean Years	10		9	10	12	10		9	10	11
Range	3-17		3-16	4-16	4-17	4-16		4-16	4-16	4-16
Time in U. S.										
Mean Years	6		5	6	7	6		6	6	7
Range	1-24		1-15	1-22	1-24	1-24		1-15	1-17	1-24
Former English										
Mean Years	1.08		.58	.83	1.83	1.17		.83	.75	1.83
Range	0-12		0-5	0-6	0-12	0-12		0-5	0-2	0-12
English in U. S.										
Mean Years	1.08		.50	1.41	1.33	1.17		.38	1.25	1.50
Range	0-12		0-1.5	0-12.0	0-3.0	0-3.0		0-1.5	0-2.5	0-3.0
Former Country										
Columbia	8	5	4	4	0	3	4	1	2	0
Cuba	123	83	42	35	46	73	90	22	25	26
Dominican R.	3	2	2	0	1	1	1	0	0	1
Ecuador	2	1	1	0	1	1	1	1	0	0
Peru	4	3	3	0	1	0	0	0	0	0
Other	8	5	0	3	5	3	4	0	1	2

Occupation*	Pretest T	Pretest 1	Pretest 2	Pretest 3	Matched T	Matched 1	Matched 2	Matched 3
1. Foreman	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
2. Craftsman	7/4	3/1	3/2	1/1	1/0	1/0	0/0	0/0
3. Semi-skilled Worker	14/44	7/15	7/16	0/13	6/20	4/4	2/7	0/9
4. Laborer	0/1	0/0	0/0	0/1	0/0	0/0	0/0	0/0
5. Household Worker	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
6. Personal Service	5/11	2/3	2/5	1/3	2/7	0/2	2/5	0/0
7. Fireman/Policeman	0/1	0/0	0/1	0/0	0/1	0/0	0/1	0/0
8. Professional	38/4	6/0	12/0	20/4	23/1	5/0	8/0	10/1
9. Technician	5/2	1/0	1/0	3/2	4/1	0/0	1/0	3/1
10. Farmer	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
11. Farm Worker	2/0	2/0	0/0	0/0	1/0	1/0	0/0	0/0
12. Business Owner	2/0	2/0	0/0	0/0	1/0	1/0	0/0	0/0
13. Manager/Official	2/1	0/0	1/0	1/1	1/1	0/0	1/0	0/1
14. Office Worker	24/9	8/0	8/4	8/5	19/7	7/0	6/4	6/3
15. Salesperson	8/7	5/1	1/3	2/3	4/6	1/0	1/3	2/3
16. Housewife	28/32	12/14	4/2	12/16	13/17	3/5	4/2	6/10
17. Unemployed	0/29	0/15	0/9	0/5	0/19	0/12	0/6	0/1
18. Student	9/1	1/1	3/0	5/0	6/1	1/1	3/0	2/0

*Previous country/USA
See Appendix A for definitions of categories.

As can be seen by scanning Table 2.1 there were more women amoung the day-school students than men, more did not have a high-school diploma than had one; the mean age was approximately 46 years and the mean level of education attained was about that of the tenth grade. On the average, students had been in this country six years and had studied English in their former country and in the United States on the average for about one year. As previously indicated practically all of the students came from Caribbean and South American countries, predominantly from Cuba. The occupations held by the students in the former country and in the United States were mainly lower middle class and lower class occupations, though there were a number of students who had had professional occupations in their former country.

There were also more females than males in the night-school sample (recall that in reading the numbers in Table 2.2 that they are the numbers of students in the six classes that were studied, but the proportions represent the proportions in the night-school sample as a whole). As in the day school, a large number of students did not have a high-school diploma and the mean age was about 42 years; the students also had completed on the average 10 years of schooling; had been in this country about six years and had about a year-and-one-half of English both in their former country and in the United States. Again the majority of students came from Spanish-speaking countries.

The day-school and night-school samples of students are similar but not identical. There is a higher percentage of males in the night-school (44 percent versus 36 percent); fewer students have their high-school diplomas (17 percent versus 43 percent); and the average age is somewhat younger (42 versus 45). The night-school students have studied English in their former

TABLE 2.2

Night School Student Background Information --
Matched Sample: Descriptive Statistics

	Frequency	Percent	Frequency By Level 1	2	3
Number	41		21	14	6
Sex: M	18	44	8	9	1
F	23	56	13	5	5
Diploma: Yes	7	17	3	2	2
No	34	83	18	12	4
Age: Mean	42		44	42	37
Range	17-64		17-64	22-56	30-45
Education: Mean Years	10		7	11	12
Range	3-18		3-18	6-16	8-16
Time in U. S.					
Mean Years	5.9		5.5	5.3	8.8
Range	1-15		1-15	1-13	4-13
Former English					
Mean Years	1.17		1.00	1.08	2.0
Range	0-8		0-7	0-8	0-5
English in U. S.					
Mean Years	1.67		1.17	2.41	1.75
Range	0-5		0-3	0-5	0-4
Former Country					
Argentina	2	5	0	1	1
Columbia	3	7	0	2	1
Cuba	25	61	15	7	3
Ecuador	4	10	3	1	0
Guatemala	2	5	0	2	0
Other	5	12	3	1	1

Occupation*	Total	1	2	3
1. Foreman	0/0	0/0	0/0	0/0
2. Craftsman	4/7	3/4	1/2	0/1
3. Semi-skilled Worker	4/24	1/15	3/8	0/1
4. Laborer	1/2	0/1	0/1	1/0
5. Household Worker	0/0	0/0	0/0	0/0
6. Personal Service	3/2	2/1	0/0	1/1
7. Fireman/Policeman	0/0	0/0	0/0	0/0
8. Professional	6/0	1/0	3/0	2/0
9. Technician	1/0	1/0	0/0	0/0
10. Farmer	0/0	0/0	0/0	0/0
11. Farm Worker	0/0	0/0	0/0	0/0
12. Business Owner	0/0	0/0	0/0	0/0
13. Manager/Official	1/1	0/0	1/1	0/0
14. Salesperson	1/0	1/0	0/0	0/0
16. Housewife	7/1	6/0	1/1	0/0
17. Unemployed	1/0	1/0	0/0	0/0
18. Student	3/0	2/0	0/0	1/0

*Previous country/USA
See Appendix A for category definitions.

country approximatley six months longer (1.67 years versus 1.17 years). Although the majority of students were from Cuba (61 percent), there were not as many Cubans in the night school as in the day school where Cubans were 90 percent of the sample. More housewives and unemployed persons attended day-school classes.

The picture that emerges from these data is of an adult immigrant population. The age range is considerable, but the group is on the average a middle-aged group. The average level of education is not high; most of the people are either housewives or unemployed, and those who are working are by and large working in semi-skilled occupations.

Teacher Characteristics

Both day- and night-school teachers filled out a questionnaire asking for the following information: age; sex; undergraduate institution; undergraduate major; graduate institution; graduate major(s); number of credits completed; degree(s) completed; number of years teaching; grade level of teaching experience; number of years teaching ESL part-time (night-school adults); number of years teaching ESL full-time to children, to teens, to adults; number of ESL seminars; number of ESL workshops; usefulness of undergraduate experiences; usefulness of graduate experience; usefulness of ESL seminars and workshops; percent of teaching which is adulolingual, silent way, and other; and a statement describing their teaching methods and philosophy. Table 2.7 presents the summaries of this information for the day-school and the night-school samples.

TABLE 2.7

Teacher Background Information

	DAY SCHOOL (N = 6)*	NIGHT SCHOOL (N = 6)*
Age	$\bar{X}$: 35 Range: 24-45	$\bar{X}$: 35 Range: 23-55
Sex	5 Females; 1 Male	3 Females; 3 Males
Undergraduate Institution	Jersey City State: 2 Douglas: 1 Ladycliff: 1 Montclair State: 1 Seton Hall: 1	Jersey City State: 3 N.E. Missouri: 1 Saint Peter's: 1 St. Elizabeth's: 1
Undergraduate Major	Elementary Ed.: 2 Spanish/Ed.: 2 Philosophy: 1 Italian: 1	Elementary Ed.: 4 History: 2
Graduate Work Institution (1st. MA)	Jersey City State: 2 Fairleigh Dickinson: 1 Montclair State: 1 Rutgers: 1 Seton Hall: 1	Jersey City State: 3
Graduate Work Major	Elementary Ed.: 1 ESL: 1 ESL and Ed.: 1 ESL and Reading: 2 Guidance and Personnel: 1	Elementary Ed.: 1 Reading: 2
Number of Credits Completed	N = 6; $\bar{X}$: 30, Range: 9-47	N = 3, $\bar{X}$: 32, Range: 30-34
1st Master's Degree Completed	N = 3	N = 3
Graduate Work Institution (2nd MA)	Rutgers: 1	Jersey City State: 1 Montclair State: 1 Fairleigh Dickinson: 1
Graduate Work Major	Italian Lang. & Lit.: 1	ESL: 2 Elementary Ed.: 1
Number of Credits Completed	N = 1, $\bar{X}$: 27	N = 3, $\bar{X}$: 26, Range 15-34
2nd Master's Degree Completed	N = 0	N = 2

** The N for each category = 6 unless otherwise stated.

TABLE 2.7 (Continued)

	DAY SCHOOL*	NIGHT SCHOOL*
Number of Years Teaching	$\bar{X}$: 5 Range: 1-18	$\bar{X}$: 10 Range: 1-26
Teaching Experience Level	Elementary: 3 High School: 1 Adults: 2	Elementary: 4 Junior High: 2
Years ESL at Night, Adults	N = 2, $\bar{X}$: 8, Range: 3-12	N = 6, $\bar{X}$: 4, Range: 1-6
Years ESL Children Full Time	N = 1, $\bar{X}$: 4	N = 3, $\bar{X}$: 4, Range: 1-7
Years ESL Teens Full Time	N = 0	N = 2, $\bar{X}$: 4, Range: 3-4
Years ESL Adults Full Time	N = 6, $\bar{X}$: 2, Range 1-4	N = 0
Number of ESL Seminars	$\bar{X}$: 4, Range: 0-8	$\bar{X}$: 1, Range: 0-5
Number of ESL Workshops	$\bar{X}$: 17, Range: 4-30	$\bar{X}$: 13, Range: 2-25
Usefulness Undergraduate Experience**	$\bar{X}$: 4, Range: 2-5	$\bar{X}$: 3, Range: 1-4
Usefulness of Graduate Experience**	N = 6, $\bar{X}$: 3, Range: 1-4	N = 3, $\bar{X}$: 2, Range: 1-3
Usefulness of ESL Seminars and Workshops**	$\bar{X}$: 1, Range: 1-2	$\bar{X}$: 1, Range: 1
Percent Audio-lingual Method	N = 2, $\bar{X}$: 62%, Range: 60-65%	N = 6, $\bar{X}$: 60%, Range: 10-90%
Percent Silent-Way Method	$\bar{X}$: 66%, Range: 30-99%	$\bar{X}$: 35%, Range: 10-60%
Percent Other Methods	N = 5, $\bar{X}$: 14%, Range: 1-30%	N = 1, $\bar{X}$: 30%

*The N for each category = 6 unless otherwise stated.

**Very useful = 1, not at all useful = 5.

TABLE 2.7 (Continued)

	DAY SCHOOL*	NIGHT SCHOOL*
Concepts Underlying Method & Philosophy:		
Audiolingual	N = 2	N = 4
Cognitive	N = 1	N = 0
Conversational	N = 0	N = 1
Counseling-learning	N = 1	N = 0
Ecletic	N = 3	N = 0
Flexible-adjust to class	N = 3	N = 2
Group work	N = 0	N = 1
Individualize	N = 1	N = 0
Listening and speaking	N = 1	N = 1
Need for English	N = 0	N = 1
Peer teaching	N = 0	N = 2
Reading and Writing	N = 1	N = 1
Silent way	N = 4	N = 4
Situational	N = 3	N = 3
Student dominated	N = 4	N = 0
Student responsibility	N = 2	N = 1
Variety	N = 2	N = 1

*The N for each category = 6 unless otherwise stated.

The two samples differ in several ways. The majority of day-school teachers were women while only half the night-school teachers were. All six day-school teachers had completed some graduate work, three had completed a Master's degree and one was working on a second degree; while only three of the night-school teachers had done graduate work--all three had completed a Master's degree and two of them had completed a second Master's degree. Night-school teachers on the average have had more experience in teaching, but this apparent difference is due primarily to one teacher who has taught for 26 years. Most of the day-school teachers' experience with ESL had been teaching adults; whereas, most of the night-school teachers also taught ESL to children or teens during the day as well as to adults in the evening.

The teachers' statements about their teaching methods reveal differences between the two groups. Only two day-school teachers said they used the audiolingual method of teaching and estimated that 60-65 percent of their teaching was of this kind. But all of the night-school teachers stated that they used the audiolingual methods and their estimates of use ranged from 10-90 percent. The two groups differed in philosophy in two respects: day-school teachers more frequently stated that they believed that teaching styles should be eclectic and that students should dominate classroom interaction.

THE MEASURES OF STUDENT PROFICIENCY IN ENGLISH

The measures of proficiency in English were administered at two different times: for the day school in November, and again during April and May; for the night school in March and early April and again in June. The reason for two administrations was to estimate the change in proficiency between the two testing times. The first test measured how proficient the students were near the beginning of the instruction that they were receiving that year or semester; the second administration measured how proficient they had become after six months of instruction (day school) or three months (night school). Test dates and procedures are discussed in Chapter II.

The Oral Proficiency Test

The major goal of the Center is to increase proficiency in speaking English. Therefore, the most appropriate measure of the effects of instruction would be a measure of the student's ability to speak English. After a search for and examination of available procedures we concluded that none of the procedures that we had examined met the requirements of a standardized testing procedure. Consequently, we developed an Oral Proficiency Test.

The first step in developing the Oral Proficiency Test was to gather from the teachers in the Center a list of objectives for each of their classes. Each teacher provided 20 objectives and across the set of day-school teachers a composite list of 132 objectives resulted. These 132 objectives were then given to the day-school teachers who rated each objective on (1) whether or not it was something they taught; (2) how important they felt it was for proficiency; (3) how much emphasis they gave it in their classes; and (4) how difficult they thought it was to learn. These objectives and ratings were then used to develop the content of the Oral Proficiency Test.

The test itself was an individually administered test that required the student to speak in English. It had three kinds of items. The first set of items were based on representative examples of the teachers' objectives for each level of proficiency. These items were in a structured, conversational format and arranged in order of difficulty by objective. The second type of item involved presenting the student with action pictures; the student was asked to describe what was occurring in the picture. Both types of items tested the student's ability to generate language freely. The third type of item consisted of three Spanish cartoons and the student was asked to state in English what was occurring in the cartoons. The purpose of this type of item was to see if the student could go from idiomatic Spanish to idiomatic English. The test was piloted and then revised on the basis of this experience.

The tests were administered individually by trained testers. A recording was made of the testing. The test itself was administered to the students near the end of the school year (it was not used as a pretest). The test yielded three kinds of scores. One score was for comprehension, meaning that the student gave evidence of understanding the questions asked of him or her but could not produce a correct English response. The second score was given for selecting the appropriate structure to use in a response even though the student made other errors in responding. The third score was for correctness as well as use of appropriate structures. Thus the first score represents a level of proficiency in which the student appears to be understanding the English spoken to him or her but responds inadequately. The second score represents a higher level of proficiency, one in which the student can respond and does so by choosing correct structures but whose response is not entirely correct. The third score represents the highest level of proficiency in which the student responds in an entirely appropriate and correct manner.

The Literacy Test

A potential side effect of learning to speak English is that students may also learn to read it. They are exposed outside of the school to English newspapers in English, directional signs in English, and a variety of other printed matter in the English language. As they acquire words in the language and understand how to speak the language, they may also acquire the ability to read it. Furthermore, during the teaching of English as a second language words are written on the board and printed materials are given to the students as a basis for conducting conversations.

Therefore, we used a Literacy Test which measures the functional reading level of the student. The materials in the test consist of pictures of signs, labels on bottles, forms, tables and so on.

There were 50 such items in the Literacy Test adapted from an original set of 170 items which had been developed to measure functional literacy of adults who spoke English.[1] The items selected for this study were pretested with two students from the Center, one of whom had little or no English proficiency and an advanced student with considerable proficiency. After this pilot testing, the test was revised and shortened.

Instructions for the test were given in either Spanish or English depending upon the students' level of proficiency. The question for each item was read to all students in English. They then read the item and indicated the answer to the question by circling the appropriate word or sentence. Testing was stopped when the student answered five consecutive items incorrectly. (The items were arranged in order of their difficulty for English-speaking adults.) The student score was the number of correct answers.

[1] Murphy, R. T. Adult Functional Reading Study. Educational Testing Service, Princeton, New Jersey, 1973.

Oral Decoding Test

We thought that students might also acquire decoding skills as a result of learning to speak English. Obviously if students acquire general skills, such as decoding, their language proficiency shoud increase rapidly, and they can become independent learners of words and structures. We administered an Oral Decoding Test developed originally to measure the decoding skills of young children.[2] Forms of this test at two difficulty levels were tried out on the same two students who had taken the literacy items and the results suggested that the more difficult test should be used.

Instructions were modified so that the item stem was read to the student rather than having the student read it to himself or herself. Instructions were prepared both in English and Spanish; however, test items were read only in English. This test was also administered as a pretest and posttest. The score on the test was the number of correct answers.

Other Measures of Proficiency

Two other measures of proficiency were used as pretests, the John Test and the Morano Test. The Center had been using both of these tests to estimate students' initial proficiency in order to place them in classes.

<u>John Test</u>

This Oral Proficiency Test (developed by Linda Kunz of Hunter College) consists of eight pictures about which the student was asked 22 questions. The student is given a comprehension score and is rated by the tester on fluency, use of structures, pronunciation and vocabulary.

[2] Developed by Robert and Kathryn Calfee, Stanford University, for the Beginning Teacher Evaluation Study conducted for the California Commission for Teacher Preparation and Licensing by Educational Testing Service. Frederick J. McDonald, Project Director.

Morano Test

The Morano Test is a paper-and-pencil test of recognition of correct use of English grammar. It has 50 items each consisting of three sentences expressing the same idea but only one of which is grammatically correct. The student is instructed to read the items and indicate the correct sentence. This test was also administered as a pretest.

Thus there are five kinds of tests all of which have face validity for measuring proficiency in English. But of the five tests used, only two are direct measures of proficiency in speaking English, the Oral Proficiency Test and the John Test. The difference betw-en these latter two tests is that the Oral Proficiency Test samples some different aspects of speaking English such as free production of language. Two of the other tests, the Morano and the Oral Decoding Test measure skills presumably related to proficiency in speaking English. The Literacy Test measures proficiency in reading English at the functional level. Proficiency on this test would be regarded as a beneficial side effect in the perspective of this study.

Student Performance on the Measures of Proficiency

Table 2.3 presents the data for both samples of students, day school and night school, on these measures. For each test information is given on the number of students taking the test, their mean score, the range of scores, the standard deviation of the scores, and the reliability of the test. As can be seen by examining Table 2.3, all the tests have high reliability, the lowest reliability being in the .70's for the Oral Decoding Test.

TABLE 2.3

STUDENT TESTS: DESCRIPTIVE STATISTICS

Instrument	Who	When	N	$\bar{X}$	Range	SD	Reliability
Literacy	Day	November (Pre)	148	29.82	3-49	12.52	.96
	Day	April (Post)	119	37.91	11-49	7.22	.89
	Night	March (Pre)	45	34.73	4-49	9.70	.93
	Night	May (Post)	43	36.81	11-48	9.98	.94
Aural Decoding I	Day	November (Pre)	148	26.72	6-38	7.40	.88
	Day	April (Post)	120	28.42	13-37	5.33	.81
	Night	March (Pre)	45	29.47	14-37	6.90	.88
II	Day	November (Pre)	148	47.84	31-57	5.28	.74
	Day	April (Post)	120	52.45	36-59	4.63	.77
	Night	March (Pre)	45	49.33	47-59	4.86	.73
Total	Day	November (Pre)	148	74.57	49-93	11.03	.89
	Day	April (Post)	120	80.88	56-95	7.90	.83
	Night	March (Pre)	45	78.80	60-96	9.47	.87
Proficiency Comprehension	Day	June (Post)	113	30.48	2-58	14.66	.96
	Night	June (Post)	43	27.65	3-53	16.21	.97
Correctness	Day	June (Post)	113	11.83	0-45	8.84	.92
	Night	June (Post)	43	10.49	0-29	8.65	.92
Structure	Day	June (Post)	113	14.53	0-41	10.16	.93
	Night	June (Post)	43	12.09	0-33	10.01	.94
Prompts	Day	June (Post)	113	9.31	0-26	4.86	.72
	Night	June (Post)	43	7.70	0-19	4.47	.71
John	Day	November (Pre)	115	37.27	0-70	20.54	.87
	Night	March (Pre)	46	30.54	0-65	21.07	.86
Morano	Day	November (Pre)	118	26.82	3-50	11.47	.93
	Night	March (Pre)	46	26.30	7-46	10.46	.92

Table 2.5 presents two kinds of information; in the upper half of the table are the intercorrelations among the tests; in the lower half of the table are the correlations of each test with the background characteristics of the students. (Note that the letter F means fall administration of the test, and the letter S means spring administration of the test.)

As can be seen by examining Table 2.5 the tests are all highly intercorrelated with the exception of Decoding 2 which is a pure measure of decoding skill. Decoding 1 measures word attack skills and understanding of word meanings, and is an estimate of general language proficiency. It is not surprising that Decoding 1 correlates with other measures of proficiency.

As can also be seen in that table the three scores of the Oral Proficiency Test are highly correlated so that it is reasonable to assume that the Oral Proficiency Test is measuring a general factor. It is also interesting to note that the Proficiency Test comprehension subscore has a high correlation with the John Test ($r = .85$). As we suggested above, both of these tests are measures of speaking proficiency.

The lower half of the table presents the correlations among the background factors and the test scores for the day-school sample. Very few of these correlations are substantial in magnitude, and only those equal to or larger than .30 are statistically significant. There are a number of interesting relations in this correlation matrix, however. There is no relation between sex of student and test scores, but there is a negative correlation of age with all of the test scores. Considering only those which are statistically significant, younger students do better on the Decoding, Literacy and Oral Proficiency Tests. Neither time spent in the U. S. or having a diploma was related to the fall Literacy Test score. Note, however, the substantial correlations between years of education and all of the test scores.

TABLE 2.5

Intercorrelations of Fall and Spring Tests and Student Background Information
Day School -- Matched Sample (N = 81)

	D_1-F	D_2-F	L-F	J-F	M-F	D_1-S	D_2-S	L-S	P-CT	P-CP	P-ST
Decoding$_1$ - F	-										
Decoding$_2$ - F	.50	-									
Literacy - F	.65	.41	-								
John - F	.56	.29	.77	-							
Morano - F	.55	.25	.59	.78	-						
Decoding$_1$ - S	.73	.40	.48	.42	.45	-					
Decoding$_2$ - S	.33	.27	.22	.10	.15	.41	-				
Literacy - S	.47	.31	.71	.64	.51	.47	.16	-			
Proficiency											
Correct	.60	.26	.59	.68	.63	.51	.22	.47	-		
Comprehension	.70	.39	.75	.85	.73	.63	.21	.61	.86	-	
Structure	.63	.30	.63	.74	.68	.54	.23	.51	.97	.90	-
Sex	.07	.10	.17	.17	.02	.03	-.06	-.02	.12	.19	.13
Age	-.32	-.15	-.30	-.13	.05	-.27	-.06	-.28	-.18	-.23	-.21
Time in U. S.	-.15	.05	.18	.13	-.07	-.13	-.08	.00	.05	.06	.04
Diploma	.21	[illegible]	.22	.11	.09	.22	.17	.09	.12	.18	.16
Former Job	.25	.17	.39	.33	.33	.29	.18	.22	.24	.32	.27
Job in U. S.	.13	.08	.21	.20	.19	.26	.16	.24	.13	.20	.17
C-Origin	.05	-.12	.02	-.07	-.15	.00	.09	.04	-.01	-.04	.01
Years Education	.21	.12	.33	.18	.16	.23	.15	.23	.22	.23	.25
Former Eng.	.16	.19	.25	.26	.17	.08	.13	.12	.12	.14	.16
Eng. in U. S.	.28	.13	.41	.51	.43	.25	.03	.31	.17	.37	.23

$p < .05$ for $r > .22$
$p < .01$ for $r > .28$

DEVELOPMENT OF THE OBSERVATION SYSTEM

Teachers' methods, materials and interaction with students were coded with an in-class observation system. This system was based on what the ESL teachers and students at the Center actually do. To develop the system, the ETS staff observed in several classrooms and then discussed their observations with the coordinating teacher. The staff spent much of the first three weeks of the project observing in Center classrooms and talking with the teachers. The last week in October each teacher was videotaped teaching a thirty-minute class.

On the basis of the in-class observations, the videotapes and discussions with the project director and coordinating teacher, we developed a preliminary form of the observation system. This first system was tried out in the Center classrooms. Repeated modifications were made and a lexicon of definitions prepared. During this time actual observations were discussed to clarify definitions of categories; categories were expanded or eliminated; videotaped and in-class examples were considered in developing the final categories. An important consideration in developing the system was to make sure that the categories would describe the differences between the two major teaching strategies used in the Center. The final draft version was used in preliminary observations of night-school classes as well as day-school classes to insure its appropriateness to both settings.

Description of the Observation System

The result of the development work described above was a categorical observation system which allowed for sequential coding of classroom behavior.

A reduced copy of the Observation Coding Sheet is shown in Figure 3.1. A complete lexicon and a sample observation are in Appendix C.

Three superordinate categories--context (instructional design), materials, and strategy (method of instruction)--describe the classroom setting within which the teacher and student behaviors are recorded. Each of these categories is subdivided; e.g., the context can be drill, writing, explanation, dictation, etc., and each subdivision has a numerical code. These categories are coded initially and recorded again only if they change during the observation period.

The first group of teacher behavior categories--questions, serial redirects, direction, models, writes on board, explanation, and other--are discrete instructional behaviors which usually initiate a teacher-student(s) interaction. Where these behaviors are carried out nonverbally, they are coded with an "N" rather than a check mark. The next column--class, group, individual--signifies to whom this behavior is directed.

The first group of student behavior categories--answers, free response, practice, writes on board, reads, chooses not to respond, asks question, (participates in) conversation, student-to-student feedback, and other--are those behaviors which either follow the teacher's initial behavior or initiate an interaction on the part of the student. Conversation and student feedback are coded with an "S" if they occur in Spanish rather than English.

The next three categories--positive, corrective, negative--describe the possible types of teacher feedback. Here again, since nonverbal feedback is an important component of "silent way" instruction, an "N" is used to indicate a nonverbal response.

FIGURE 3.1

Observation Code Sheet with an Illustration of a Possible Classroom Behavior Sequence

DATE — TEACHER — TIME — NO. OF STUDENTS 8 OBSERVER —

CLASS SET UP [T] STRUCTURE simple future

				TEACHER BEHAVIOR							STUDENT BEHAVIOR														TEACHER BEHAVIOR											STUDENT BEHAVIOR											COMMENT
	1 CONTEXT	2 MATERIAL	3 STRATEGY	4 QUESTION	5 SERIAL REDIRECT	6 DIRECTION	7 MODELS	8 WRITES ON BOARD	9 EXPLANATION	10	11 CLASS/GROUP/IND.	12 ANSWERS	13 FREE RESPONSE	14 PRACTICE	15 WRITES ON BOARD	16 READS	17 CHOOSES NOT TO RESP.	18 ASKS QUESTION	19 CONVERSATION	20 S/S FEEDBACK	21	22 POSITIVE	23 CORRECTIVE	24 NEGATIVE	25	26 MODELS	27 PROMPTS	28 ASKS TO REPEAT	29 REPEATS	30 EXPLANATION	31 WRITES ON BOARD	32 DIRECTION	33 QUESTION	34 ALTERNATE RESP.	35 STUDENT-MODELS	36 STUDENT PROMPTS	37 ANSWERS	38 FREE RESPONSE	39 PRACTICE	40 WRITES ON BOARD	41 READS	42 CHOOSES NOT TO RESP.	43 ASKS QUESTIONS	44 CONVERSATION	45 S/S FEEDBACK	46	47	48 COMMENT
1	1	1	1	✓							✓	✓										✓				✓		N									✓											
2																											N										✓											
3																										✓					✓								✓									
4							✓				I			✓																																		
5																																																

- 228 -

The second group of teacher behavior categories--models, prompts, asks to repeat, repeats, explanation, writes on board, direction, question, and other--designate response behaviors on the part of the teacher. Again, they are coded to indicate whether they are given verbally or nonverbally.

The second group of student behavior categories--student models, student prompts (these two usually follow a direction from the teacher), answers, free response, practice, writes on board, reads, chooses not to respond, asks question, (participates in) conversation, student-to-student feedback and other--describes those student behaviors given in response to the teacher's response to the student's initial behavior or response. Again, an "S" is used to code those responses which were given in Spanish.

The comment column allows the observer to indicate what the "other" behavior coded on that line is or to note some unusual classroom occurrence.

Use of the Observation System

The observation system was designed for in-class use. When the observer first entered the room, he or she took a few minutes to fill out the top of the observation sheet, indicating the date, teacher, time, number of students, observer, class set-up (diagram) and language structure being taught. During this orientation time it was possible to code the three superordinate categories and then quickly proceed to accurate coding of the classroom interaction behavior (See Chapter Three for an example of the coding.)

Observer Training

Using the videotapes of the Center teachers, the coordinating teacher and both tester-observers were trained on the system. In-class training was accomplished by placing two observers in the same classroom and comparing their interaction codes for agreements and disagreements: Those categories where the highest disagreement occurred were in several cases further modified and in others better clarified to improve inter-observer agreement. Once these final modifications were completed, reliability studies were begun.

Observation Schedule

Day School

A schedule for observations was prepared which allowed for a 20-minute observation of each class on four different days of the week. For those classes meeting only two or three times a week, observations were scheduled for each meeting. (See the latter portion of this chapter on the reliability of the observations.) Observation of the day-school classes began January 19th and continued through April 1, 1976.

Several factors affected the actual number of observations made; e.g., school holidays and closings as well as teacher and observer illness. The number of observations per class and observer are presented in Table 3.1. Note that for all but three of the classes (two of which met only two or three times-a-week), we had an average of 30 observation sessions per class, giving us a data base of some 600 minutes of observation per class.

Night School

A similar observation schedule was drawn up for the night school. Here, however, since class meetings were held for two hours, three times-a-week, one observer observed during the first half of a class and a second observer observed during the second half. Night-school observations were made by the

TABLE 3.1

Number of Classroom Observations: By Class and Observer

Day Classes	Observer 1	Observer 2	Observer 3	Total
0221	8	0	21	29
0233	9	0	6	15
0312	1	1	16	18
0313	15	0	1	16
0323	15	0	14	29
0412	15	0	15	30
0422	17	2	13	32
0433	16	0	15	31
0511	16	0	15	31
0522	15	1	13	29
0621	15	1	17	33
0711	16	1	14	31
0721	14	0	14	28
0731	16	0	14	30
TOTAL	188	6	188	382

Night Classes	Observer 1	Observer 2	Observer 3	Total
0901	5	3	4	12
1203	7	2	2	11
1402	5	4	5	14
1601	5	1	4	10
1701	7	3	4	14
2102	4	1	4	9
TOTAL	33	14	23	70

same tester-observers and the coordinating teacher. Observations were made at two different times during the semester: between February 5th and March 17th, 1976, and again between May 3rd and May 25th, 1976.

Again, the number of observations per class and observer are listed in Table 3.1. For the six night-school classes we had an average of 12 observations per class, giving us a data base of approximately 200 minutes per class.

Processing of Observation Data

As classroom observation data were being collected, the data were being transferred and coded onto forms which were set to ETS. These data were processed and data sets were made for use as input to the computer. An editing procedure similar to the one used for student data was used on these data.

Data for analyses were then created and this process is described in the section entitled, "Analysis of Behavioral Observations." Many analyses including summary statistics, correlations and factor analyses were completed. Factor scores were then created for each teacher and were placed on each student's record (the appropriate student and teacher were matched) for the final analyses. Appendix D contains the details of the data processing.

Descriptive Statistics Pertaining to the Observational Data

Table 3.2 gives the observation item numbers, category labels, codes and means for day and night-school classes. The category means may be interpreted as the proportion of observation episodes[3] in which the event was observed.

[3] An episode is defined as a sequence of behaviors between the teacher and any particular student. It may be initiated by either teacher or student and ends when the teacher addresses or responds to another student.

TABLE 3.2

Item Categories for Classroom Observation

Item	Number of Categories	Category Labels	Codes	Day School Means+	Night School Means+
1 Context	2	1 Drill	DRILL	.99	1.00
		2 Other	OTHR	.01	.00
2 Materials	7	1 No Materials	NO MATLS	.53	.65
		2 Rods	RODS	.11	.12
		3 Cartoons	CARTOONS	.02	.00
		4 Pictures	PICTURES	.05	.05
		5 Mimeo Sheets	MIMEO	.23	.17
		6 Objects	OBJECTS	.04	.00
		7 Sight Words	SIGHTWDS	.02	.00
3 Strategy/Model	5	1 Question & Answers	Q&A	.94	.99
		2 Free Response	FREE RES	.03	.01
		3 Repetition	REPEAT	.00	.00
		4 Directed Dialogue	DIRECTED	.02	.00
		5 Discussion	DISCUSSN	.01	.00
4 Initial Teacher Behavior	2*	2 Question	TB1-QUES	.32	.40
5 Initial Teacher Behavior	2**	2 Models	TB1-MODL	.10	.11
6 Initial Teacher Behavior	2**	2 Writes on Board	TB1-WOB	.05	.06
7 Initial Teacher Behavior	4	1 Serial Redirect	TB1-SR	.01	.02
		2 Direction	TB1-DIR	.18	.18
		3 Explanation	TB1-EXP	.04	.09
		4 Other	TB1-OTHR	.03	.01
8 Object of Teacher Behavior	3	1 Class	CLASS	.29	.31
		2 Group	GROUP	.01	.00
		3 Individual	INDIVDL	.70	.69
9 Initial Student Behavior	2*	2 Answers	SB1-ANSR	.35	.45
10 Initial Student Behavior	2*	2 Practices	SB1-PRAC	.11	.11
11 Initial Student Behavior	8	1 Free Response	SB1-FREE	.09	.05
		2 Writes on Board	SB1-WOB	.01	.02
		3 Reads	SB1-READ	.09	.12
		4 Chooses Not to Respond	SB1-CNOT	.00	.01
		5 Asks Question	SB1-AQ	.12	.06
		6 Conversation	SB1-CONV	.01	.03
		7 Student-Student Feedback	SB1-SFBK	.02	.04
		8 Other	SB1-OTHR	.02	.03

TABLE 3.2 (Continued)

Item	Number of Categories	Category Labels	Codes	Day School Means+	Night School Means+
12 Corrective Feedback	2*	2 Corrective	FB-CORR	.50	.30
13 Quality of Feeback	4	1 OK	FB-OK	.07	.02
		2 Positive	FB-POS	.05	.06
		3 Negative	FB-NEG	.01	.01
		4 Other	FB-OTHR	.00	.00
14 Successive Teacher Behavior	2*	2 Models	TB2-MODL	.22	.22
15 Successive Teacher Behavior	2*	2 Prompts	TB2-PROM	.47	.20
16 Successive Teacher Behavior	7	1 Asks Student to Repeat	TB2-ATR	.16	.07
		2 Repeats	TB2-REP	.04	.07
		3 Explanation	TB2-EXP	.07	.07
		4 Writes on Board	TB2-WOB	.06	.09
		5 Direction	TB2-DIR	.02	.02
		6 Alternate Response	TB2-ALT	.02	.00
		7 Other	TB2-OTHR	.01	.00
17 Successive Teacher Behavior	2*	2 Question	TB2-QUES	.09	.12
18 Other Student Behavior	2*	2 Student Models	SB2..SMODL	.09	.05
19 Other Student Behavior	2*	2 Student Prompts	SB2-SPROM	.05	.04
20 Successive Student Behavior	2*	2 Answers	SB2-ANSR	.75	.40
21 Successive Student Behavior	2**	2 Writes on Board	SB2-WOB	.02	.02
22 Successive Student Behavior	8	1 Free Response	SB2-FREE	.02	.01
		2 Practices	SB2-PRAC	.22	.17
		3 Reads	SB2-READ	.02	.05
		4 Chooses not to Respond	SB2-CNOT	.00	.00
		5 Asks Question	SB2-AQ	.04	.01
		6 Conversation	SB2-CONV	.00	.00
		7 Student-Student Feedback	SB2-SFBK	.04	.03
		8 Other	SB2-OTHR	.01	.00
23 Observer Comments	7	1 General Comment	C1-GENLC	.02	.02
		2 Student Response in Spanish	C2-SRSPN	.03	.04
		3 Teacher Action Non-verbal	C3-TA-NV	.10	.01
		4 Backwards Buildup Exercise	C4-BLDEX	.01	.00
		5 Teacher Reads	C5-READG	.00	.00
		6 Teacher Response in Spanish	C6-TRSPN	.00	.00
		7 Comment on Back of Summary Sheet	C7-OTHR	.07	.04

+Mean number of occurrences per episode. In most instances these may be interpreted as proportion of episodes in which the event was observed. They will only add to 1.00 when groups of behaviors are mutually exclusive and exhaustive.

*Binary items are designated appropriately as having two categories, but only the "behavior present" category is labeled and scored for analysis.

**Some of the binary items which could be repeated in rapid succession were simply counted instead of making a separate data in the day-school or night-school classes. However, these will add to 1.00 only when the behaviors within a group are mutually exclusive and exhaustive.

Note that the means for day- and night-school classes in most instances are highly similar. The categories where this is not the case are: night-school teachers are more likely to work without materials (item 2); day-school teachers are more likely to use corrective feedback (item 12), prompts (item 15), and their actions are more likely to be nonverbal (item 23); night-school students' initial behavior is more likely to be answering or questioning (item 9), while these behaviors are more likely to be successive behavior for day-school students (item 20).

Overall, the similarity between means for the remaining categories indicates that similar amounts and the same kinds of behavior were observable in both the day- and night-school classes. A description of how the observation data were analyzed follows.

ANALYSIS OF THE OBSERVATIONS OF TEACHING BEHAVIOR

The system of recording classroom behaviors employed in this study yielded a 23-item response record for each episode of interaction observed in the classroom. Each such episode could be initiated either by the teacher or by a student, and could continue through an extended sequence of dyadic interchanges. The participation of other students in the basic pattern of teacher-student interactions was also recorded. Every interchange in each sequence of interaction (episode) was initially recorded and coded individually; these individual instances were then aggregated by summing to obtain one data-record per episode. An episode began at the initiation of any new interaction either by the teacher or by a student.

Methods of Organizing Dyadic Chains of Interaction

As seen in Table 3.2, some of the things a teacher could do to initiate an interaction with a particular student, group of students, or the class were: ask a question, model correct usage, or give directions (use of the Observation Lexicon in Appendix C in conjunction with Table 3.2 will be helpful throughout this section). A student could initiate interaction by asking a question, for example, or could continue an episode of interaction initiated by the teacher by answering a question, following instructions (e.g., to read or to write on the board), practicing, and so forth. If the sequence of interaction continued, it could lead to further teacher behavior in the form of corrective feedback, modeling, prompting, questioning; which could lead, in turn, to the student's second attempt to answer correctly, to more practice, and so on. Long chains of cyclical, dyadic interaction could thus be, and indeed were, coded as "successive" behaviors in each episode of classroom interaction.

For purposes of data analysis we chose to treat each distinct interaction sequence or classroom episode as the basic unit. This means that all teacher and student behaviors after the initiation of any particular interaction sequence were aggregated. We thus retained only the distinction between <u>initiating</u> and <u>subsequent</u> behaviors. By maintaining the distinction between initiating behaviors and subsequent consequences, however, we left some opportunity for the most basic aspects of sequential organization among classroom behaviors to emerge in our analyses.

Objectives of the Analysis of Classroom Interaction

Our main interest in the analysis of the behavioral observation data was to find, if they occur, fairly stable and distinct patterns of classroom interaction. That is, we wanted to find out if certain teacher behaviors give rise to or are associated with certain specific student behaviors. From the point of view of analyzing categorical data we were interested in the patterns of association shown between all pairs of multicategory items (subsections) given in Table 3.2.

Methods of Analyzing Classroom Interaction

The availability of data on many individual behavioral episodes made it possible to conduct factor analyses of a large number of binary variables separately for the day-school and night-school samples, even though only six teachers were studied in each sample. Several different methods were used to determine how many common factors should be extracted from each correlation matrix.

Factor analysis is a method of studying the association among variables. The method uses the intercorrelations among all the original variables. In this study the original data are the individual codes; that is, the entries in each category in Figure 3.1. Successive behaviors were summed within each episode of interaction; different categories within episodes are correlated to varying degrees.

Suppose that modeling and practice were associated, that is, teachers who modeled also elicited practice in the same episode, and teachers who used modeling infrequently also elicited practice infrequently in the same episode. This association would be represented by a correlation coefficient such as r = .80. The correlation table (matrix) is made up of numbers of this kind. The numbers will vary in size, and there will be a number for each pair of categories. Smaller numbers (closer to .00) mean the association is weak; the larger the numbers (closer to + 1.00 to - 1.00), the stronger the association.

Factor analysis tries to find patterns of association in the correlation matrix. A pattern found, for example, after the teacher asks a question and the student attempts an answer, was: corrective feedback, prompt, student answers again. This pattern was found because the three pairs of categories were highly associated, and this association would be represented in the correlation matrix by substantial r's; for example:

corrective feedback-prompts: r = .57
prompts-student answers: r = .70
corrective feedback-student answers: r = .62

Corrective feedback also correlates with teacher modeling (r = .50), and with student practice (r = .47). Modeling also correlates with student practice (r = .83). Another factor was found: corrective feedback, model, practice.

Thus, six of the original variables are intercorrelated; but the factor analysis breaks these intercorrelations into two patterns or factors. (Note that corrective feedback appears in both patterns.)

The method is a quantitative way of looking for these patterns which are called factors. The methods may yield none or many factors. The number of factors produced depends on the number of distinct patterns of association in the data.

Factors of Classroom Interaction

Rather than discussing factors found in our analyses in a rigid but superficial way with respect to the relative amount of variance accounted for, size of loadings, etc., we prefer to try to convey to the reader the outcome of our own attempt to infer how classroom behavior seems to be organized. After all, the point of undertaking these analyses is not to generate numbers that we are then compelled to take seriously, but to help us form a conceptual framework based upon empirical results through which we can come to understand and discuss classroom interaction.

Comparison of Day School and Night School in Patterns of Initiating Classroom Interaction

Although there are many similarities between the patterns of classroom interaction identified by factoring day-school and night-school data, it is equally informative to note some of the major differences. It seems that a major difference between these samples (or analyses) is related to the sequencing of classroom behavior.

Teacher model-student practice factors. We found "teacher model"-"student practice" factors in both samples. In the day-school analysis separate "model-practice" factors break out for the initiation (V) and follow-up (II) phases of interaction.[4] But in the night-school analysis one factor (I) involves both phases of interaction, while the other (V) is for follow-up. The distinction between phases of the "model-practice" interaction pattern in the day-school analysis is further accentuated by

[4]The Roman numerals in parentheses refer to the factor columns in the respective tables. Note that the factors are printed in arbitrary order but are sequence-numbered in terms of their relative sizes.

some indication that initiation of the sequence leads to a request to repeat the response rather than to automatic continuation of the "model-practice" sequence.

Teacher direct-student read factor. Like "model-practice", another major pattern of classroom interaction which seems to display differential sequential organization in the two samples is "teacher direct" - "student read." In the night school, again we find that initiation of this sequence (IX) leads to its continued expression or repetition--but the teacher may "ask the student to repeat" in continuation of the reading sequence. This result suggests rather sustained, continued reading sequences under the teacher's direction, with repetition as necessary. In the day school, on the other hand, factor (IX) suggests a pattern of initial "teacher direct = student read" behavior, but if anything, a low tendency for further teacher direction or explanation; that is, we see a short reading episode which is not continued.

Teacher direct-student read or ask question factor. In addition to what was said immediately above about the "direct-read" sequence in each sample, there seems to be another sequence (VII day, VII night) of "teacher direct" - "student read and/or ask question." This behavior pattern seems to be in contrast to one of "teacher ask question" - "student answer." There is the further complication of a class vs. individual distinction in the night school. There "teacher ask question" applies to the class-as-a-whole, whereas individual students are "directed to read." One way of interpreting what is going on here is to regard this factor as a contrast between two situations which can alternate within the same classroom during any given session of observation.

In the night-school sample there is some indication that students may be reading mimeo materials but are being asked to answer questions about pictures. In the day school there is some indication that the teacher's behavior after a student's attempt to read or after asking a question is an explanation; this, in turn, may be followed by more reading and question-asking on the part of the student--a tutorial arrangement.

Teacher question-student answer factor. In each sample there is also another, distinct, "teacher question"--"student answer" factor (III day, IV night) that is not contrasted with student question-asking, just as we already saw distinct "direct-read" factors in each sample. In both samples it is clear that this somewhat more pure "question-answer" paradigm is directed at individual students rather than to the class at large. In both samples there is some indication that the instructor might subsequently repeat the question.

Let us now briefly name and codify the four factors introduced above before moving on to some of the major dimensions of classroom interaction which seem to follow these initiating events: (1) "teacher model"-"student-practice"; (2) "teacher direct"-"student read"; (3) "teacher direct"-"individual student read or ask question" vs. "teacher asks question"-"student in class answers"; (4) "teacher questions individual"-"individual student answers".

Patterns of Classroom Interaction Which are Sustained Once Initiated

The "model-practice" pattern seems to lead to subsequent cycling in the night school. There are also several other factors of classroom interaction which seem to continue once started and therefore cut across the "initiating" vs. "subsequent" dichotomy which we have set up.

Other factor. One factor which emerged in both analyses (IV day, II night) is simply characterized by the "other" category of both initial and subsequent behavior on the part of both student and teacher. The observer made a special notation about the content of the behavior on the observation form and the behavior tended to be of long duration. These notations were so diverse that any further classification would have added nothing to the analysis. Examples are given in the Lexicon under "Comments - Other." (See Appendix C.)

Student-student feedback factor. Another factor which showed up in both samples is largely student initiated. "Student-student feedback" is the behavior involved here, both initiating and subsequently continued, and it is associated with the observer comment that the feedback occurred in Spanish (VI day, VI night). In the night school this feedback might be associated with "teacher explanation" as well as repeating verbatim what the student has said. It is possible that in this pattern of interaction both the teacher and other students are trying to help an individual get something straight.

Free response factor. A factor which emerged only in the day-school sample (VIII) is related to the "free response" strategy of instruction as opposed to question and answer. A "free response" arises when the teacher has indicated only the structure and not the idea or vocabulary to be used.

Comparison of Day School and Night School in Patterns of Follow-Up Classroom Interaction

As for the clearly "subsequent" patterns of interaction detected, we see two major patterns connected with "corrective feedback."

Corrective feedback-model-practice factor. The first factor mentioned in this discussion was "model-practice" and we stated there that it "breaks up" into an initiating and subsequent factor in the day school, but appears as a cycle of continued interaction in the night school.

Corrective feedback-prompt-answer factor. Another, quite distinct pattern of follow-up interaction involving corrective feedback was recovered in both samples (I day, III night). In this case there is some reason to believe that the feedback is encouraging (especially from the day-school data) and the sequence goes: "corrective feedback"-"teacher prompt or asked to repeat"-"student answer." Since prompting or cueing the student on how to modify his or her response to make it correct predominates here, we will call this factor "CF-prompt-answer."

However, there are other possible teacher reactions in addition to the ones mentioned above; from the night-school data we see that the teacher might further direct the student; in the day school we see the possibility of questioning the student. In either case, the observer comment indicates a strong tendency toward a nonverbal teacher action, especially in the day school. Here also we see the possibility that other students will get involved in modeling, prompting, or student-student feedback (day school). In the night school we see some indication that this particular pattern of interaction may occur while the student is writing at the board.

Teacher question-student answer factor. This description completes our discussion of the classroom interaction factor patterns except for one extra "subsequent" behavior found only in the night-school sample: "teacher question"-"student answer" (VIII). This factor is relatively highly correlated (.26) with the "CF-prompt-answer" factor just discussed, but "teacher question" is included on that factor for the day school.

Classroom Differences in Interaction Patterns

Once the factor analyses of behavioral observations were completed, we investigated the differences in classroom behavior patterns--the ultimate aim being, of course, to identify interesting contrasts among classrooms which might account for differences in student achievement. We were interested in between-classroom variation in behavior for its own sake, however, and therefore sought a technique which would permit us to optimally discriminate between classrooms on the basis of behavior observed therein.

We settled upon the individual classroom as the focus of study at this stage because our interest centered upon teacher differences as well as upon differences in the ways in which any given teacher might approach students of various proficiency levels. For purposes of assessing variation within classrooms, that is, the interaction of a particular instructor with a more-or-less homogeneous group of students, we chose individual sessions of observation as the basic units of analysis. We contrasted overall mean differences between classrooms on each pattern of interaction (factor) with the amount of day-to-day variation observed within classrooms for that pattern of interaction. This was done simultaneously for scores on all nine factors, separately for the day-school and night-school samples.

The technique of analysis of variance was used to estimate the statistical significance of the variation in interaction patterns between classrooms to the variation within classrooms. This technique yields a ratio of the between-classes variance to the within-class variance, which is called the F-ratio. The probability of obtaining an F-ratio of a certain magnitude is provided by standard statistical tables; in Table 3.4 these values have been provided for the significant F-ratios.

Table 3.4 presents this information for this study. The patterns of interaction of factors which we described in the preceding sections are listed in the left-hand column. The F-ratios are listed in the columns to the right. The way to read these numbers is as follows:

1. Note that two names are provided in each row. In the first row, two different names are listed, the upper one being the name of the factor found in the day-school sample, the lower being the name of the factor found in the night-school sample. Recall that similar and sometimes identical factors were found in both samples. In the next several rows, identical names are listed because these factors were found in both samples.

2. Read across to the columns labelled, "F-Ratio." Reading in the first row, we see that the F-ratio for "Model-Practice" for the day-school sample is 4.05. The asterisks indicate that the probability of obtaining a number this large by chance alone is less than one in a hundred ($p < .01$).

An F-ratio will be greater than 1 whenever the numerator is larger than the denominator. The numerator in this analysis is the between classes variance, an estimate of how much the teachers differ from each other. The denominator is the within-classes variance, an estimate of how much each teacher differs from day-to-day. Thus when the F-ratio is greater than 1, we know that the variance between-classes is greater than the within-classes variance; or the teachers differ more from each other than they do from themselves from day-to-day.

TABLE 3.4

Univariate Analysis of Variance Tests for Classroom Differences on Factor Scores

Factor Label	Day School F-Ratio	Night School F-Ratio
"Model-Practice"	4.05**	
"Model-Practice-Model-Practice"		.83
"Direct-Read"	4.32**	
"Direct-Read"		.63
"QA-Individual" vs. "Direct-Read/AQ"	2.33**	
"QA-Individual" vs. "Direct-Read/AQ"		1.24
"QA-Individual"	17.40**	
"QA-Individual"		4.81**
"Student Feedback"	7.70**	
"Student Feedback"		2.10
"Other"	1.73	
"Other"		1.20
"Free Response"	7.88**	
"CF-Model-Practice"	11.50**	
"CF-Model-Practice"		1.60
"CF-Prompt-Answer"	11.98**	
"CF-Prompt-Answer"		4.84**
"Question-Answer" (successive)		3.40*
Between Classrooms Degrees of Freedom	13	5
Within Classrooms Degrees of Freedom	367	44

*$p < .05$

**$p < .01$

QA: Teacher Question-Student Answer

AQ: Student Asks Question

CF: Corrective Feedback

But we must also consider the probability that the F-ratio's value is a chance phenomena. Therefore, before we infer that the contrast being studied actually describes what is happening in these classes, we must consider the statistical significance of the F-ratio which is represented by the p-value. In the case of the first factor, we may conclude that the differences found between teachers are likely to be "true" differences; that is, the differences we have found in our observations describe what is actually happening and is not an artifact or error in our observation system.

Note, however, that the F-ratio for the night-school factor is less than 1, F = .83, and, of course is not statistically significant. There is more variation from day-to-day for each teacher than there is between teachers.

Notice that in the day-school sample the F-ratio is significant for all but one factor. In the night-school sample, the F-ratio is significant only for three factors. The day-school teachers differ more from each other than they do from themselves day-to-day. The night-school teachers differ in this way only in three respects.

We will not undertake a detailed discussion of the univariate analysis of variance results presented in Table 3.4 until after we have completed discussion of the multivariate analysis (canonical discriminant function analysis). The latter will provide an overall test of classroom differences which takes into account any dependencies among various classroom interaction factors. The results in Table 3.4 are presented here mainly to justify our emphasis upon analyzing the day-school data in what follows, and to give the reader a familiar frame of reference prior to undertaking discussion of the multivariate analysis.

From the summary results in Table 3.4 it can be seen that many classroom differences were detected in the day-school sample but, presumably due in part to the smaller number of observation sessions per classroom, few differences were detected in the night-school sample. There were from 15 to 33 observation sessions per classroom in the day school compared to from 9 to 14 such sessions per classroom in the night school.

Session-to-Session Covariation
of Classroom Interaction Patterns

Because of the greater precision of day-school classroom comparisons, we will focus there in the following discussion with only a brief discussion later of the night-school results. In considering the univariate analysis of variance results given in Table 3.4, it must be remembered that the various classroom interaction factors are not uncorrelated; therefore, there is apt to be some overlap or redundancy in the F-test results reported. In order to get a feeling for this association between classroom interaction factors as they vary from day to day (observation session to observation session) we can examine the pooled within-classrooms correlation matrix among the factor scores which is presented in Table 3.5.

TABLE 3.5

Pooled Within Day School Classrooms Correlations Among Factor Score Session Means

	QA IND	CF-P-A	CF-M-P	MOD PRAC	OTHER	DIR READ	QA vs. DIR	SFBK	FREE
QA IND	1.00								
CF-P-A	.45	1.00							
CF-M-P	.05	.16	1.00						
MOD PRAC	- .15	.00	.30	1.00					
OTHER	- .18	.02	.27	.56	1.00				
DIR READ	- .24	- .18	- .08	.21	- .02	1.00			
QA vs. DIR	- .01	.10	.03	- .07	- .15	.06	1.00		
SFBK	.02	- .01	- .02	- .03	- .09	- .26	- .16	1.00	
FREE	- .05	- .03	- .14	- .21	- .25	- .25	.00	.13	1.00

Mnemonic label	Interaction pattern
QA IND	"teacher question - individual" - "student answer"
CF-P-A	"corrective feedback" - "teacher prompt" - "student answer"
CF-M-P	"corrective feedback" - "teacher model" - "student practice"
MOD PRAC	"teacher model" - "student practice"
OTHER	"other"
DIR READ	"teacher direct" - "student read"
QA vs. DIR	"teacher ask question" - "student answer" vs. "teacher direct" - "student read and/or ask question"
SFBK	"student - student feedback"
FREE	"free response"

For purposes of discussion we have ordered the rows and columns of the pooled-within-classroom correlation matrix in Table 3.5 so as to best reveal the clustering of classroom interaction patterns. Note that most of the positive intercorrelations are near the main diagonal and that several clusters of associated interaction patterns have been blocked off there.

Note in Table 3.5 a tendency for the "corrective feedback-prompt-answer" interaction pattern to occur on those days when the "question-answer" pattern is used with individual students (r = .45). These two factors thus go together to form a "macro" pattern which links individual episodes of classroom interaction.

A second cluster which emerges has as its focus the "teacher model"-"student practice" interaction pattern. The inclusion of "corrective feedback"-"teacher model"-"student practice" as well as "other" in this cluster indicates day-to-day covariation in these repetitive, rehearsal-based sequences of interaction.

It is interesting to note that the bipolar factor which contrasts teacher questioning with student question-asking is relatively independent of the other interaction patterns as they vary from day to day within classrooms. Finally, we see that the more flexible and open patterns of classroom interaction ("student-student feedback," "free response") show a slight tendency to prevail on the same days (.13) but are inconsistent with the "direct-read" (-.25)/"model-practice" (-.21)/"other" (-.25) syndrome.

Multivariate Analysis of Classroom Differences in Interaction Patterns

Moving now to a consideration of overall classroom differences averaged across days of observation we must consider the outcome of canonical discriminant function analysis. This analysis is designed to identify those linear combinations of observed scores (in this case, classroom interaction factor scores) which maximally discriminate between groups relative to within-group variation. As far as this study is concerned, the role of canonical discriminant function analysis is to identify ways in which classroom interaction patterns can be combined to yield stable overall contrasts among the classrooms studied; that is, to find how the classrooms differ most clearly from one another on the average while displaying minimal day-to-day fluctuations.

The essence of the canonical discriminant function technique is to find the bases on which the classes are most sharply discriminated from each other on the average. The results of this type of analysis are portrayed in terms of discriminant axes. There may be several such axes. The axes can then be described numerically in terms of the original factors.

In this study, the canonical discriminant function analysis generated eight discriminant functions, but the first two accounted for most of the variance in the factor score means. Each factor could be "located" with respect to these axes. These numbers are provided in Table 3.6 in Chapter Three.

An inspection of the results of the canonical discriminant function analysis among day-school classrooms summarized in Table 3.6 reveals that there are two major dimensions of variation between classrooms and two or three more minor dimensions.

One way to describe the results of these analyses is to portray them visually. Each class will be represented in a space defined by the discriminant axes. Each class will be represented by a point. By projecting this point out onto the axes we can obtain an idea of the characteristic patterns of interaction in that class.

Because the first two canonical discriminant functions account for the majority (78 percent) of all differences between day-school classrooms, in terms of stable patterns of teacher-student interaction, we have chosen to plot the classes with respect to these axes in Figure 3.2. In the figure each classroom is identified by a letter-number combination. The initial letters range from A through F and identify the six day-school instructors. The numbers range from 1 to 3 and refer to proficiency level of the class being taught as measured by the John Test. The lower case letters identify different classes at any given level where they are taught by the same teacher.

We have drawn vectors in Figure 3.2 to represent the way the nine original factors of momentary classroom interaction project into the discriminant space. Note that the important features of these vectors are their directions and their relative lengths--their absolute lengths are arbitrary and have been scaled merely for convenience of plotting. By visualizing the direct perpendicular projection of classroom centroids onto these vectors (extended through the origin if necessary) one can get a feeling for the relative ordering among classrooms in terms of the patterns of interaction. By noticing the relative collinearity or perpendicularity of these vectors one can get a feeling for how classroom behaviors associate or disassociate in the profiles which distinguish classrooms from one another.

As for the classroom centroids plotted in Figure 3.2, the most outstanding feature is the obvious clustering of classrooms taught by the same teacher. This phenomenon is particularly striking in view of the fact that no information about who taught what class entered into any of the analyses leading to these results. The evidence is thus incontrovertible that teachers have consistent

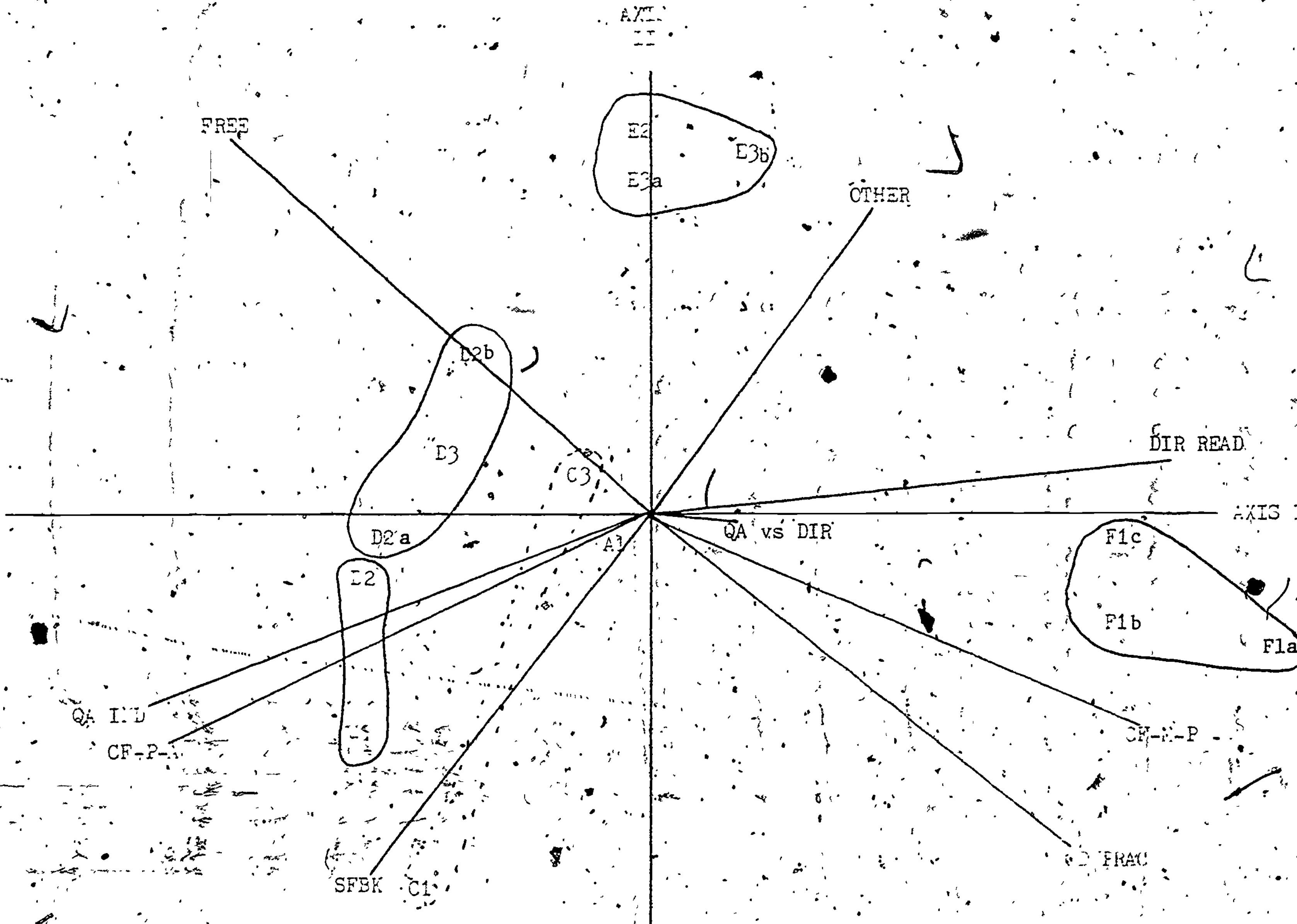

Figure 3.2

Plot of Day-School Classroom Centroids and Classroom Interaction Factors in the Space Defined by the Two Largest Canonical Discriminant Axes and distinct "styles" of interaction with students--styles which do not in general vary markedly even when teaching classes of quite different initial ability level.

Only one teacher, C, appears to employ widely different strategies when teaching students of differing ability levels; however, it must be pointed out that no other teacher had the opportunity to teach classes containing students of such widely different ability levels.

Because composite variables in canonical form are notoriously difficult and dangerous to interpret we will not attempt to label the axes in Figure 3.2. We can note, however, that the first (horizontal) axis is a contrast between about the same clusters of variables that we identified in the pooled within-classrooms correlation matrix in Table 3.5. That is, a general pattern of DIR READ, CF-M-P, MOD PRAC, and OTHER is contrasted with QA IND and CF-P-A.

Yet another method was used to transform (rotate) the canonical variates ~~into a more~~ theoretically informative and interpretable position. For this purpose we chose to use only the first four discriminant axes since they account for most (93 percent) of the variation between groups and little significant variation remains (χ^2 - 59.4; 45 df) after the fourth axis is considered. (Table 3.7 in Chapter Three presents the DIRECT GEOMIN transformed between-classrooms canonical-variate factor pattern matrix.) Upon transformation of the first four between-classrooms canonical variates we came up with two major contrasts between teacher-student interaction patterns (still accounting for the majority of the between-groups variance) and two smaller variates each relating essentially to only one or two patterns of momentary interaction. Because the transformed canonical variates are fairly highly correlated it is clear that there is a second order general factor in operation here.

It must be kept in mind that the four contrasts presented were arrived at because they discriminate quite sharply between behavior patterns observed in different day-school classrooms in this particular study. We have already seen that the classroom differences detected are largely a function of teacher differences, however, so it is quite likely that other contrasts in teaching behavior would become salient given a different sample of teachers.

<u>First axis</u>. The first axis is a specific distinction between the "question-answer-corrective feedback-prompt-answer" paradigm of classroom interaction and the "direct-read" and/or "other" paradigm. At the level of contrasting individual aspects of classroom interaction, however, interchangeability characterizes the <u>follow-up</u> aspect of the positive pole of this contrast. A "question-answer" pattern directed to individual students and followed by "corrective feedback" may lead to quite a variety of teacher behaviors in addition to "prompt". "Student-student feedback" likewise figures prominently as a follow-up pattern, as does "student question-asking". The "direct-read", "other" pole of this first contrast between classroom interaction patterns is also characterized by many subtleties at the level of individual episodes of interaction.

<u>Second axis</u>. The second transformed discriminant function axis picks up a contrast between the "free response" mode of interaction and the "model-practice-corrective feedback-model-practice" paradigm. As for all four transformed axes, the positive pole applies more to individualized instruction while the negative pole refers to class or group-oriented instruction.

<u>Third axis</u>. The third transformed canonical variate is not a strong bipolar contrast, unlike the others, but relates quite simply to the presence of "student-student feedback" as a component of classroom interaction. Along with this student feedback is some indication of corrective feedback on the part of the teacher and other follow-up behaviors similar to those seen for the first axis.

Fourth axis. Finally, the fourth transformed canonical variate is interesting in that it relates strongly and negatively to the original bipolar contrast found when we factor analyzed classroom interaction patterns as they varied from episode to episode: "teacher question"-"student answer" vs. "teacher direct"-student read and/or ask question". We now reiterate our earlier conjecture that the "direct-read/ask question" pole of this factor is, indeed, an interaction pattern characteristic of relatively free, unstructured classrooms, whereas the "question-answer" pole characterizes more highly structured classrooms.

What the transformed canonical discriminant function axes seem to reveal is that several patterns of manifest classroom interaction are combined in different ways to form "higher-order" patterns. Thus, the observed interaction pattern "teacher direct"-"student read" can either be part of a highly structured, "programmed" classroom interaction climate (the negative pole of the first axis) or part of a more open-ended, unstructured setting for interaction (the positive pole of the fourth axis). Looking at the opposing poles of these same axes, respectively, we can infer that "teacher asks question" can likewise be part of an encouraging, prompting, follow-up pattern or it can form part of a more demanding pattern of drill.

Summary of the Results of the Analysis of the Observation of Classroom Interaction Patterns

At this point it will be useful to summarize the major steps in the analysis. The first step produced nine patterns of interaction (factors). Since these nine factors were intercorrelated, a second analysis was conducted, a canonical discriminant function analysis. The nine factors could be located principally in terms of two discriminant analyses. Each class could be described in terms of its location with respect to these axes and with respect to the original nine factors (see Figure 3.2). These canonical variates were then transformed with respect to four axes, two of which account for most of the variance among the canonical variates.

The four axes produced in this last analysis are bipolar, which means the data for a set of observations will reveal a position closer to one end than the other of an axis. Thus by knowing a class' or teacher's discriminant function scores with respect to these axes, we visualize the classroom interaction patterns in that class.

The first discriminant axis contrasts two paradigms; one, the "question-answer-corrective feedback-prompt-answer" style; the other, the "direct-read or other paradigm. The first paradigm is a structured style, almost as if it were programmed. The other paradigm is less structured, requiring more responsiveness on the part of the students.

The paradigms are not "pure" in practice. The <u>follow-up</u> phase of the first paradigm has several variations. Other patterns occur as part of it such as "student questions". Similarly, the "direct-read or other" style has components of "teacher explanation", or "modeling" or "writing on the board" in it. The principal features of each pole of this axis are those listed in its label.

The second axis contrasts a "free response" mode with a "model-practice-corrective feedback-model-practice" paradigm, the most structured of the interaction patterns. Thus what this as well as the first axis seems to be describing is a contrast between a more and a less structured style.

Since classroom patterns are not pure types in practice, a class or a teacher will usually have scores that place them closer to a pole of each axis. Teacher A and Teacher B will be alike or different depending on where they are located with respect to these axes. Also any one teacher may include other patterns within the overall style.

The third axis is not a strong bipolar factor. The axis locates a class with respect to whether the "student-student feedback" pattern appears in the interactions.

The fourth axis is a contrast between "teacher-question and student-answer" and "teacher direct-student read or ask question." The "question-answer" pole is characteristic of the more structured classrooms.

It may appear to the reader that the classroom interaction patterns simply are either structured or unstructured. Such is not literally the case. There are three paradigms of structured instruction and two of unstructured instruction. But these paradigms obviously do not occur in pure form in this sample of teachers and students.

Perhaps the most illuminating way to think about what these axes mean is to think of them as dimensions along which a teacher will have a score. Thus each teacher will have a profile of scores. It is this profile that describes each teacher's characteristic style.

What is most important to notice about these styles is their stability. We found that differences between teachers were greater than the teachers' day-to-day variations. Further, with one exception, the teacher's style did not vary with respect to the proficiency level of the class.

We now turn to whether these styles had differential effects on the acquisition of proficiency in English by the students. First, however, we will describe the reliability of the observations. Had they not been reliable, we would most likely not have found significant relations between classroom interaction patterns if they existed.

Reliability of Classroom Interaction Pattern Contrasts

The succession of data reduction procedures applied to classroom behavioral observations has ultimately led to only 56 essential numbers, the classroom centroid coordinates on each of four transformed discriminant function axes. These coordinates define four contrasts between day-school classrooms which can now be used in an attempt to predict student achievement over the school year from classroom interaction experiences. How reliable are these indices of classroom interaction since unreliability limits predictive validity? The issue of reliability also provides us an opportunity to return, as promised, to the univariate analysis of variance results presented in Table 3.4, since we will discuss reliability in terms of variance components.

Components of Variance Analysis of Classroom Interaction Factors

The day-school sample comprises a completely nested four-level analysis of variance design having approximately 20 episodes nested within each session, approximately 25 sessions per classroom, and about two classrooms per teacher.

The results of the analysis of variance are presented in Table 3.8. The original nine factors derived from the first factor analysis are listed at the head of each column. Down the left-hand side are the sources of variance. The

TABLE 3.8

Components of Unit Total Sample Variance of Day-School Classroom Interaction Factors Due to Teacher, Classroom, Session, and Episode Differences

Source	QA IND	CF-P-A	OTHER	DIR READ	FREE	MOD PRAC	CF-M-P	SFBK	QA VS DIR	df
Teacher	16.0**	9.4**	1.8*	5.9*	11.6**	4.6**	4.6*	.5	.5	5
Classroom	0.0	.2	0.0	0.0	1.6*	0.0	1.5**	1.8**	1.0	8
Session	15.2**	12.7**	30.1**	24.4**	33.9**	11.0**	10.8**	7.5**	33.8**	368
Episode	68.8	77.6	68.1	69.7	52.[illegible]	84.4	83.1	90.2	64.7	6704
Discriminant Axis	+I	+I	-I	-I	+II	-II	-II	+III	-IV	

* $p < .05$

** $p < .01$

episode is the smallest unit of analysis; episodes occur within sessions. Recall that an episode is a distinct pattern of interaction. Sessions are the hours of observation on different days. Classrooms are the different groups taught by a teacher; and teacher, of course, refers to the individual teacher.

The numbers in the columns are percentages--percentages of the variance in the factor attributable to a source of variance. Consider the factor, "Question-Answer-Individual" (QA IND). This factor comprises those performances in which the teacher asks an individual student a question. Teachers vary on how much they use this performance. Some of this variation may be associated with episodes; that is, some episodes will include this performance while others will not. Whether or not the performance appears may be due to what the teacher is trying to do at that moment. Or the variation in the data may be due to the day or time of the period in which the class was observed. Or, the variation may be related to the class being taught. Or, it may be related to differences in teachers' styles.

By reading down the column under QA IND we can see what percentage of the variance is attributable to each source. Thus episodes account for 69 percent of the variance, 15 percent is accounted for by session, and 16 percent is attributable to teachers. But the 69 percent attributable to episodes is not statistically significant. Therefore, the variation in this performance is related to who is being observed (teacher) and when they are being observed (session). The teachers differ from each other in this respect; and an individual teacher will differ in this respect from day-to-day.

From Table 3.8 it can be seen that from 53 to 90 percent of the total variance of each classroom interaction factor can be attributed to within-session variation (see line labeled Episode in Table 3.8); that is, to variation from interaction episode to interaction episode during the same session (day) of observation. The factors are thus primarily factors of within-session variation.

In Table 3.8 we have indicated the probability levels associated with each component of the total sample variance, along with the associated degrees of freedom. From these results many aspects of variation in classroom interaction patterns can be clarified. It is clear, for instance, that only a very small and generally nonsignificant contribution to variation in interaction patterns can be attributed to differential treatment of separate classrooms by the same teacher (see line labeled Classroom in Table 3.8)--an outcome already suggested by the compact clustering of classroom centroids for most teachers in Figure 3.2 of the multivariate analysis.

The relatively large components of variance for Sessions suggest that a number of these factors fluctuate substantially from day-to-day.[5] Especially noteworthy are the large components of variance for interaction patterns involving "free response" (33.9 percent), "other" (30.1 percent), and teacher "direction" (24.4 percent to 33.8 percent) behaviors. These results suggest that any given teacher might be inclined to devote certain sessions to these "specialized" activities but refrain from doing so in other sessions.

But, considerable session-to-session variance characterizes many of the classroom interaction factors. This result indicates that each session of

[5] Some of the session-to-session fluctuation could be attributed to observer biases, since various observers collected data on different sessions. This suspicion will be dispelled shortly, however, when we consider the inter-observer agreement approach to reliability.

observation provides a rather narrow view of what is going on in the classroom in general and may explain why stable teacher differences could not be well established with the limited night-school observation schedule.

In summary, differences among teachers account for from .5 to 16 percent of the total variance in the observational data (see line labeled Teacher in Table 3.8). The differences among the teachers are statistically significant on all but two factors (SFBK and QA VS DIR). In ordinary language the teachers teach differently, and they differ most on the factors "question-answer--individual student" (QA IND) and "free response" (FREE). There is not much evidence that the particular classroom being taught by a teacher has any influence upon the patterns of interaction which take place therein, as we have seen earlier in connection with Figure 3.2. The major component of overall variation in classroom interaction patterns can be attributed to alternation in interaction patterns from episode-to-episode, which is not surprising, but session-to-session changes also play an important role. Notice that session differences are statistically significant on all nine factors.

Fortunately, predicting what might take place in each individual classroom episode or even in any given session of observation was not our task in this study, so we can be content to account for such variation by reference to hypothetical determinants or "factors" of momentary classroom interaction. Our aim was to detect stable classroom or teacher differences and it is to assess our attainment of that goal that a reassessment of the components of variance in Table 3.8 is required. That significant teacher differences were obtained despite the fact that much session-to-session and episode-to-episode variation occurred is not hard to understand because so many sessions and episodes were recorded for each teacher.

When the components of variance in Table 3.8 are weighted to take into account the fact that approximately 1000 episodes of interaction were observed per day-school teacher, about 500 per classroom, and about 20 per session, we see that a very large fraction of the observed variation among teacher means on each classroom interaction factor is reliable (i.e., can be attributed to true differences between teachers). Since substantially the same issue is addressed by the F-tests among teachers (given in Table 3.4), it is clear why so many significant results were found in the day-school sample.

Components of Variance Analysis of
Transformed Discriminant Function Scores

The multivariate analysis of variance (canonical discriminant function analysis; see Table 3.6) revealed only a few significant contrasts among day-school classrooms. From the first four of these canonical discriminant functions we derived the four transformed axes discussed earlier. Since these are the classroom interaction contrasts which we wish to employ to predict student achievement, it is in order to determine how sensitive teacher scores on these contrasts are to various possible underlying sources of variation: teacher differences, classroom differences, session differences, episode differences, and observer differences as well. The issue of inter-observer differences will be taken up shortly.

First, however, let us look at the results of a second component of variance analysis for the transformed discriminant axes, just as was done for the nine original factors of classroom interaction in Table 3.8. The upper section of Table 3.9 gives percentages of variance for the transformed canonical discriminant function axes comparable to the entries in Table 3.8. From these

TABLE 3.9

Components of Unit Total Sample Variance of Transformed Discriminant Axes Due to Teacher, Classroom, Session, and Episode Differences

Source	I	II	III	IV
Teacher	22.9**	25.7**	4.5*	4.0*
Classroom	.8*	.3	2.1**	1.0**
Session	13.4**	15.8*	7.1**	15.1**
Episode	63.0	58.2	86.2	79.9

Relative Contributions of True Score Differences Between Teachers, Classrooms, Sessions, and Episodes to Observed Variance Among Teacher Means on Transformed Discriminant Axes

Source	I	II	III	IV
Teacher	97.2%	98.2%	79.8%	83.4%
Classroom	1.5%	.5%	16.5%	9.2%
Session	1.1%	1.1%	2.3%	5.9%
Episode	.2%	.2%	1.4%	1.5%

$^{*}p < .05$

$^{**}p < .01$

entries we can see that there is substantial episode-to-episode variation in all four of the classroom interaction contrasts (from 58 to 86 percent of the total variance). Note, however, the large components of variance attributable to teachers (Table 3.9, upper section).

Teacher differences account for approximately one-fourth of the total observed variance on each of the first two transformed axes (I and II). What this means, in a practical sense, is that we could reduce our uncertainty about which pole of either contrast might appear in any given episode of classroom interaction by a substantial amount merely by knowing which teacher is in charge. If the teacher is high on the first contrast (teacher B in Figure 3.2), we would do well to predict "question-answer-corrective feedback-prompt-answer" interaction episodes. On the other hand, if the teacher is low on the first contrast (teacher F or perhaps E), then we would do well to predict "direct read" and "other" episodes.

Considering the second contrast we would proceed in a similar fashion, predicting the "free response" mode of individualized instruction for the positive pole (teachers D, B, and E), and group "model-practice-corrective feedback-model-practice" chains of episodes for the negative pole (teacher F).

Even the last two transformed canonical variates ("student-student feedback" and "direct read and/or ask question vs. question-answer", respectively) show larger teacher and classroom components of variance in Table 3.9 than do the original factors of classroom interaction from which they are largely derived. However, it is clear that the latter are not major contrasts in teaching styles so we could not use knowledge of average teacher performance on these variates to predict much about individual episodes of classroom interaction. Let us turn, therefore, to a consideration of how reliably the teacher means on these contrasts can be estimated *per se*.

The lower section of Table 3.9 gives a breakdown of the weighted contribution of each factor to observed variation in teacher means on each transformed discriminant function axis. The first two contrasts are almost perfectly reliable indicators of teacher differences; 97 to 98 percent of the observed variation in teacher means can be attributed to true differences among teachers. The last two contrasts are somewhat sensitive to differential interaction in various classrooms taught by the same teacher.

The components of variance discussed above reveal that we are dealing with highly reliable indices of teacher variation; the reliabilities of teacher means on transformed discriminant axes range from .80 to .98 in Table 3.9. But a more importnat feature of this investigation is the evidence it provides that those classroom interaction contrasts which show differential treatment of different classrooms by the same teacher are diagnostic of variation in the ability levels of the classes involved. Whereas our initial impression of Figure 3.2 holds true--teachers do have distinct and consistent styles of interaction with students, there is also evidence that certain more limited aspects of classroom interaction vary from classroom-to-classroom taught by the same teacher. That the latter variation might be in response to student characteristics is of great interest and will be taken up later.

Inter-Observer Reliability

A final question about reliability remains before we can proceed to our main task of relating student achievement to teacher performance: do different observers record the same aspect of variation in classroom interaction patterns? In order to study the issue of possible observer bias, data was collected in some of the early observation sessions by pairs of independent observers observing the same sessions. Since three observers were used in the study, it was arranged to have concurrent observations made by each of the three pairs in eight different classrooms.

We are not in a position to make a strict comparison of inter-observer agreement in the usual sense, since observers worked independently and there is no way to know which of their data-sheet entries should correspond. (The relevance of this correspondence is frequently ignored in estimating inter-observer reliabilities.) However, we can compare data on the basis of individual sessions. In Table 3.10 we have summarized the results of this comparison for each of the three pairs of observers, using as the basic units of analysis, session means on the four transformed classroom interaction contrasts. It is appropriate to look at possible observer influences on these scores because we have used these scores to predict differences in student achievement.

Note that we are discussing observer assessment of session-to-session variation in these interaction patterns, but we have already seen from Table 3.9 that enough sessions of observation were obtained in the day-school sample to yield highly reliable teacher means despite any session-to-session variation. Since each classroom was visited by all observers, the influence of any possible observer biases on classroom mean scores is indicated by session-to-session variation, which as we have seen, accounts for less of the variance than teacher differences (lower section of Table 3.9).

From Table 3.10 it can be seen that session-to-session correlations among transformed discriminant function scores derived from the data collected

TABLE 3.10

Means, Standard Deviations, and Correlations for
Pairs of Observers on Transformed Discriminant Function Scores

Variate		Observer Pairs 1	2	1	3	2	3
I	$\bar{X}$*	.93	.70	.13	.32	.33	.52
	s	2.36	1.89	1.62	1.47	1.45	1.58
	r	.98		.99		.99	
II	$\bar{X}$	.65	.54	.00	.04	.13	.32
	s	1.01	1.13	1.39	1.73	1.17	1.59
	r	.95		.96		.98	
III	$\bar{X}$	.47	.08	.08	.59	.26	.47
	s	1.42	.96	1.82	1.82	1.54	1.03
	r	.66		.81		.37	
PV	$\bar{X}$	.87	.55	-.13	-.09	-.11	.23
	s	2.44	2.29	1.50	1.65	1.83	1.63
	r	.97		.97		.97	

*$\bar{X}$ = mean, s = standard deviation, r = product moment correlation; eight observations.

concurrently by different observers range from .95 to .99 for all but the third axis. For the latter, the correlations suggest that certain observers may have had difficulty either recognizing or recording "student-student feedback in Spanish." Since the lowest correlations involve Observer 2, it is possible that this one individual is the problem;[6] however, none of these correlations are as high as for the other axes.

One problem is that this variable (as we mentioned earlier) occurs relatively infrequently--particularly in the more advanced classes. For this reason each such occurrence is highly weighted (as seen in Figure 3.5) and the failure to record any given instance could have a sizeable influence upon the outcome for that session.

When correlations comparable to those given in Table 3.10 are studied for all nine of the original classroom interaction factors, a similar conclusion is reached; only for "student-student feedback" is there a serious problem with inter-observer correlation. These results are briefly summarized in Table 3.11.

Finally, it can be pointed out that inter-observer correlations for different scores were much lower than for the same score, indicating that the observers were indeed discriminating between different classroom interaction patterns.

Just as for the day-school data, we have plotted classroom centroids on the first two canonical discriminant function axes in Figure 3.7. Night-school teachers have been identified in the figure with the letters G through L and we have also included vectors representing the original nine classroom

[6]Notice from Table 3.1 that observer 2 collected very little data in the study.

TABLE 3.11

Interobserver Correlations for Nine
Classroom Interaction Factors (8 Observations)

Variate	Observer Pairs		
	1 2	1 3	2 3
GF-P-A	.97	.99	.99
DIR READ	.97	.93	.86
MOD PRAC	.96	.98	.99
OTHER	.79	.99	.96
QA vs. DIR	.98	.94	.96
QA-IND	.99	.98	.98
GF-M-P	.82	.96	.52
SFBK	-.13	.42	-.09
FREE	.94	.96	.97

interaction factors. Interpretation of this figure can proceed in the same manner as for Figure 3.2. Again, we find the night-school results to be less intuitively compelling than the day school results.

Relation of Teacher Background Characteristics to Teaching Styles

Considerable information was obtained about each teacher in both the night and day-school samples through a questionnaire administered at the outset of the study. It is of interest to see whether a teacher's performance in the classroom might be related to this prior information about his or her education, experience, and preference among teaching styles.

Only 12 teachers were studied, background characteristics were often badly skewed (e.g., number of years of teaching in Table 2.7), and information was available for more background variables than teachers. We therefore decided to use rather crude data reduction techniques in order to break these data down to a minimal set of contrasts among teachers. The outcome of this crude approach to metric multidimensional scaling is presented in Figure 3.8. In this figure teachers are represented by the same letters used in earlier figures (A-F for day school; G-L for night school) and their background characteristics are represented as vectors.

Despite the approximate nature of the analyses under discussion, Figure 3.8 reveals interesting information about the teachers included in this study. The horizontal axis is aligned with the variable, "percent audio-lingual", and it marks a rather strong contrast between day-school and night-school teachers. Only one day-school teacher, J, is placed toward the "silent way" (left) pole of the horizontal axis.

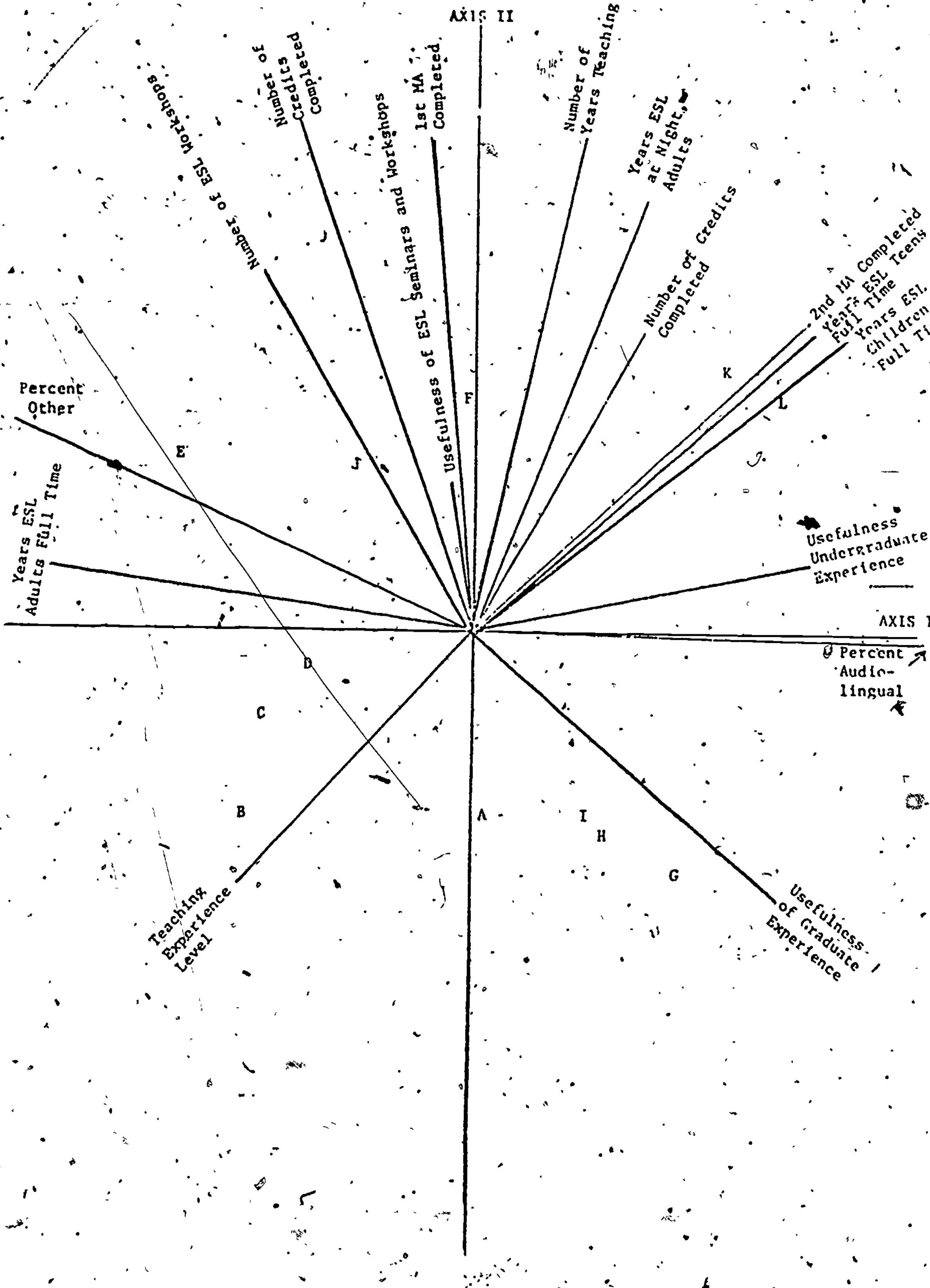

Figure 3.8

Multidimensional Scaling of ESL Teachers on the Basis of Similarity in Their Background Characteristics; Day School Teachers (A-F), Night School Teachers (G-L).

The vertical axis is clearly an indication of educational level which differentiates teachers within both the day and night-school samples. "Years of teaching" is closely aligned with this vertical axis but amount of experience teaching English as a second language to adults ("ESL day adult") is aligned with the "silent way" pole of the horizontal axis. In other words, the more experience a teacher has had teaching English as a second language to adults, the more likely he or she is to use or approximate "the silent way." The more experience a teacher has had in teaching in the elementary or secondary schools, the more likely they are to use the audiolingual method or an approximation to it. It seems that the lower the level (from adult to child) at which an individual has had experience teaching, the more likely they are to be educated, experienced, and a practitioner of the audiolingual method (e.g., teacher K and L).

Comparing Figure 3.8 to Figure 3.2, there appears to be some association between classroom interaction patterns (teacher performance) and teacher background characteristics. Notice that the day-school teachers fall into roughly comparable circular patterns (from A through F) in Figures 3.8 and 3.2.

THE RELATION OF TEACHING PERFORMANCE TO STUDENT LEARNING

The major problem to be solved in the analysis of the student achievement data was to determine how we could account for variation in posttest scores. There are three domains of variables which are potential predictors of posttest scores: background characteristics of the students, students' initial level of proficiency as measured by the pretests, and classroom experiences.

Variables describing the students' backgrounds are worth considering as potential predictors of final achievement because these measures may be indirect indicators of aptitude for learning, of academic skill, or of prior achievement of proficiency in English. Obviously, the students' initial proficiency (the second domain mentioned above) may be associated with their final achievement status. Pre- and post-instruction scores on the same test are usually highly correlated with each other because the experience acquired in the interval which separates these measures does not greatly alter the related order of students with respect to their abilities; the latter have, of course, been built up over an entire lifetime of experiences, for which the background measures are indicators or proxies. Nevertheless, the relationship between pre- and posttest scores may be altered through the influence of intervening events, including classroom interaction experiences. Classroom interactions is the third domain of predictor variables and is of most interest in this study because it is the only domain over which some degree of control can be exercised.

Other relevant experiences which might intervene between pre- and posttesting (such as use of English at home, on the job, and in the community) unfortunately could not be objectively measured in this study. The possibility of controlling these extracurricular sources of experience is slight in any case, but it would be useful from a theoretical point of view to take them into account. The best we can do at this point, however, is to bear in mind that certain "background" characteristics might serve as proxies for sustained extra-curricular experiences (e.g., occupational level for the need to speak English in the workplace; length of time in the United States for assimilation into an ethnic community).

Given these three domains of predictor variables, which are organized in an obvious temporal sequence (background experiences ⟶ pretest performance ⟶ classroom instruction), the analytic problem is to find how the information can be most parsimoniously combined to predict final achievement. Multiple linear regression can be applied to this task, since the squared multiple correlation (R^2) between a set of predictors and a criterion variable indicates what proportion of the observed variance in the criterion can be accounted for by a given set of predictors.

Background characteristics must be taken into account first because these variables are direct measures of or proxies for educational attainment, competence to cope with the processes of schooling, motivation and aptitude. If such variables account for most of the variance in final scores, then classroom experiences can have little differential influence on final status; this is not to say that large pre- to posttest gains could not have occurred, only that they are not likely to have changed the relative ordering of students in terms of achievement.

Next we must consider how much initial status in terms of pretest scores adds to the prediction of outcomes beyond what we have been able to learn from a knowledge of background characteristics alone. A related issue here is how well initial status *per se* can be predicted from background characteristics. This analysis tells us how adequate and useful our information about background experiences is in the first place.

The final step in the regression analysis is to assess the unique contribution of classroom experiences in the prediction of final achievement, above and beyond any predictive utility of initial status and background experiences. We also must determine whether different forms of classroom interaction appear to have different effects upon achievement. The purpose of relating classroom interaction to achievement is to find out how much this information adds to the prediction of achievement and to identify any components of classroom interaction which can be hypothesized to have an impact upon specific forms of achievement.

Notice that the classroom experience variables differ fundamentally from all of the other variables in the analysis in that they are not individual measures, but apply to all individuals in each classroom equally. From the point of view of this study, these variables are the "independent" variables whose possible effect upon final achievement we are most interested in learning. These variables take the form of contrasts among classrooms in terms of observed patterns of teacher-student interaction, the four transformed canonical variates developed and discussed in Chapter Three. Every effort was made to arrive at a small set of independent variables which are reliable and not too highly intercorrelated, in line with the requirements of the multiple regression model. On the other hand, the background and pretest variables are more error prone; but they only play the role of covariates to adjust for preexisting differences among classrooms in the final analysis. Moreover, these data are available for every individual, so stable estimates of the required regression parameters can be obtained even with fallible measures.

Predicting Fall and Spring Student Achievement from Student Background Characteristics

In Table 2.5 we saw many significant zero-order correlations between student characteristics and student achievement scores, both in the fall and in the spring. The background and achievement domains are clearly related at the level of individual pairs of variables, so it is in order to see how background information can be optimally combined (weighted) to predict individual achievement scores and to determine how effective this prediction can be when simultaneous use is made of all available background information.

Each achievement test score was predicted from the background characteristics using multiple linear regression. The results of these analyses for the day-school sample are summarized in Table 4.1. In that table the labels across the tops of the columns designate the fall and spring achievement scores; note that logarithmic transformations of the oral proficiency scores were used. The rows of the table list particular background characteristics used in the regression analyses. The number in any cell of this matrix is the standardized regression weight for a particular background variable (row) as a predictor of a given achievement score (column). These entries complement the zero-order correlations already given in Table 2.5. The former show the direct contribution of variation in each background characteristic to achievement, whereas the entries in Table 2.5 do not take into account the fact that the background characteristics are correlated among themselves.

The line in Table 4.1 labeled R^2 gives the squared multiple correlation of each achievement test score with all of the background variables, taken simultaneously. These numbers indicate the proportion of the variance in each achievement score predicted by the complete set of background characteristic variables; for example, the background characteristic variables account for 26 percent (.26) of the variance in the fall Decoding 1 scores; for 13 (.13) in the fall Decoding 2 scores; for 44 percent (.44) of the variance in the John Test scores, and so on.

TABLE 4.1

Standardized Regression Weights, Squared Multiple Correlations and Statistical Tests for Background Characteristics as Predictors of Pretest and Posttest Scores; Day-School Sample

	Pretests					Posttests					
	D_1-F	D_2-F	L-F	J-F	M-F	D_1-S	D_2-S	E-S	Log_{10} P-CT	Log_{10} P-CP	Log_{10} P-ST
Sex	.03	-.01	-.02	.03	-.01	-.01	-.09	-.14	.08	.11	.06
Age	$-.32^{**}$	$-.22^{*}$	$-.31^{***}$	-.16	.03	$-.26^{*}$	-.06	$-.29^{**}$	-.17	$-.20^{*}$	$-.20^{*}$
Time in U. S.	-.12	.08	$.27^{**}$	$.22^{*}$	-.04	-.14	-.07	.10	.04	.05	.04
Former Job	$.22^{*}$	.10	$.21^{*}$	.13	$.20^{*}$	$.27^{*}$	.15	.13	$.23^{*}$	$.28^{**}$	$.26^{*}$
Job in U. S.	.00	-.02	.00	.00	.08	$.19^{*}$	.12	.13	-.01	.02	.01
C-Origin	-.03	-.17	-.01	-.11	-.16	-.05	.07	-.03	-.02	-.07	.04
Years Education	.05	-.02	.14	-.01	.01	.07	.09	.09	.03	.02	.03
Former Eng.	.11	$.22^{*}$	$.22^{*}$	$.27^{**}$	.11	-.05	.04	.09	.07	.07	.10
Eng. in U. S.	$.18^{*}$	.11	$.38^{***}$	$.52^{***}$	$.37^{***}$	.11	-.05	$.24^{*}$	$.24^{*}$	$.34^{***}$	$.27^{**}$
R^2	.26	.13	.47	.44	.29	.25	.08	.25	.20	.33	.24
F (9, 71)	2.76^{**}	1.14	7.13^{***}	6.25^{***}	3.23^{**}	2.66^{*}	.73	2.58^{*}	1.94	3.82^{***}	2.51^{*}

$^{*}p < .05$

$^{**}p < .01$

$^{***}p < .001$

Labels Code:

D_1: Decoding, Part 1
D_2: Decoding, Part 2
L : Literacy Test
J : John Test
M : Morano Test
P : Oral Proficiency Test
CT: Correctness Score
CP: Comprehension Score
ST: Structure Score
F : Fall
S : Spring

The F-ratios presented in Table 4.1 (degrees of freedom in parentheses) provide a statistical test of the null hypothesis of no predictability of each dependent variable (Achievement Test score) from the set of independent variables (background variables). In other words, is the mean score on each test the best available estimator of how any given student will perform, or can we learn more about that performance by taking his or her background characteristics into account? The probability levels associated with these statistics are coded with asterisks.

A prediction equation could be written with the regression weights in any given column of Table 4.1 as the coefficients of the background variables. The actual standard scores of a particular student for each of these variables would be entered in this equation to predict his or her standardized score on the corresponding achievement test. It can be seen that variables with larger weights have more influence on the predicted outcome. The following example will illustrate this concept. We will use the numbers in the column labeled P-CT, the correctness score on the Oral Proficiency Test which was taken in the spring. The prediction equation for this score after logarithmic transformation is:

(1) $$Z_{\log_{10}(P\text{-}CT)} = .08\ Z_S - .17\ Z_A + .04\ Z_T + .23\ Z_{FJ} - .01\ Z_J - .02\ Z_{CO} + .03\ Z_{YE} + .07\ Z_{FE} + .24\ Z_E$$

As will be seen shortly, two of the above weights are significantly different from zero in the statistical sense. The three which are underlined in the P-CT column of Table 4.1 are large enough to give some feeling for which background factors might affect correctness scores (P-CT). They are age, which has a negative weight (- .17), status of job in former country (.23) and amount of English taken in the U. S. (.24). In other words, students who are younger, had better jobs in their former country, and have taken more course work to learn English after arriving in this country will have higher predicted correctness scores.

There are two other features of the information in Table 4.1 which should be noted. Those regression coefficients which are significantly different from zero in a statistical sense are marked with asterisks. In addition to significance in a statistical sense, however, it is important to consider how much each predictor variable contributes to the total predictability signified by R^2, without regard to sample size. The regression weights for those variables which contribute at least .01 to the magnitude of the squared multiple correlation, R^2, have therefore been underlined in Table 4.1 (i.e., these variables account for at least one percent of the total variance in the Achievement Test score).

Implications of the Regressions of Achievement Test Scores on Background Characteristics

The R^2 for prediction equation (1) is rather low (.20). This equation is not a very accurate predictor of Oral Proficiency correctness scores. We may still speculate, however, that students with three background characteristics (Youth, Former Job, Eng. in U. S.) are likely to have higher correctness scores in general as long as we remember that the prediction equation will be highly in error in any given instance.

By way of contrast, consider the comprehension scores (P-CP). The prediction equation in this case is likely to be a little more accurate (R^2 = .33 for comprehension vs. .20 for correctness). In that equation amount of English studied in the U. S. has the largest weight. Three other characteristics are worth noting: age, which again has a negative weight (- .20), status of job in former country (.28), and amount of English in the U. S. (.34).

Background characteristics predict comprehension scores best ($R^2 = .33$), structure scores less well ($R^2 = .24$) and correctness scores least accurately. A reasonable hypothesis from these data, however, is that those students will be more proficient in spoken English by the end of the year who are younger, had a higher status job in their former country, and have taken more English courses or programs since coming to the United States.

It should be remembered that predictive background characteristics do not necessarily "cause" higher or lower scores. Age, for example, does not necessarily make a person less proficient. But age is associated with several other characteristics which, taken together, may give us some idea about why age is a negative predictor of proficiency. Older students are more likely to have a low status job in this country, hence, may have less opportunity to speak English. They are more likely to have had course work in English in the previous country than in the U. S., hence, may have learned English originally as a school subject rather than as something to be used in their daily life.

Amount of English studied in this country is more likely to be a direct influence on achieving proficiency. But, again, students who have had more work in English here are also younger, have arrived more recently, and have a history of higher level employment both here and abroad. They probably have both more need to speak English and more opportunity to do so. Thus a combination of previously acquired proficiency and opportunity to speak English is likely to facilitate the progress of these students during their time at the Adult Learning Center.

Differential Prediction of Fall and Spring
Achievement from Student Background Characteristics

Of the five squared multiple correlations for predicting fall tests from background variables, all but one are statistically significant at the .01 level. Note also that the multiple correlations are generally lower in magnitude and of lower statistical significance for the spring tests. These results suggest that the role played by background experiences in achievement is substantially diminished over the course of the year, as new experience with English as a second language is acquired through classroom instruction and extracurricular activities. That students gain substantially in their average test scores has already been seen in Table 2.3.

That background characteristics do help predict performance in the fall (R^2 = .44 for the John Test; R^2 = .47 for the Literacy Test) but that their predictive utility is diminished in the spring (R^2 = .33 for Proficiency-Comprehension; R^2 = .25 for Literacy) is the most important piece of information to note from Table 4.1. Only the comprehension score on the spring Proficiency Test retains a highly significant relationship to background characteristics. This fall-spring distinction suggests that something beyond the students' background experiences prior to instruction accounts for their final spring achievement. The next question is whether or not we can identify any component of classroom interaction which accounts for the change in achievement.

Note that the regression weights for "age" are generally negative in sign. In this case we see weights which are significantly different from zero in predicting both fall and spring achievement scores, just as for former employment level and English studied in the United States.

It is obvious from this analysis that a number of background characteristics are significant predictors of fall achievement and a few of them are of importance for predicting spring scores. Most of the multiple correlations are significant when all of the predictors are used in the regression analysis, regardless of whether fall or spring achievement is the criterion.

Predicting Posttest Scores Using Pretest Scores in Addition to Background Characteristics

The next step in the analysis was to predict the posttest scores using information about the pretest scores in addition to information about background characteristics. The results of these analyses for the day-school sample are presented in Table 4.2. This table is read in much the same way as Table 4.1. Across the top of the columns are listed the posttest scores, beginning with the two Decoding scores, then the Literacy score, and the last three columns contain logarithmic transformations of the three scores on the Oral Proficiency Test. The predicted scores used in this analysis were all scores taken from tests administered in the spring.

Down the left-hand column of the middle section of Table 4.2 are listed the labels for the pretest scores; included are the two parts of the Decoding Test, the Literacy Test, the John Test, and the Morano Test. The first line of this table gives the squared multiple correlations resulting from prediction of posttest scores from background characteristics; these numbers are repeated from the right-hand portion of Table 4.1. The entries in the line labeled, "R^2 with Addition of Pretests," can be compared to the corresponding entries in the first line. Note under D_1-S that R^2 when only background characteristics

TABLE 4.2

Standardized Regression Weights, Squared Multiple Correlations and Statistical Tests for Posttest Scores Predicted from Pretest Scores Adjusted for Background Characteristics; Day-School Sample

	D_1-S	D_2-S	L-S	P-GT	P-CP	P-ST
R^2 from Background	.25	.08	.25	.20	.33	.24
D_1-F	.64***	.24	-.16	.27**	.29***	.25**
D_2-F	.08	.15	.10	-.01	.09	.01
L-F	-.07	.07	.47*	.09	.02	.07
J-F	.00	-.14	.39*	.49***	.60***	.52***
M-F	.07	.03	-.04	.09	-.06	.11
R^2 with Addition of Pretests	.60	.19	.61	.72	.82	.76
F (5, 66) Test of Information Increase	11.17***	1.72	12.23***	24.83***	35.86***	27.85***

* $p < .05$

** $p < .01$

*** $p < .001$

are used is .25, but R^2 increases to .60 when the pretest information is added. In other words, if only background characteristics are used to predict posttest scores, then 25 percent of the variance in the Decoding 1 posttest is accounted for. But an additional 35 percent is accounted for if information from the pretest scores is added. Note, for example, that the R^2 for predicting transformed Oral Proficiency comprehension scores from background characteristics is only .33, but with the addition of pretest information it becomes .82. A similar pattern is apparent for the other two scores of the Oral Proficiency Test. Thus, the students' initial proficiency plus some information about their background characteristics accounts for a substantial proportion of the variance in the comprehension scores. In summary, to understand how much the prediction is improved by adding pretest data to background information, simply compare the R^2 in the first line with the R^2 in the third portion of the table.

It should be noted that most of the squared multiple correlations for background characteristics plus pretest scores are substantial (.60 to .82). Moreover, the increase in all of the R^2's except for part two of the Decoding Test are highly significant when compared to prediction from background characteristics alone.

Except for part two of the Decoding Test, all of the squared multiple correlations of posttest scores with background and pretest variables are at or above .60. From Table 2.3 we can see that the posttest reliabilities for all except Decoding 2 range from .88 to .94 (Decoding 2 has lower internal consistency, yielding a reliability of .77). If we take these reliabilities to mean that roughly 90 percent of the variance in most posttest scores is reliable, then it can be said that over 60/90 or two-thirds of that variance can be accounted for by the background and pretest variables which we have measured. This leaves up to one-third of the reliable posttest variance which might be accounted for by classroom experiences. However, we must keep in mind that we do not have separate measures of classroom experience for each of the students in any classroom—only contrasts between classrooms. This means that we cannot expect to increase the accuracy of our prediction to the limit of the reliability of posttest measures on the basis of classroom experiences. Even if classroom experiences are highly influential, their effect cannot be detected unless there is sufficient homogenity among students within classrooms with respect to these experiences.

Posttest Scores Predicted from Classroom Interaction, with Adjustment for Pretest Scores and Background Characteristics

Table 4.3 presents information from the regression analyses for the day-school sample in which posttest scores were predicted from classroom interaction contrasts, pretest scores, and background characteristics. This table is organized in the same way as Table 4.2.

The first line of Table 4.3 gives R^2 from the regression of posttest scores on background characteristics and pretest scores, as seen already in Table 4.2. Next are four lines with Roman numerals I through IV; within each line can be found the standardized regression weight of the corresponding transformed canonical variate (the four "higher order" bipolar contrasts described earlier and in Chapter Three).

The sixth line in Table 4.3 gives R^2 once the four transformed canonical variate classroom interaction contrasts have been included in each prediction

TABLE 4.3

Standardized Regression Weights, Squared Multiple Correlations and Statistical Significance Tests for Posttest Scores Predicted from Classroom Interaction Contrasts, Adjusted for Student Background Characteristics and Pretest Scores; Day-School Sample

	D_1-S	D_2-S	L-S	$\log_{10}$ P-CT	$\log_{10}$ P-CP	$\log_{10}$ P-ST
R^2 for Background and Pretests	.60	.19	.61	.72	.82	.76
I	-.27	-.29	-.28	.25	.37$^+$	.40$^+$
II	-.22	-.23	-.34*	.28$^+$	.34*	.21
III	.40$^+$	.27	.42$^+$	-.18	-.36$^+$	-.19
IV	.32**	.25*	.27*	.10	.02	.09
R^2 with Addition of Classroom Interaction Contrasts	.65	.28	.66	.76	.85	.81
F (4, 62) Test of Information Increase	2.26$^+$	2.03$^+$	2.35$^+$	2.74*	2.73*	3.91**

$^+p < .10$

$^*p < .05$

$^{**}p < .01$

equation. Again, the lines containing the initial and final R^2 values can be compared. From this comparison we can see that the addition of classroom interaction contrasts adds little in an absolute sense to the accuracy of prediction of posttest scores (increases in R^2 vary from .04 to .09).

Despite the relatively small absolute increases in the accuracy of the prediction obtained by supplementing background and pretest information with classroom interaction information, it can be seen from the F-test results in Table 4.3 that the increases in predictability are statistically significant.

As we pointed out earlier, the classroom interaction contrasts apply equally to all individuals within each classroom, so great homogeneity of experiences would be required in order for these contrasts to yield a large absolute increase in the accuracy of prediction ~~of posttest achievement~~. The F-test results relate, however, to relative increases in accuracy of prediction. These relative increases in the accuracy of prediction are substantial enough, even given our small sample size, to support further exploration of hypotheses about classroom interaction.

Recall that the first two classroom interaction contrasts reflect broad stylistic differences among teachers. The negative pole of each contrast relates to more highly structured class or group-oriented activities. The prevalence of "teacher model-student practice" interaction patterns on the negative poles suggests that a variant of the audiolingual method was being used. But, the "Free" and "Other" patterns of classroom interaction on the positive pole of the second transformed axis suggests that this variate is a contrast between the silent way and audiolingual methods. The flexible, individualized, supportive, and encouraging aspects of the positive pole of the first axis suggests an eclectic style of classroom interaction.

The third axis relates most directly to the occurrence of "student-student feedback" during instruction. The fourth axis can be ignored for the moment.

The most striking aspect of the pattern of standardized regression coefficients presented in Table 4.3 is that it appears that those features of classroom interaction which are associated with higher than would otherwise be expected posttest scores on the Oral Proficiency Test (i.e., axes I and II) are associated with lower than would otherwise be expected posttest scores on the Literacy and Decoding Tests (compare the weights in the left-hand columns of Table 4.3 opposite I and II with those in the right-hand columns opposite I and II). If we were to take these coefficients as the basis for formulating hypotheses about the effectiveness of individualized instruction vs. group instruction, the silent way vs. the audiolingual method, an open and supportive vs. a highly organized and directive classroom climate, then we should expect the former (individualized, supportive, silent way) to lead to increased oral proficiency while expecting the latter (grouped, directive, audiolingual) to lead to increased literacy and decoding skills.

It appears, therefore, that there is a trade-off in terms of the achievement goals--what appears to be incremental for oral proficiency appears to be detrimental for literacy and decoding, and vice versa. The same can be said for axis III, "student-student feedback," since its occurrence in a classroom is predictive of lower than might otherwise be expected oral proficiency (mainly comprehension) but higher than might otherwise be expected literacy and decoding.

As for that ubiquitous bipolar contrast in classroom interaction patterns, "teacher direct-student read and/or ask question" vs. "teacher question-student answer," we again see it playing an important role. In this case there is no apparent trade-off, however. On the contrary, the data suggest quite clearly that those classrooms in which students are directed to read are also those classrooms in which we can predict that the students will achieve higher than would otherwise be expected on the Literacy and Decoding posttests. The contrast seems to bear no real relationship to oral proficiency, suggesting that nothing is to be lost and literacy and decoding skills may well be gained by using the "teacher direct-student read and/or ask questions" paradigm (line labeled IV in Table 4.3).

The Analysis of Differences in Student Achievement by Classrooms

We can now carry the process one step further than was done with the data as presented in Table 4.3, by allowing parameter estimates for <u>any possible</u> differences between classrooms or teachers. Thus we no longer restrict ourselves to the four classroom contrasts derived from observations of teacher and student interactions over the course of the study.

Predicting Posttest Scores Using All Possible
Contrasts Among Classrooms with Adjustment for
Student Background Characteristics and Pretest Scores

In the first line of Table 4.4 we have presented again the squared multiple correlations of posttest variables with background and pretest scores. These entries are the same as those seen on the first line of Table 4.3. The second line of Table 4.4 contains comparable R^2 values. These R^2's were obtained by entering a complete set of 13 contrasts (one for each degree of freedom) among classrooms into the prediction equation. The F-tests of information-increase associated with these R^2 values are all statistically significant, <u>indicating that there are real differences in student posttest achievement among the various classrooms in the study.</u>

TABLE 4.4

Regression Analysis Results when Possible Classroom Effects are Entered; Compared to Results when only Four Classroom Interaction Contrasts are Entered; Day-School Sample

	D_1-S	D_2-S	L-S	$\log_{10}$ P-CT	$\log_{10}$ P-CP	$\log_{10}$ P-ST
R^2 from Background and Pretests	.60	.19	.61	.72	.82	.76
R^2 from Background, Pretests and full set of Classroom Contrasts	.73	.48	.92	.82	.88	.86
Overall F (13, 53) test of Adjusted Classroom Means	1.95*	2.21*	14.96***	2.19*	1.95*	2.99**
I	- 9.34	-15.63+	- 1.22	.22	.35	.44
II	4.92	-27.90+	- .30	.70	.73+	.22
III	5.09	26.77*	- 2.93	- .12	- .48	.12
IV	13.94**	9.67+	16.94***	.24	.00	.17
% Reduction in Residual Sum of Squares per df, Canonical Classroom Contrasts	10.0	8.3	2.6	5.7	6.5	6.9
F (4, 53) Contribution of Canonical Contrasts	2.53*	2.38+	5.10**	1.62	1.64	2.70*
% Reduction in Residual Sum of Squares per df, Remaining Classroom Contrasts	6.7	7.4	9.9	8.6	8.2	8.0
F (9, 53) Contribution of Remaining Contrasts	1.70	2.15*	17.99***	1.80+	1.51	2.26*

+ $p < .10$
* $p < .05$
** $p < .01$
*** $p < .001$

The last four lines in Table 4.4 contain information about the separate contributions of our original four classroom interaction contrasts (transformed canonical variates), as well as the nine possible remaining classroom contrasts, to variation in adjusted classroom means on the respective posttests.

From the last four lines of Table 4.4 we can learn a good deal about the possible bases of differences among classrooms on posttest performance, beyond what is accounted for by the background characteristics and pretest performance of the students involved. Only for the Decoding 1 posttest is it safe to conclude that the four classroom interaction contrasts of the classroom interaction data are sufficient to account for all classroom differences in adjusted posttest performance.

From the last two lines of Table 4.4 we can see that the Literacy, Proficiency Structure, Decoding 2, and, perhaps, Proficiency Correctness posttests all show classroom-to-classroom variation in adjusted means which must be attributed to something beyond the reliably measured aspects of classroom interaction embodied in the four transformed canonical variates. This fact provides justification for looking at mean performance data for the individual classrooms in order to seek an explanation for their differences.

Simple Comparison of Pretest and Posttest Performance in Day-School Classrooms

In order to study the mean performance of individual classrooms we reverted to simple analysis of covariance procedures in which each pretest was the sole covariate for its respective posttest. An exception: in the case of the Oral Proficiency Test, only the correctness score was investigated and both the

Decoding 1 and the John Test were considered, individually, as covariates. For the sake of completeness, the Decoding 1 Tests were also studied with simple analysis of covariance, even though there is no evidence in Table 4.4 for classroom differences beyond the four classroom interaction contrasts.

Figures 4.1 through 4.5 are plots of the 14 day-school classroom means on pretest scores (horizontal axis) vs. posttest scores (vertical axis) for the Decoding 1, Decoding 2, Literacy, and Proficiency Tests (the proficiency means are in the $\log_{10}$ units and are plotted against John Test scores in Figure 4.4 and fall Decoding 1 scores in Figure 4.5 because the Proficiency Test had not been administered as a pretest). Individual classrooms are coded in the figures by the same letter and number system used in earlier figures.

The line drawn through the points in each figure is the pooled-within-classrooms regression line, and may be taken as the point of reference for determining the adjusted gains or losses of individual classrooms. The way to use this figure is as follows (refer to Figure 4.4):

1. Along the horizontal axis are units of scores on the John Test. Assume that a <u>class</u> had a score of 20 on the John Test in the fall.
2. From this point, draw a vertical line to the regression line, and from where this vertical line meets the regression line draw a horizontal line to the Proficiency Score axis. When this procedure is followed, the horizontal line from the regression line would intersect the vertical axis at about .75.
3. This score (.75) is the predicted score for classes whose mean score on the John Test in the fall was about 20.

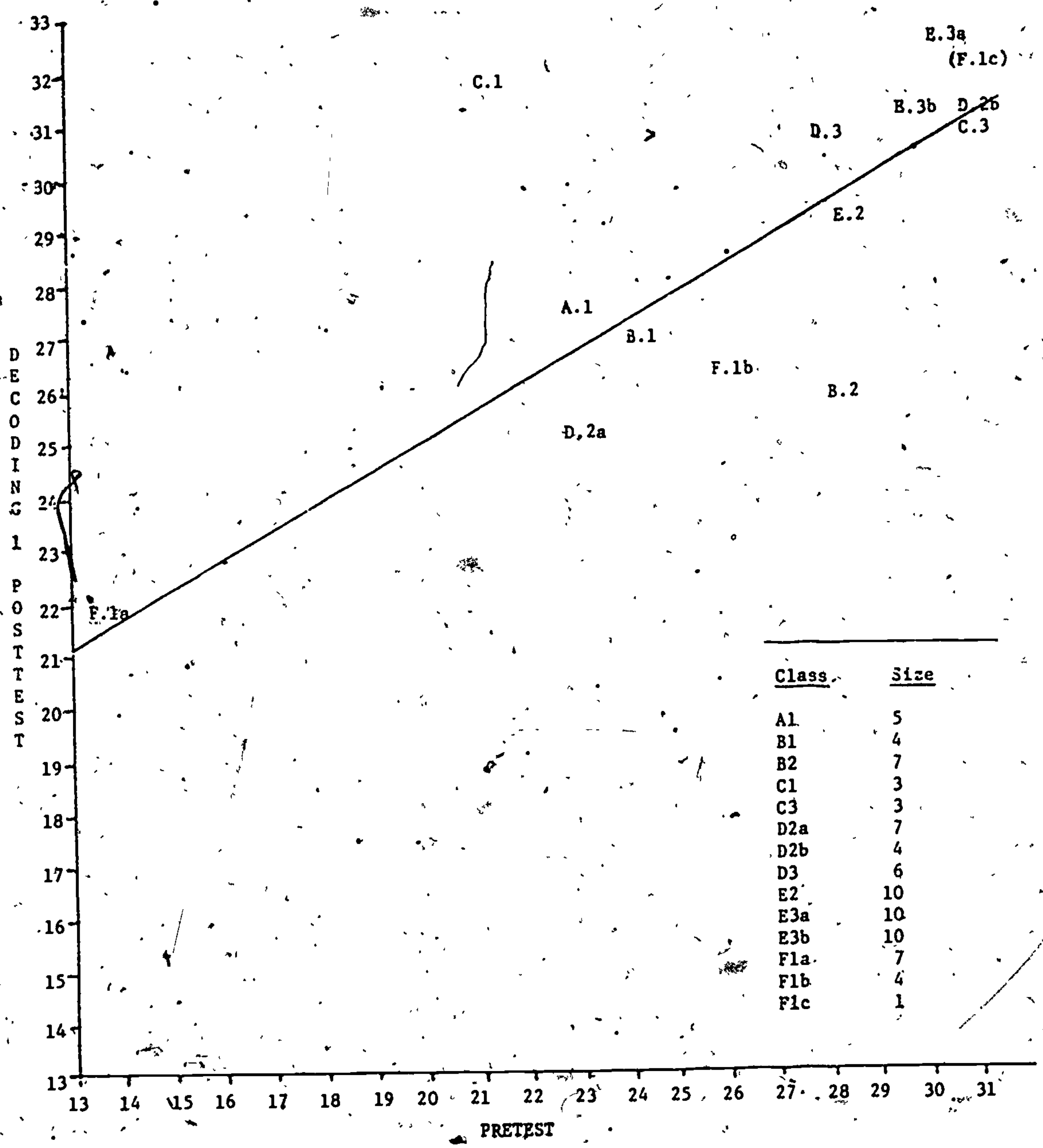

FIGURE 4.1

Decoding 1 Classroom Means for Fall and Spring Administrations, Day-School Sample (Pooled Within-Classrooms r = .69; Total Sample r = .73)

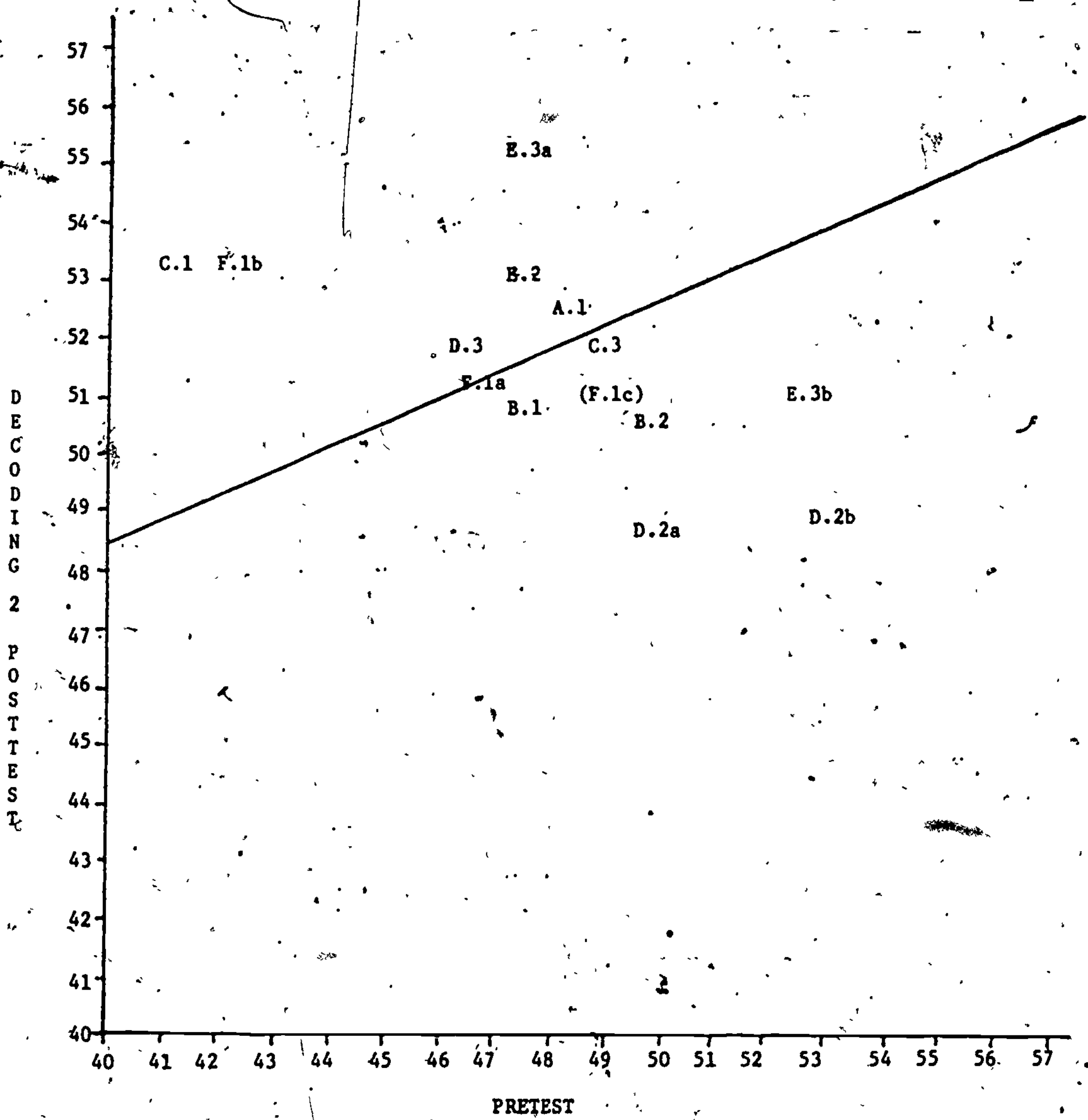

FIGURE 4.2

Decoding 2 Classroom Means for Fall and Spring Administrations, Day-School Sample (Pooled Within-Classrooms r = .48; Total Sample r = .27)

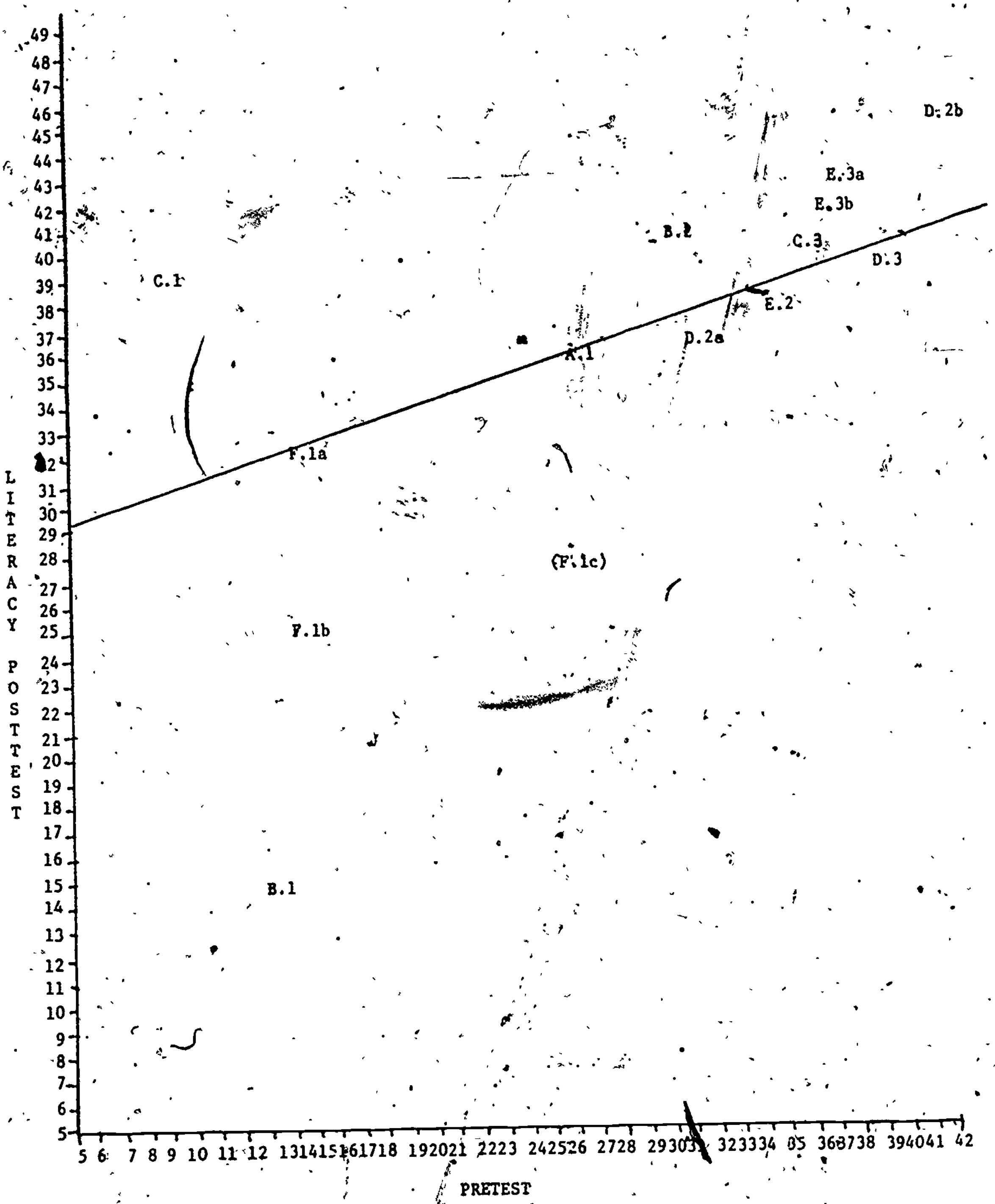

FIGURE 4.3

Literacy Classroom Means for Fall and Spring Administrations, Day-School Sample (Pooled Within-Classrooms r = .59; Total Sample r = .71)

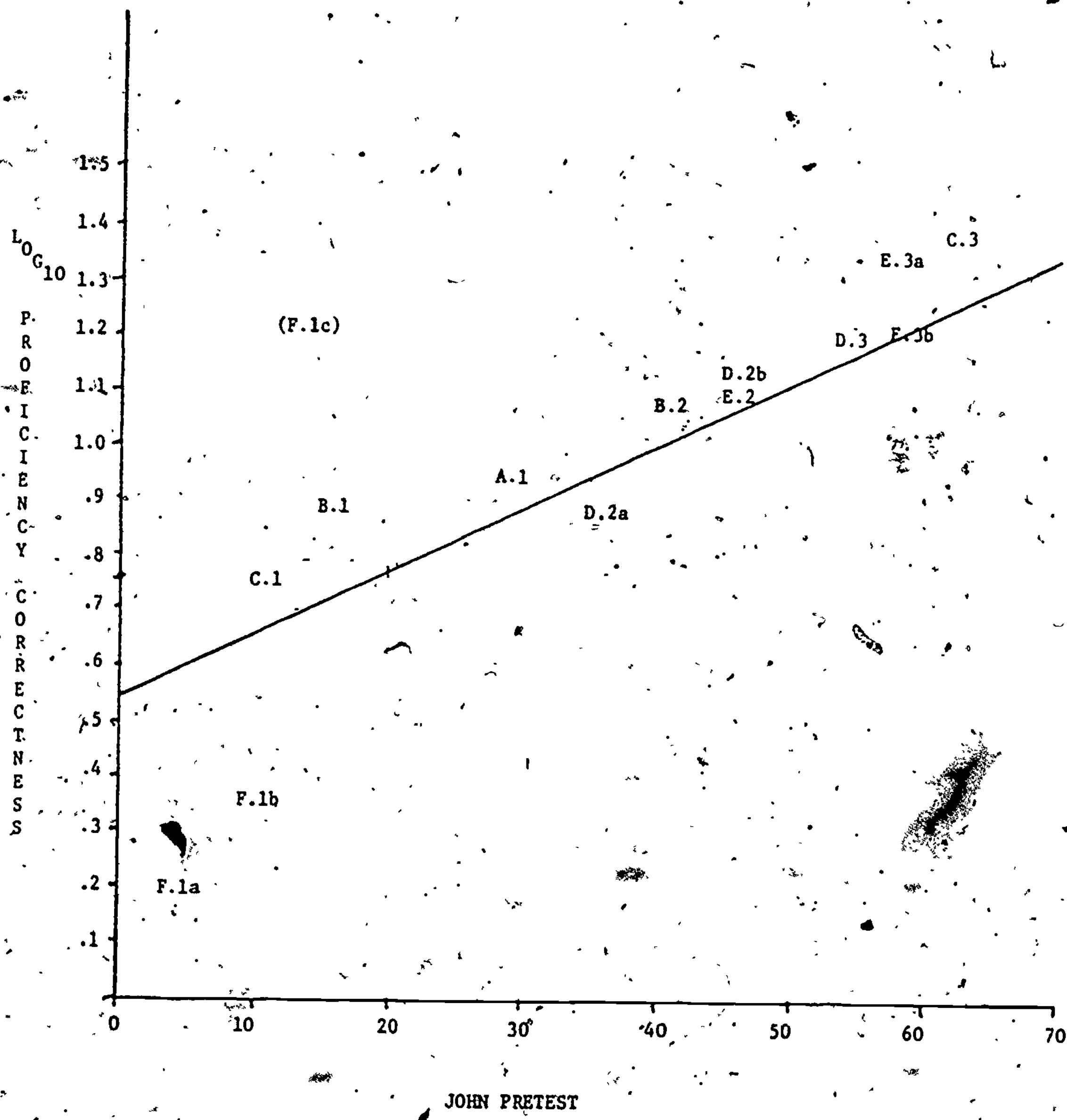

FIGURE 4.4

Log_{10} Spring Proficiency Correctness Classroom Mean Scores x Fall John Test Classroom Mean Scores, Day-School Sample (Pooled Within-Groups r = .28, Total Sample r = .77)

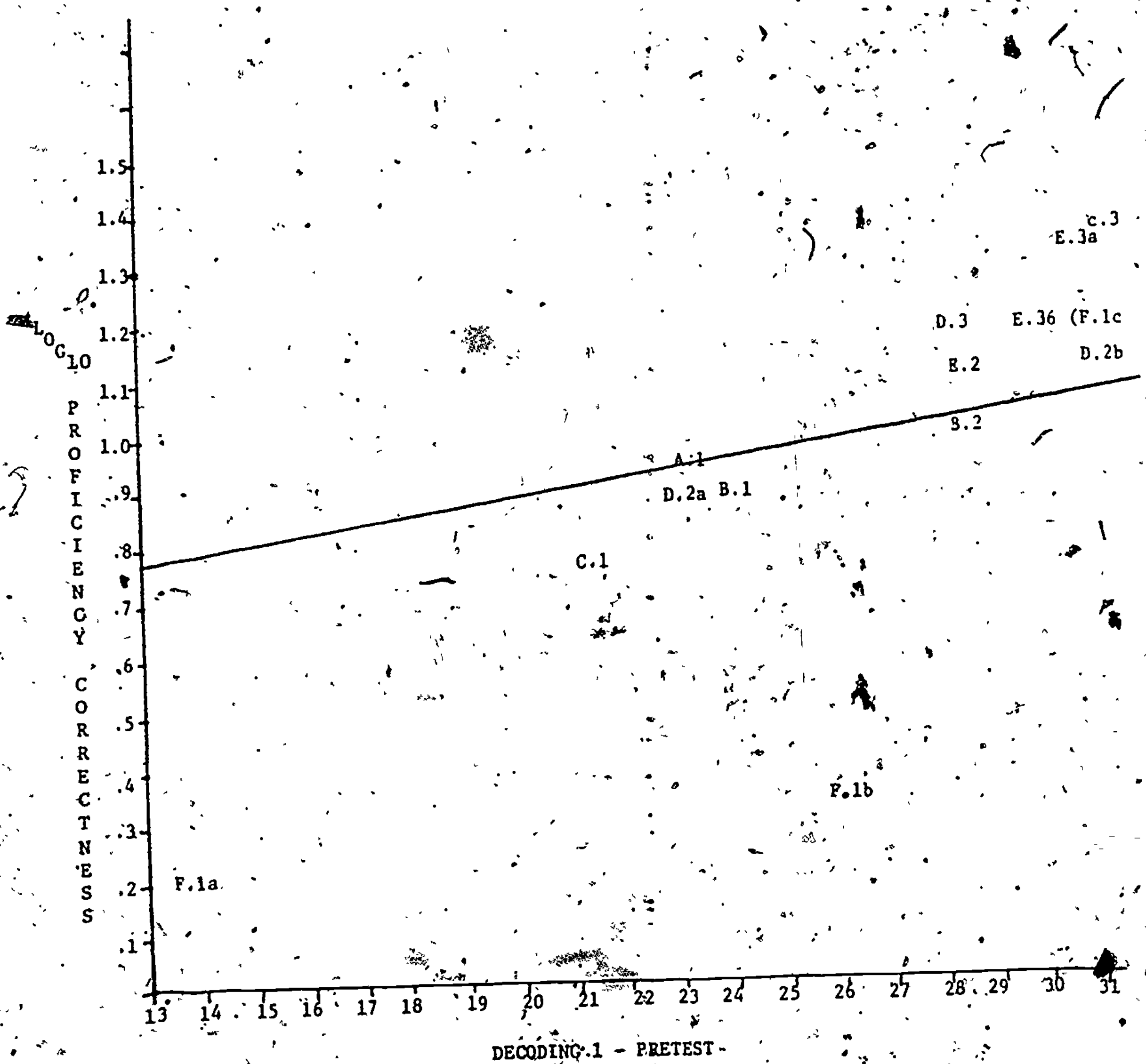

FIGURE 4.5

Log_{10} Spring Proficiency Correctness Classroom Mean Scores x Fall Decoding 1 Classroom Mean Scores, Day-School Sample (Pooled Within-Groups r = .44, Total Sample r = .67)

4. Thus, points on the regression line represent predicted spring scores on the test being considered.
5. The points in the figure around the regression line represent the actual scores. A.1, for example, has fallen above the line. Its actual spring score is better than would be predicted on the basis of the fall John Test mean for this class.
6. As can be seen, some classes fall above the line (F.1c, B.1, C.3 and others); others fall below (F.1a, F.1b and others).

These figures may be used in several different ways: (1) to compare the same class on different measures; (2) to compare different classes of the same teacher; and (3) to compare the performances of classes at different levels of proficiency. Using Figures 4.3 and 4.4, compare the relative positions of F.1c on the Literacy and Proficiency correctness measures. This class is doing better than predicted on the Proficiency measure and poorer on the Literacy measure. Now note in Figure 4.4 the three classes of teacher F; two are doing poorer than predicted (F.1a and F.1b) and one better than predicted (F.1c). Again referring to Figure 4.4 we see that 4 of the 6 Level 1 classes are doing better than predicted; 3 out of 4 of the Level 2 classes and all of the Level 3 classes are doing as well as or better than predicted.

Differences in Achievement of Literacy

Referring back to Table 4.4, we see that the Literacy Test shows evidence of much variation which cannot be accounted for by reference solely to the four reliable classroom interaction contrasts. From Figure 4.3 we see that this variation is probably due to the exceptionally good posttest performance of the level 1 students in classroom C1. These students begin the year at the very lowest level of performance on the Literacy Test, but, by the end of the year, they are as literate as students in several level 2 and level 3 classrooms. This performance contrasts markedly with that of students in classroom B1, where spring Literacy Test performance is essentially unchanged from that in the fall. Obviously, teacher C is increasing the English language literacy of level 1 students more than other teachers at that level. A similar conclusion holds for Decoding 1 in Figure 4.1.

Differences in Achievement of Decoding Skills

The Decoding 2 test has been revealed to be a bit anomalous in the foregoing analyses in that performance on the test is not highly predictable from a knowledge of student background characteristics, pretest performance, or classroom interaction experiences. Moreover, there are real differences among classrooms in terms of Decoding 2 scores which cannot be accounted for by reliable differences in observed classroom interaction patterns. Figure 4.2 helps clarify what might be going on with Decoding 2 scores, which in part measures how well students can identify sounds and words in English. It seems that certain classrooms in which students have the poorest relative grasp of phoneme-grapheme correspondences in the fall are the very classrooms in which the students have achieved a relatively superior grasp of these correspondences by the end of the school year; and *vice versa*. One obvious way in which we can interpret these results is to hypothesize that teachers in some sense overreact to their students' initial abilities in English pronunciation: if these abilities are initially minimal, the improvement is sought; if these abilities are initially superior, then other aspects of performance are emphasized.

Differences in Achievement of Oral Proficiency

As for the Oral Proficiency Test structure scores, little is revealed by inspection of Figures 4.4 and 4.5 except that the achievement of level 1 students in classes F1a and F1b is lower than might be expected. Notice, however, that neither the John Test nor the Decoding 1 Test is an optimum "pretest" for oral proficiency.

Between Classrooms Relationships Among Student Background Characteristics, Achievement, and Teacher/Student Interaction Variables

A comparison of Table 4.1 with the corresponding values in Table 4.5 reveals substantial change in the predictive role played by age as well as history of employment in the country of origin. This suggests that classrooms are differentiated not only by level of proficiency, as measured by initial John Test scores, but by certain student background characteristics as well. In order to explore this possibility we turned to univariate analysis of variance and canonical discriminant function analysis.

Not too surprisingly, one highly significant discriminant axis emerged from the analysis, and it accounted for 59 percent of the between-classroom variation. Details of the analysis need not be presented here; suffice it to say that amount of English studied in the U. S., English studied in the previous country, length of time in the U. S., years of education, age (younger), and previous occupational level were weighted in the first discriminant function in the order listed. By taking into account the second canonical axis, the percentage of variance accounted for could be brought up to 73 percent, and the resulting solution is plotted in Figure 4.6. The second axis is of marginal statistical significance, it is a weighted linear combination of sex, English studied in the previous country, and national origin, in that order.

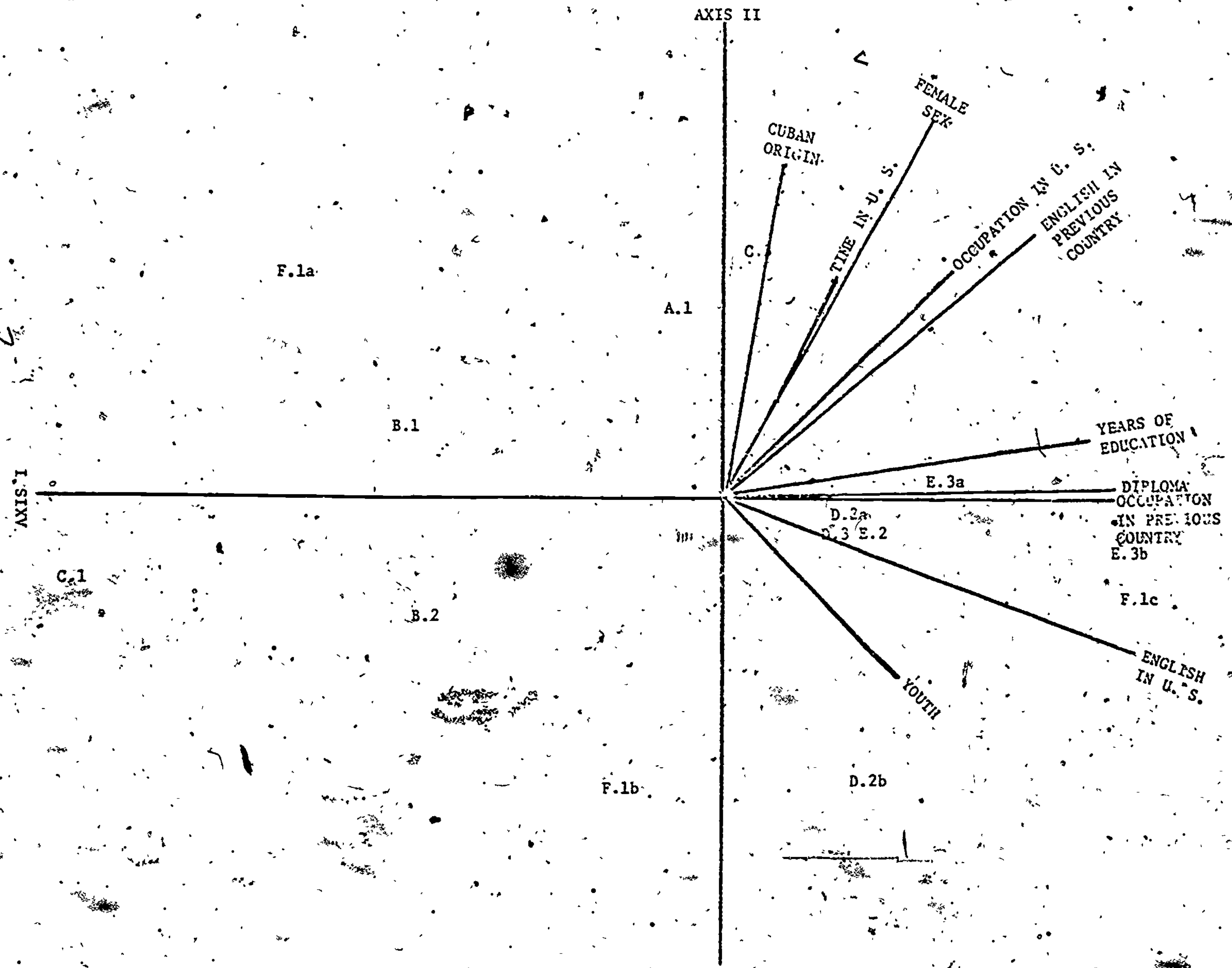

FIGURE 4.6

Plot of Day-School Classroom Centroids and Student Background Characteristics in the Space Depicted by the Two Largest Canonical Discriminant Axes

As for other figures relating to discriminant function analysis, we have represented the classroom centroids in Figure 4.6 with unique letter-number codes, while the student background characteristics have been represented by vectors. Inspection of the figure reveals wide variation among day-school classrooms in terms of student background characteristics. It can also be seen that much of the variation among classrooms is related to level differences as determined by John Test scores.

The background characteristic vectors are also plotted in Figure 4.6. There seem to be some rather suggestive clusters: youth and amount of English studied in the U. S.; educational and occupational level in the previous country; amount of English studied in the previous country and occupational level in this country; sex (female), origin (Cuban), and length of time in the United States. All in all, these results suggest the possibility that students have been assigned to more-or-less homogeneous classrooms on the basis of background characteristics in addition to their John Test scores.

It is enlightening to compare Figure 4.6 to the earlier figures in this chapter, in which pre- and posttest achievement are plotted without taking any notice of student background characteristics within the classrooms involved. This comparison makes the exceptional Literacy gains of students in classroom C1 seem all the more impressive, for instance (Figure 4.3).

When Figure 4.6 is compared to Figure 3.2--a similar analysis of classroom differences, but in terms of teacher and student interaction patterns--several fascinating points of agreement emerge. Notice that the two classrooms taught by teacher C are widely discrepant with regard to student background characteristics (Figure 4.6), and that these two classrooms are taught in markedly different ways by teacher C (Figure 3.2).

These data suggest that the manner in which a particular class is taught may reflect a special adaptation on the part of the teacher to the background characteristics and ability level of the students involved.

Although a bit more speculative than our reasoning with regard to teacher C's behavior, notice that the three level 1 classrooms taught by teacher F are spread widely with regard to background characteristics (Figure 4.6) and that their respective locations in Figure 3.2 suggest the application of a more directive teaching strategy in the less well prepared classrooms. As for teachers B and D, there seems to be a possible association between the amount of student-student feedback in Spanish which occurs in a classroom (Figure 3.2) and its composition in terms of country of origin, sex, and length of time in the U. S. (vertical direction in Figure 4.6); a similar inference can be drawn for teacher C.

In summary, we have accumulated considerable evidence suggestive of relationships between classroom interaction variables and student achievement variables. The regression analyses presented in earlier sections of this chapter were designed to reveal how classroom interaction experiences might come to affect student performance. Now we see, however, that student performance and background characteristics may well affect classroom interaction; that is, a model of reciprocal causation is probably most appropriate for these data.

The Structure of Between Classroom Variation in the Adult Learning Center

The canonical correlations reported above give evidence of rather strong relationships between student background/performance characteristics and teacher/student interaction factors at the classroom level. It is in order to ask whether we might be able to find these relationships back down at the level of individual background characteristics, test scores, and classroom interaction factors. The answer is yes, although we must keep in mind that the results are dependent upon only 14 classrooms studied in the Adult Learning Center and may not generalize widely. Only where a similar system of assignment of students to classrooms, a similar range in teaching styles, and a similar population of students can be assumed to exist will these results generalize to other ESL systems.

In this analysis all the information obtained was used in one analysis: students' background characteristics, their pretest scores, the factor scores for patterns of classroom interaction, and posttest scores. The purpose of this analysis is to see if there were distinctive patterns of these domains of variables. The results are described by four factors in terms of which the classes may be differentiated by background, classroom interaction patterns, pretest performance and final achievement.

We found a tendency for superior Adult Learning Center classrooms to contain students who have a history of higher than average English Study in the U. S., Level of Former Job, and Educational Level; while having lower than average Age and Level of Job in the U. S. We found an association of these characteristics with superior performance on all initial achievement measures.

These students were taught with the "free response" mode of interaction, as might seem appropriate for students of higher than average initial ability. "Student-student feedback" either did not occur or was not allowed; nor was the "model-practice-corrective feedback-model-practice" paradigm used. There is a tendency for students within classrooms with such a high level of prior training and initial ability not to be expected to read and/or ask questions under direction of the teacher; on the contrary, a (group?) question-answer interaction pattern goes along with the "free response" format of instruction.

But in these classes somewhat above average Oral Proficiency correctness is associated with below average Decoding 2 performance. Here we see again the phenomenon noted in connection with Figure 4.2; students in higher-level classrooms appear not to be exposed to instruction in phoneme-grapheme correspondence--they are consistently surpassed in these skills by students in the initially lower-level classes.

There was a tendency for some superior classrooms to contain students who have been in this country for a relatively long while and who are established in higher level jobs. They also have a history of English study in the former country which supports our earlier conjecture that prior English language training may be instrumental in the acquisition of higher level employment in this country. These classes have exceptional achievement of literacy, along with an above-average achievement of proficiency.

One classroom interaction pattern characteristic of these classes is the "direct-read and/or ask questions" interaction factor which is highly associated with posttest literacy. Other characteristic classroom interaction patterns are "Other" and "Free," along with the "question-answer-corrective feedback-prompt-answer" sequence.

A third group of superior classrooms contains well employed females who have been relatively well educated. Individuals in these classrooms, who may well have ~~arrived from~~ Cuba some time ago (relatively speaking), have had superior English language training in their country of origin, but not necessarily in the United States.

Above average performance on the Morano pretest characterizes this group. The Morano score is an indicator of grammatical skill. The Oral Proficiency correctness score also has its highest loading on this dimension as do the posttest decoding skills. This substantial achievement could well be attributed to the high current as well as prior level of employment, former ESL training, educational level, or sex (female) of the individuals involved.

Two interaction patterns predominate in these classes: the "Other" factor and the "teacher-direct-student read and/or ask question" factor. Other classroom interaction patterns are deemphasized, including "model-practice" as well as "question-answer-corrective feedback-prompt-answer."

The fourth factor is strongly related to all but one of the original nine classroom interaction factors. The pattern of loadings suggests the familiar contrast between classroom interaction factors seen along the horizontal axis of Figure 3.2: QA-IND, CF-P-A, SFBK, FREE VS. MOD-PRAC, CF-M-P, DIR READ, and OTHER. At earlier points in our discussion we have referred to the positive (left in Figure 3.2) pole as flexible, free, supportive, eclectic, individualized while referring to the negative pole as structured, class or group oriented.

Those classrooms in which a more supportive and individualized pattern of teacher-student interaction prevails tend to contain students who are recent arrivals from Western Europe, perhaps not too highly educated, a bit younger than average, and male. This means, of course, that the more highly structured classrooms (e.g., those of teacher F in Figure 3.2) tend to contain older females of Cuban or Carribean origin who have been in the United States for some time.

As for test performance, it is clear that the recent arrivals from Western Europe are very deficient in Decoding 2 skills (phoneme-grapheme correspondences) in the fall but become exceptionally proficient by the end of the school year.

What these four factors of between-classrooms variation reveal is the complexity and richness of the ESL training in terms of student background characteristics, initial test performance, student teacher interaction patterns, and final achievement. The Adult Learning Center must be viewed as a dynamic system in which students are channeled into classrooms which promise to provide them with an optimal learning experience. There are at least four ways in which classrooms can come to be above average in the achievement of English language proficiency at the end of the period of instruction. It is clear that initial proficiency as determined by the John Test is a good across-the-board guarantee of final proficiency, not surprisingly; but, among other aspects of pretest performance, background characteristics and classroom interaction patterns must be taken into account in order to characterize fully the variety of patterns of achievement seen in different classrooms.

CONCLUSIONS AND RECOMMENDATIONS

The single most important analysis in this study revealed the interactions among students' characteristics, their initial proficiency, classroom interaction patterns, and final achievement. Some classes performed better than others. These classes fell into three distinct groups in which different methods of instruction were used.

Before we comment on these arrangements, it is important to point out that we are discussing differential or <u>relative</u> achievement and not <u>absolute</u> achievement. All of the classes made gains in proficiency. Some, however, made greater gains than would have been predicted from their initial scores. We were able to differentiate among these classes in terms of the students' characteristics, their initial proficiency and the classroom patterns of interaction.

One type of class of above average achievement was typically composed of younger students, who had attained a higher educational level, had studied more English in the United States, and who in their native country had held higher status jobs. This type of class had higher than average Oral Proficiency correctness scores, but lower scores on the measure of phonic skills (Decoding 2). They were largely taught in the "free response" mode which meant that the students were encouraged to generate English statements. The interaction pattern is characterized by the teacher asking a question and the student answering or also asking questions.

A second type of class was composed of students who had been in this country for a longer time, who had studied more English in their native country, and held higher level jobs in this country. These students achieved better than average proficiency and performed exceptionally well on the literacy measure. They were taught largely in the "direct-read and/or ask questions" pattern in which the teacher directs the students to read something and ask questions about it or the students ask questions. The teacher works from a set of materials that form the basis for asking questions. The teachers of these classes also used the "Other" and "Free response" and the "Question-answer-corrective feedback-prompt-answer" modes. These variations reflect the use of different strategies within a context of talking about materials.

A third group was composed of females who had higher status jobs, who were relatively better educated, were from Cuba, had been here longer, and who had studied more English in their native country. Their performance was superior on all measures that required correct usage--the Oral Proficiency correctness scale, the Morano, and the Decoding tests. They were taught largely in the "Other" mode and the "Teacher direct-student read and/or ask question" modes.

These differences suggest a hypothesis. Assume that some teachers had chosen the appropriate methods for the type of student. Then the hypothesis is that proficiency is increased to the degree that the appropriate method is chosen for the type of student. While this conclusion is hardly startling, guides for practical action are apparent in the data. The three types described above comprise the majority of students in the Adult Learning Center. Presumably the easiest way to adapt methods to types of students is to organize classes in terms of the students' characteristics and to have teachers use the methods appropriate to the type of class.

The Center presently places students in classes on the basis of their level of proficiency as measured by the John Test. <u>It is important to note that it is not the initial level of proficiency that alone determines how much additional proficiency is achieved.</u> If appropriate methods are used in relation to the students' characteristics, greater than expected proficiency is achieved. We recommend, therefore:

1. That students be placed in classes in terms of their initial proficiency and that they also be grouped within levels of proficiency, as much as possible, by common background characteristics, and in clusters like those described above.

2. That teachers be assigned to these classes whose teaching style is appropriate for the type of student.

This second point needs some additional explanation. The variations on the audiolingual method have limited effectiveness for the types of students attending the Center. In some of the first analyses the interactions characteristic of this method were associated with greater achievement. But, when we analyzed the interaction patterns in conjunction with the students' characteristics, it appeared that interaction patterns requiring more free responding were more effective, particularly with students who had studied some English (and again irrespective of the actual level of proficiency).

The reader has undoubtedly noticed the frequency with which the "Other" interaction pattern appeared significant in these analyses. When we examined the variety of specific instances in this category, many of them seemed to require the student to generate language, to think about the language (for example, sorting words into their grammatical categories), and to use the language for reading and discussion. This category and "free response" appear frequently as significant interaction patterns particularly in conjunction with students' characteristics. The activities in these categories seem to have in common increasing the frequency with which students use the language, but use it not by imitating it but by generating it.

A practical plan might be to use the variants on the audiolingual method with students who have the least proficiency, but even then to mix it with the other interaction patterns as quickly as possible. Perhaps the next type of interaction pattern to use after some minimal proficiency has been acquired is the forms of the "Teacher-direct-student read and/or ask questions." This type of procedure seems to be a way of helping students to use the language naturally. But the direction of using these procedures should be to bring the students to free responding.

It is clear that no one method or teaching performance is uniformly effective. One cannot really take sides in an audiolingual vs. "silent way" debate on methodology since the effective patterns of teacher behavior cut across the elements of both. This conclusion is particularly important because the teachers tend to use a consistent style. The effects of this consistency were apparent in the analysis of different classes of the same teacher. One class of a teacher achieved better than predicted; another, taught in the same way, did less well than predicted. Only when we analyzed the interaction patterns with the student characteristics data did we find that methods probably must be adapted to specific characteristics of the students.

The regression analyses indicated that students' characteristics were a major predictor of their subsequent achievement. This factor cannot be ignored if one wishes to make the system of instruction even more effective. We are not implying that the Center's teachers do not adapt to the type of class; some do, some have selected the appropriate method for the class which they are teaching. But obviously there were classes for which the instructional methods were inappropriate or less effective. Given that the Center's teachers work closely with their students, it seems likely that giving the teachers more information about their students (such as that provided by the questionnaire developed for this study), and learning which procedures are more effective with certain types of students may be sufficient to increase the effectiveness of the system markedly.

The reader should recall that this study was not an evaluation of the Center. It was an intensive, in-depth study of the students and teachers of the Center to find out how best to organize instruction. The results have provided some hypotheses on which plans of action may be built.