DOCUMENT RESUME

ED 146 201 TM 006 581

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. Technical Summary.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	New Jersey State Dept. of Education, Trenton.
REPORT NO	ETS-PR-77-5
PUB DATE	Jun 77
NOTE	73p.; For related documents, see TM 006 580, 582, and 583 ; Tables are marginally legible due to small type
EDRS PRICE	MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS	Academic Achievement; Adult Basic Education; *Adults; Classrooms; *English (Second Language); Interaction; Language Instruction; *Language Proficiency; Language Tests; Oral English; Performance Factors; *Predictor Variables; Spanish Speaking; Statistical Analysis; *Student Characteristics; Student Teacher Relationship; Teacher Behavior; Teacher Characteristics; Teacher Qualifications; Teaching Skills; Teaching Styles; *Teaching Techniques

ABSTRACT

This volume provides a technical summary of research conducted at the West New York (New Jersey) Adult Learning Center with adults in English as a second language classes. The study was conducted in order to determine the relation of teaching strategies and performances to the acquisition of skills in spoken English. Student proficiency was measured at two points in time, and the intervening instruction was measured daily. 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. Classroom observations provided a continuous record of both teacher and student behavior during class sessions and between the two tests assessing the students' speaking proficiency. 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. This study revealed the interactions among student characteristics, initial proficiency, classroom interaction patterns, and final achievement. The classes with higher achievement fell into three distinct groups which used different methods of instruction, and whose students shared similar background characteristics. (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

TECHNICAL SUMMARY

By

Frederick J. McDonald
Meredith K. Stone
Allen Yates

June 1977

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

TECHNICAL SUMMARY

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.

This volume is a summary of the Final Report of the research conducted at the West New York Adult Learning Center, and appears in two volumes. Because the Final Report presents detailed and complex statistical analyses, we have prepared this executive summary which describes the purpose, methods and results of the study with the basic statistical data. Readers interested in the details of the statistical analyses may send for the full report.

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

INTRODUCTION

The West New York Adult Learning Center provides a training program in which adults learn to speak English as a second language. This report describes the research conducted during the 1975-76 school year in the West New York Adult Learning Center to determine the relation of teaching strategies, methodologies, and performances to the acquisition of facility in speaking English as a second language. 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.

The general methodology used in the study 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.

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.

THE MEANING OF PROFICIENCY IN SPEAKING ENGLISH

The operational definition of proficiency used in this study was derived by an analysis of the goals of the West New York 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.

Three levels of performance in speaking English 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 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.

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. We assigned observers to make daily observations of each of the teachers in the classes of the Adult Learning Center.

The method of observation 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. 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 because behavior relevant to them had been observed in the classes of the Center. These two methods are the audiolingual and the "silent way."

Since these two methods are used by different teachers in the Center, the category system 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 do they 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?

THE SAMPLE OF CLASSES STUDIED[1]

The students in the Adult Learning Center ranged 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.

---

[1] Both day and evening classes were studied; however, the most informative results came from the data collected on the day-school classes and therefore it is this sample that is described here. For a complete report, contact the senior author.

There were 14 day-school classes taught by six different teachers which were studied in this research project. The majority of classes in the day school of the Adult Center met for an hour-and-a-half a day, five days a week.

Table 1 presents the data on students who were in the Center during the entire period of the study 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.

As can be seen by scanning Table 1, there were more women among 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. The occupations held by the students in their country of origin 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.

Teacher Characteristics

The 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 (to night-school adults); number of years teaching ESL full-time to children, to teens, to adults;

TABLE 1

Day School Student Background Information -- Descriptive Statistics

	Total Freq.	Total Percent	Freq. by Level[2] 1	2	3
Number	81		24	28	29
Sex: M	26	32	9	13	4
F	55	68	15	15	25
Diploma: Yes	32	40	8	10	14
No	49	60	16	18	15
Age: Mean	46		48	45	45
Range	19-70		21-69	19-70	24-67
Education:					
Mean Years	10		9	10	11
Range	4-16		4-16	4-16	4-16
Time in U. S.					
Mean Years	6		6	6	7
Range	1-24		1-15	1-17	1-24
Years of English Studied in Former Country					
Mean Years	1.17		.83	.75	1.83
Range	0-12		0-5	0-2	0-12
Years of English Studied in U. S.					
Mean Years	1.17		.58	1.25	1.50
Range	0-3.0		0-1.5	0-2.5	0-3.0
Country of Origin					
Columbia	3	4	1	2	0
Cuba	73	90	22	25	26
Dominican R.	1	1	0	0	1
Ecuador	1	1	1	0	
Peru	0	0	0	0	0
Other	3	4	0	1	2

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

*Previous country/USA; e.g., there were 23 students who were professionals in their country of origin; but only one was a professional at the time of this study.

[2]"Level": students are assigned to three instructional levels on the basis of their proficiency in speaking English when they enter the Center.

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.

The majority of teahers were women. All six teachers had completed some graduate work, three had completed a Master's degree and one was working on a second degree. Most of the teachers' experience with English as a Second Language (ESL) had been teaching adults.

Only two teachers said they used the audiolingual method of teaching and estimated that 60-65 percent of their teaching was of this kind. The majority of the teachers 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: in November and again during April and May. The first test measured how proficient the students were near the beginning of the instruction that they were receiving that year; the second test measured how proficient they had become after six months of instruction.

The 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 a composite list of 132 objectives resulted. These 132 objectives were then given to the 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 rating were then used to develop the content of the Oral Proficiency Test.

The Oral Proficiency Test as an individually administered 60-item 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.

One score on this test 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.

The Literacy Test

A potential side effect of learning to speak English is that students may also learn to read it. 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, lables 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.[3]

Aural Decoding Test

We thought that students might also acquire decoding skills as a result of learning to speak English. A test developed originally to measure the decoding skills of young children[4] was adapted and used to see if the students were also acquiring decoding skills.

Other Measures of Proficiency

Two other measures of proficiency were used, 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.

---

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

[4] 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.

John Test

This oral proficiency test (developed by Linda Lunz 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.

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.

Student Performance on the Measures of Proficiency

Table 2 presents the data on these measures. Information is given for each test 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.

Description of the Observation System

A categorical observation system was developed which allowed for sequential coding of classroom behavior. A copy of the Observation Coding Sheet is shown in Figure 1.

TABLE 2

STUDENT TESTS: DESCRIPTIVE STATISTICS

Instrument	When	N	$\bar{X}$	Range	SD	Reliability
Literacy	November (Pre)	148	29.82	3-49	12.52	.96
	April (Post)	119	37.91	11 49	7.22	.89
Aural Decoding I	November (Pre)	148	26.72	6-38	7.40	.88
	April (Post)	120	28.42	13-37	5.33	.81
II	November (Pre)	148	47.84	31-57	5.28	.74
	April (Post)	120	52.45	36-59	4.63	.77
Total	November (Pre)	148	74.57	49-93	11.03	.89
	April (Post)	120	80.88	56-95	7.90	.83
Proficiency Comprehension	June (Post)	113	30.48	2-58	14.66	.96
Correctness	June (Post)	113	11.83	0-45	8.84	.92
Structure	June (Post)	113	14.53	0-41	10.16	.93
Prompts	June (Post)	113	9.31	0-26	4.86	.72
John	November (Pre)	115	37.27	0-70	20.54	.87
Morano	November (Pre)	118	26.82	3-50	11.47	.93

FIGURE 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	= CLASS/GROUP/IND.	12 ANSWERS	13 FREE RESPONSE	14 PRACTICE	15 WRITES ON BOARD	16 READS	17 CHOOSES NOT TO RESP.	18 ASKS QUESTIONS	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
1	1	1	✓							✓	✓										✓						N								✓										
2																										N								✓										
3																									✓					✓						✓									
4							✓			1			✓																								✓									
5																																														

- 13 -

18

19

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

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.

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. 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. 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.

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.

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

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. Observation of the classes began January 19th and continued through April 1, 1976.

Table 3 gives the observation item numbers, category labels, codes and means for the 14 day-school classes observed. The category means are the proportion of observation episodes[5] in which the event was observed in the classes. These will, however, add to 1.00 only when the behaviors within a group are mutually exclusive and exhaustive.

ANALYSIS OF THE OBSERVATIONS OF TEACHING BEHAVIOR

The system of recording classroom behaviors employed in this study yielded a 23-item response record (see Table 3) for each episode of interaction observed in the classroom. Each such episode began when either the teacher or a student initiated an interaction, 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

[5] 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

Item Categories for Classroom Observation

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

TABLE 3 (Continued)

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

+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.

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.

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. 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 would 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.

Our main interest in the analysis of the behavioral observation data was to find, if they occurred, fairly stable and distinct patterns of classroom intearction. We wanted to find out if certain teacher behaviors gave rise to or were associated with certain specific student behaviors.

Methods of Analyzing Classroom Interaction

The availability of data on many individual behavioral episodes made it possible to conduct analyses of a large number of variables, even though only six teachers and their students were studied. Several different methods were used to determine how many patterns of interaction could be found in the observational data. These patterns were analyzed by studying the association among the variables.

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 $r = .80$. A 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 (a statistical procedure) was used 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 ($r = .83$). Thus, six of the original variables are intercorrelated; but the factor analysis breaks these intercorrelations into two patterns or factors.

Factor analytic methods are a quantitative way of looking for these patterns. The methods may yield none or many patterns (factors). The number of factors produced depends on the number of distinct patterns of association in the data.

PATTERNS OF CLASSROOM INTERACTION

We found four patterns of classroom interaction which characterized the beginning of an episode. They were:

1. Teacher Model--Student Practice: The teacher illustrates verbally an English-language structure or pronounciation. The student in turn attempts to imitate the teacher.

   Teacher: "New Jersey."
   Student: "New Jersey."

2. Teacher Direct--Student Read: The teacher directs the student to read printed material containing the structure being learned. The student in turn reads the sentence containing the structure.

   Teacher: "George, would you read the first sentence?"
   George reads: "If I had time, I would go with you."

3. Teacher Direct--Student Read or Ask Question: This pattern is the same as the above pattern except that the student may ask a question about the material to be read.

   Teacher: "George would you read the first sentence?"
   George: "I fill in the space with the past tense?"

4. Teacher Question--Student Answer: The teacher asks a question; the student answers it.

   Teacher: "What is the short way of saying 'I would,' the contraction?"
   Student: "I'd."

These patterns of beginning an episode were followed by three distinctive patterns of continuing the interaction. They were:

5. Corrective Feedback--Model--Practice: The student has responded by reading (2 and 3 above), or by imitating (1 above) or by answering

a question (4 above). The teacher corrects the student's response and gives the correct form. The student tries the response again.

Teacher: "The short form of 'I will,' is 'I'll.' I'll go with you, if I have time."
Student: "I'll go with you, if I have time."

6. Corrective Feedback--Prompt--Student Answer: This pattern is very similar to the one above, except that the teacher does not model the appropriate response. Rather he or she encourages the student to try the response, prompting him or her in the process until the student provides the appropriate response.

Student: "I will go to the store tomorrow."
Teacher: "I ___?" (indicates short form with fingers)
Student: "I'll go to the store tomorrow."

7. Teacher Question--Student Answer: (This pattern appears both as initiating and in continuing an episode.) One of the initiating patterns has begun the episode and the student has read, asked, or answered a question. The teacher then asks a question which the student answers.

Teacher: "Can you put it in the past tense?"
Student: "I went to the store yesterday."

A set of these initiating and subsequent patterns might be one of these:

(1) Teacher Direct--Student Read ⟶ Teacher Question--Student Answer:

Teacher: "George, would you read the first sentence?"
George reads: "If I had time, I would go with you."
Teacher: "Can you use the short form of I would?"
George: "If I had time, I'd go with you."

(2) Teacher Model--Student Practice ⟶ Corrective Feedback--Model--Practice:

Teacher: "I'll go tomorrow."
Student: "I'...ll go tomorrow."
Teacher: "O.K. But slide it together more. "I'll go tomorrow."
Student: "I'll go tomorrow."

(3) Teacher Question--Student Answer ⟶ Corrective Feedback--Prompt--Student answer:

Teacher: "How would you make it negative?"
Student: "I will not go tomorrow."
Teacher: "O.K. But use the short form. I wo..."
Student: "I won't go tomorrow."

Three other patterns were found which sometimes occurred at the beginning of an episode and sometimes after an episode had been started in one of the ways described above. These were:

8. Free Response: The teacher indicates the structure, but not the ideas or vocabulary.

Teacher: "If I were _____, I could _____."
Student: "If I were taller, I could play basketball."

9. Student-Student Feedback: The students prompted each other or one or more students repeated what another student has said. This interaction usually occurred in Spanish.

10. Other: This category includes a variety of teaching activities such as games; students adding to lists of adjectives, nouns, and verbs on board, then making sentences from lists; a student reading for several minutes; small groups discussing a reading assignment or cartoon; students making a list of everything one has to know to go to the gas station or grocery store.

Classroom Differences in Interaction Patterns

Once the analyses of the patterns of behavioral observations were completed, we investigated the differences in classroom behavior patterns--the aim being to identify interesting contrasts among classrooms which might account for differences in student achievement. We studied the individual classroom at this stage because we were interested in differences among the teachers as well as differences in the ways in which any given teacher might approach students of various proficiency levels. We contrasted overall differences between classrooms on each pattern of interaction with the amount of day-to-day variation observed within classrooms for that pattern of interaction. This analysis indicated that teachers differed more from each other than each did in their day-to-day teaching.

A second analysis was performed to find the bases on which the classes were most sharply discriminated from each other on the average. The results of this type of analysis (canonical discriminant function analysis) are portrayed in terms of axes, with each class having a score with respect to each axis.

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

In Figure 2 each classroom is identified by a letter-number combination. The initial letters range from A through F and identify the six instructors. The numbers range from 1 to 3 and refer to the proficiency level of the class being taught as measured by the John Test. The lower case letters identify

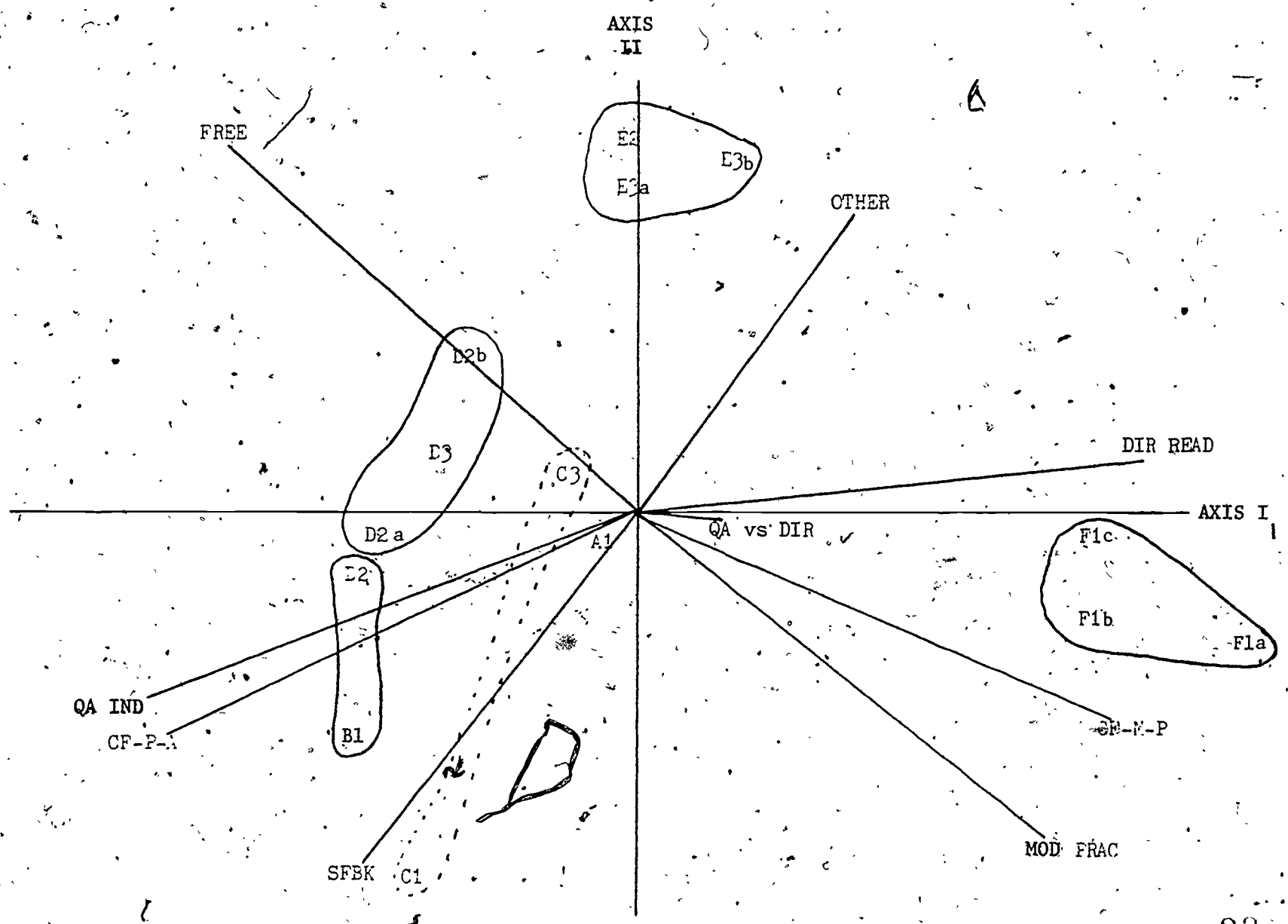

Figure 2

The Plot of Differences Between Classrooms in Interaction Patterns

different classes at any given level where they are taught by the same teacher. (We have drawn vectors in Figure 2 to represent how the nine original patterns of classroom interaction project into this space.)

The most outstanding feature of Figure 2 is the obvious clustering of classrooms taught by the same teacher. 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.

The differences between the classrooms as portrayed in Figure 2 may be understood by looking at the patterns which characterize each quadrant. The upper right-hand quadrant is characterized mainly by the pattern, "Other," and somewhat by the pattern, "Teacher Direct--Student Read" (DIR READ). One of Teacher E's classes falls in this quadrant.

The upper left-hand quadrant is characterized by "Free Response" (FREE). Teachers D and E appear in this quadrant.

The lower right-hand quadrant is characterized by the two variations around "Modeling": "Corrective Feedback--Model--Practice" (CF-M-P) and "Teacher Model--Practice" (MOD PRAC). One teacher, F, appears in this quadrant.

The lower left-hand quadrant includes the patterns of Corrective Feedback--Prompt--Answer" (CF-P-A) and the "Teacher Question--Student Answer" (QA IND) pattern. Three teachers appear in this quadrant.

This figure illustrates the characteristic style of each teacher. The patterns below Axis I require the student to imitate or practice those above the axis require the student to generate responses within a structure.

Consistency in style is evident in those teachers who taught more than one class. Teacher F uses essentially the same style with all three of his or her Level 1 classes. Teacher E, with a different style, is as consistent as Teacher F. Although Teachers B and D vary their style by class, the differences are variations on the same basic style. Only Teacher C shows remarkably different styles by level; however, C is the only teacher who had classes both at Level 1 and Level 3.

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. Since these nine patterns were intercorrelated, a second analysis was conducted. The nine patterns could be located with respect to two axes. Each class could be described in terms of its location with respect to these axes and with respect to the original nine patterns.

The first axis contrasts two teaching 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 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 is 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 have scores that place them relative to either 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 his or her overall style. 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.

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. These paradigms do not occur in pure form in this sample of teachers and students.

The Reliability of Differences in the Patterns of Classroom Interactions

Figure 2 illustrates the differences among the teachers in the kinds of interaction patterns that were observed in their classes. But, how reliable are these differences--are they "true" differences? Or might the observed differences be due to who observed a teacher, or on what day the teachers were observed?

Table 4 presents the results of an analysis performed to answer these questions. The statistical method which produced these results is a way of estimating how much of the differences among the teachers is due to sources which might have caused the observed differences. Down the left-hand side are the potential sources of the differences.

TABLE 4

Percentage of Differences in Observed Classroom Interaction Patterns Attributable to Potential Sources of These Differences

Source	QA IND	CF-P-A	OTHER	DIR READ	[illegible]EE	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	6
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

* $p < .05$

** $p < .01$

The 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 observed differences attributable to a source of variance. Consider the pattern, "Question-Answer-Individual" (QA IND). This pattern 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 related 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 sessions, and 16 percent is attributable to teachers. Therefore, 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 4 it can be seen that from 53 to 90 percent of the total variance of each classroom interaction pattern can be attributed to variation from interaction episode to interaction episode during the same session (day) of observation. In Table 4 we have indicated the probability levels associated with each source.

From these results many aspects of variation in classroom interaction patterns can be understood. It is clear, for instance, that only a very small and generally nonsignificant contribution to the variation in interaction patterns can be attributed to differential treatment of separate classrooms by the same teacher (see line labeled Classroom in Table 4).

The relatively large components of variance for Sessions suggest that a number of these fluctuate substantially from day-to-day. 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). These results suggest that a 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 patterns. This result indicates that each session of observation provides a rather narrow view of what is going on in the classroom.

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 4). 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 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.

The reader will recall that the original patterns were analyzed a second time regrouping them in terms of higher-order axes (represented mathematically by discriminant functions). Each teacher was given scores, called contrast scores, with respect to these axes. The two most useful axes are portrayed in Figure 2, although four axes could have been used.

The question we ask is, how reliable are these second set of scores? The upper portion of Table 5 is set up like Table 4, but the numbers of the axes are given instead of the labels for the original patterns. (Only Axes I and II are represented in Figure 2.)

The upper section of Table 5 gives percentages of variance for the contrast scores. 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, the large components of variance attributable to teachers. Teacher differences account for approximatley 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, by a substantial amount, our uncertainty about which pole of either contrast might appear in any given episode of classroom interaction merely by knowing which teacher is teaching. If the teacher is high on the first contrast (Teacher B in Figure 2), we would do well to predict "question-answer-corrective feedback-prompt-answer" interaction episodes. On the other hand,

TABLE 5

Percentage of Differences in Transformed Pattern Scores Attributable to Sources of These 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

Percentages of True Score Differences in Teachers' Mean Contrast Scores Attributable to Teachers, Classrooms, Sessions, and Episodes

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%

* $p < .05$

** $p < .01$

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.

Even the last two axes ("student-student feedback" and "direct read- and/or ask question vs. question-answer," respectively) show larger teacher and classroom components of variance than do the original patterns 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.

Each teacher has a score with respect to each axis. These scores were averaged across all the days of observation. The resulting mean score for each teacher was used in analyses of the relation of these scores to students' scores. Thus the mean contrast scores represent how the instruction was organized and conducted in each teacher's class. How reliable are these mean scores? Do they represent "true" differences among the teachers?

The lower section of Table 5 gives a breakdown of the weighted contribution of each potential source of differences to observed variation in teacher means for the contrast scores, that is, the scores with respect to the axes. 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.

This analysis shows that we are dealing with highly reliable indices of teacher variation; the reliabilities of teachers' means range from .80 to

.98. 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 related to differences in the ability levels of the classes involved. Whereas our initial interpretation of Figure 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 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 were 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.

In Table 6 we have summarized the results of this comparison for each of the three pairs of observers, using as the basic units of analysis the session for the nine original classroom interaction factors. Note that we are discussing observer assessment of session-to-session variation in these interaction patterns, but we have already seen from Table 5 that enough

TABLE 6

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

sessions of observation were obtained in the 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 5).

From Table 6 it can be seen that session-to-session correlations among scores concurrently by different observers range from .80 to .99 with few exceptions. The student-to-student feedback correlations suggest that certain observers may have had difficulty either recognizing or recording "student-student feedback in Spanish." The lowest correlations involve Observer 2; this observer, however, made very few observations.

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 were 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 posttest 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 relations 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 extracurricular 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.

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. Another purpose of relating classroom interaction to achievement is to identify any components of classroom interaction which may have an impact upon specific forms of achievement.

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. These variables take the form of contrasts among classrooms in terms of observed patterns of teacher-student interaction. 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

The background and achievement domains were clearly related at the level of individual pairs of variables, so it was in order to see how background information could be weighted to predict individual achievement scores. It was also important 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 are summarized in Table 7. In that table the labels across the tops of the columns designate the fall and spring achievement scores; note that logarithmic transformations or 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).

The line in Table 7 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) of the variance in the fall Decoding 2 scores; for 44 percent of the variance in the John Test scores and so on. The probability levels associated with these statistics are coded with asterisks.

In Table 7, under Posttests and under Log_{10} P-CT, note the two numbers with an asterisk, .23 and .24. These weights are statistically significant, and appear opposite Former Job (status level of job in country of origin) and Eng. in U. S. (amount of English studied in the United States). Thus, only two background characteristics by themselves predict the Correctness Score on the Oral Proficiency Test. All of the background characteristics predict only 20 percent of the variance in the Correctness scores. But these characteristics predict 33 percent of the Comprehension scores' variants (see R^2 under Log_{10} P-CP).

The three weights which are underlined in the P-CT, P-CP and P-ST columns of Table 7 are large enough to give some feeling for which background factors affect proficiency scores. They are age (which has a negative weight), status of job in former country, and amount of English taken in the U. S.

A reasonable hypothesis from these data, then, 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.

TABLE 7

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

It should be remembered that predictive background characteristics do not "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, give us some idea about why age is a negative predictor of proficiency. Older students are more likely to have had a low status job in this country, hence, may have had 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, etc.

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 pretest scores and background characteristics. The results of these analyses are presented in Table 8. This table is read in much the same way as Table 7. The predicted scores (column headings) 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 8 are listed the labels for the pretest scores. 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 7. The entries in the line labeled, "R^2 with Addition of Pretests," can be compared to the corresponding entries in the first line. For example, the R^2 for predicting transformed Oral Proficiency comprehension scores from background characteristics is only .33, but with the addition

TABLE 8

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-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***	17.85***

*p < .05

**p < .01

***p < .001

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 account for a substantial proportion of the variance in the comprehension scores.

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.

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

Table 9 presents information from the regression analyses 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 8.

The first line of Table 9 gives R^2 from the regression of posttest scores on background characteristics and pretest scores, as previously seen in Table 8. Next are four lines with Roman numerals I through IV; within each line can be found the standardized regression weight of the corresponding "higher order" bipolar contrasts described earlier.

The sixth line in Table 9 gives R^2 once the four higher-order patterns of classroom interaction (the contrasts, see Figure 2) 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

TABLE 9

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**

$^{+}p < .10$

$^{*}p \leq .05$

$^{**}p < .01$

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 9 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.

The most striking aspect of the pattern of standardized regression coefficients presented in Table 9 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 9 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 helpful 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.

Comparison of Pretest and Posttest Performance in Day-School Classrooms

In order to study the mean performance of individual classrooms we used analysis of covariance procedures in which each pretest was the sole covariate for its respective posttest. The only exception was for the Oral Proficiency Test, in which only the correctness score was investigated and both the Decoding 1 and the John Test were considered, individually, as covariates.

Figures 3.1 through 3.5 are plots of the 14 classroom means of 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 3.4 and fall Decoding 1 scores in Figure 3.5 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 Figure 2 (page 23).

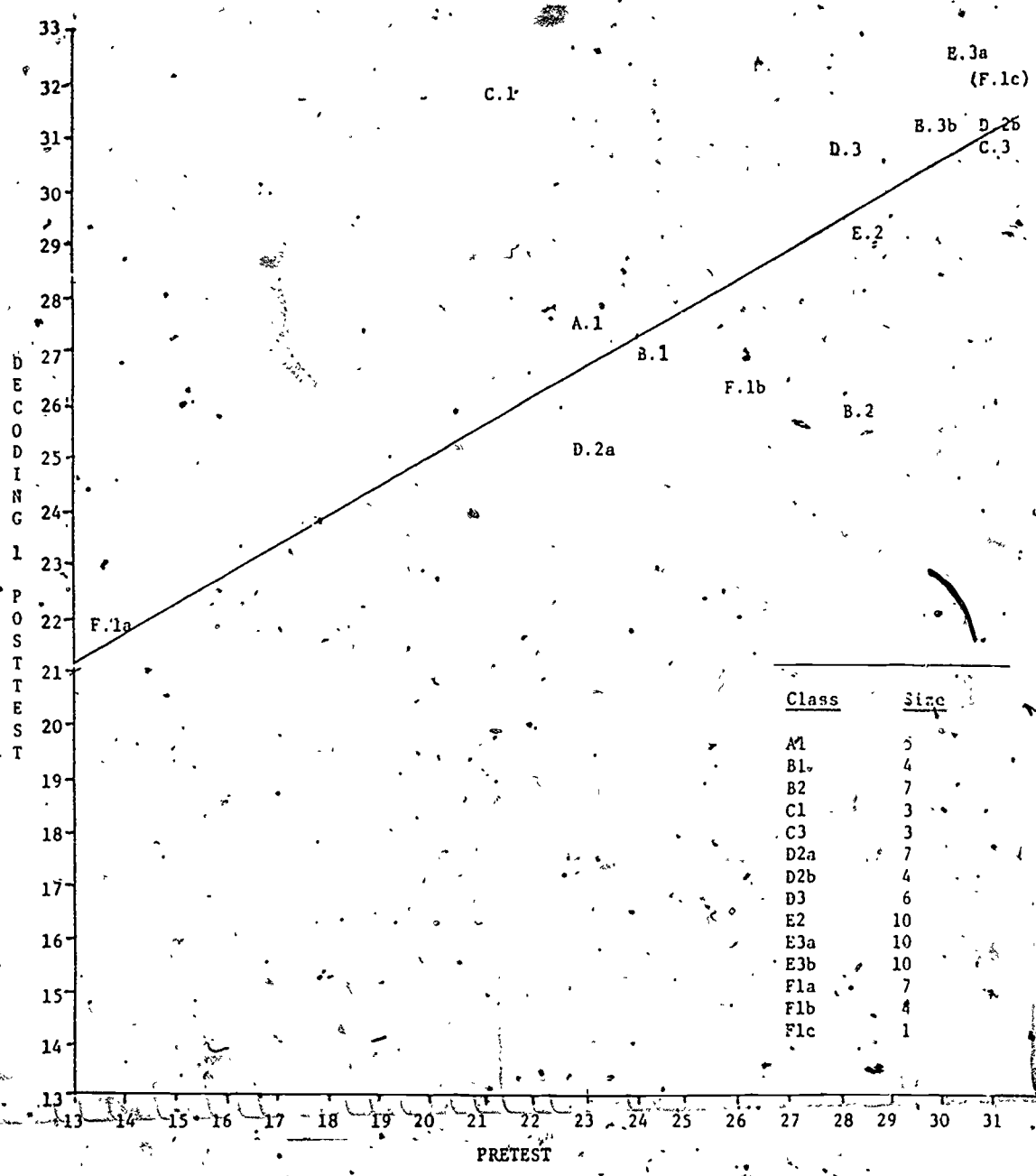

FIGURE 3.1

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

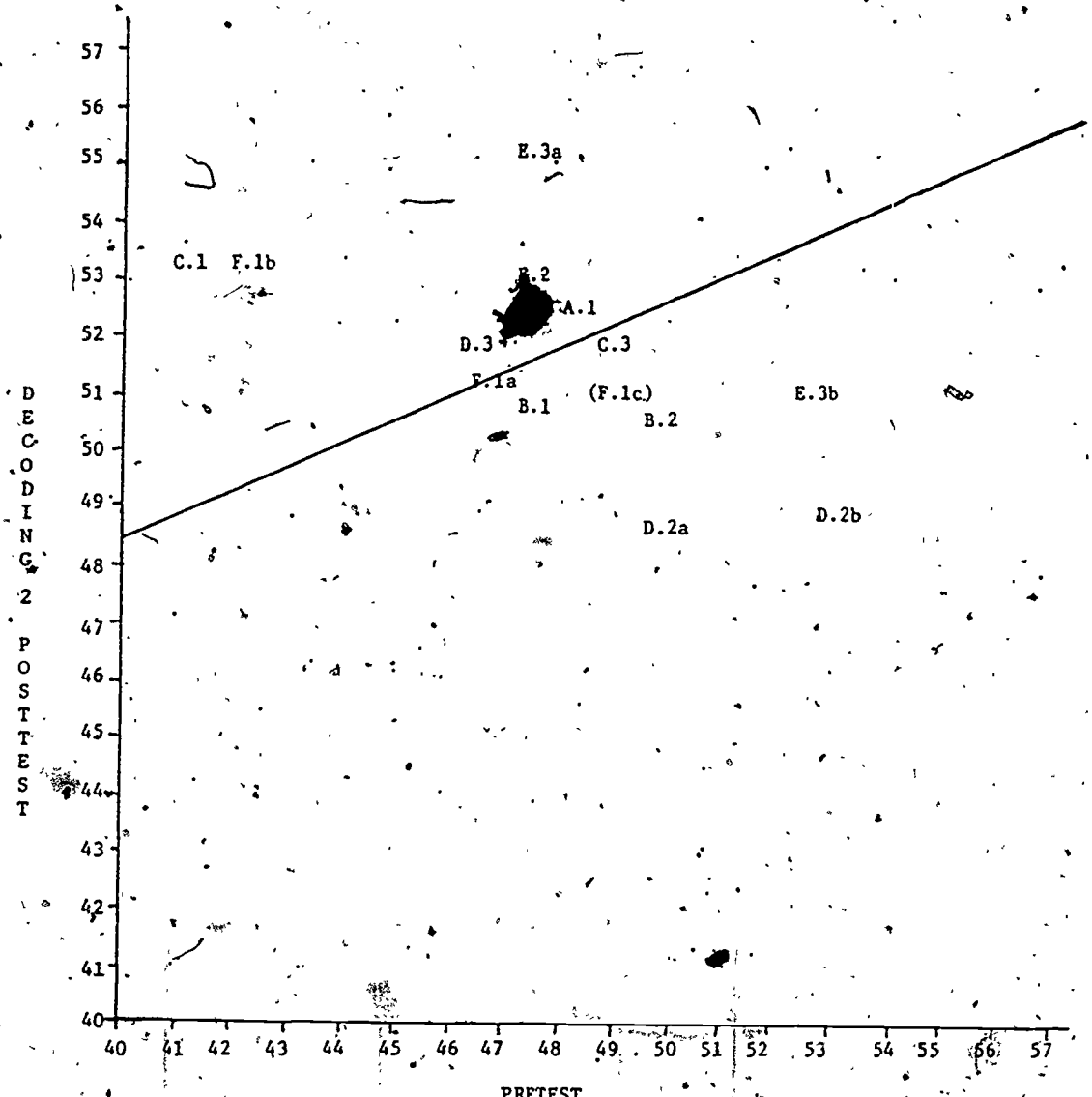

FIGURE 3.2

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

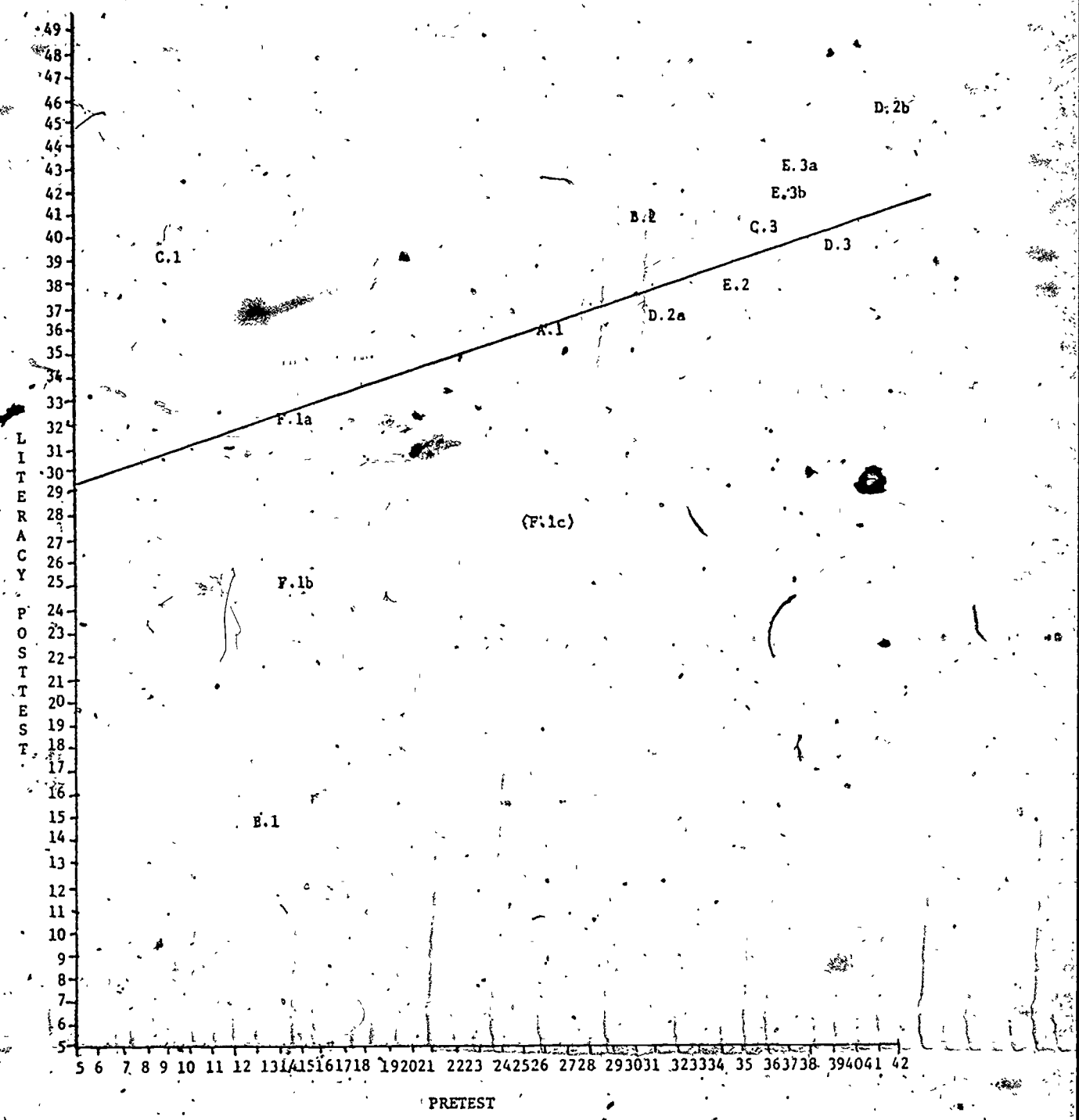

FIGURE 3.3

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

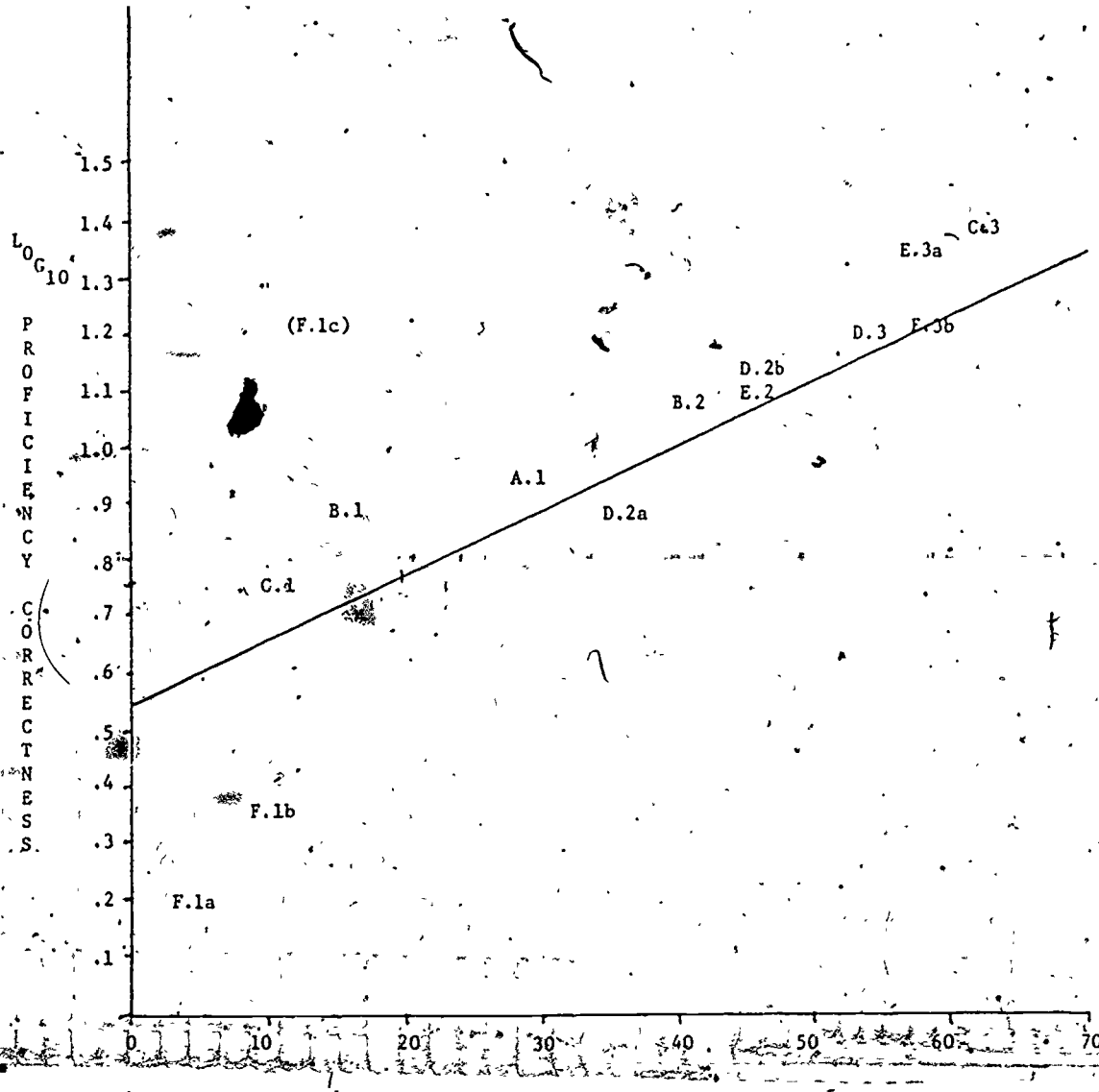

FIGURE 3.4

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)

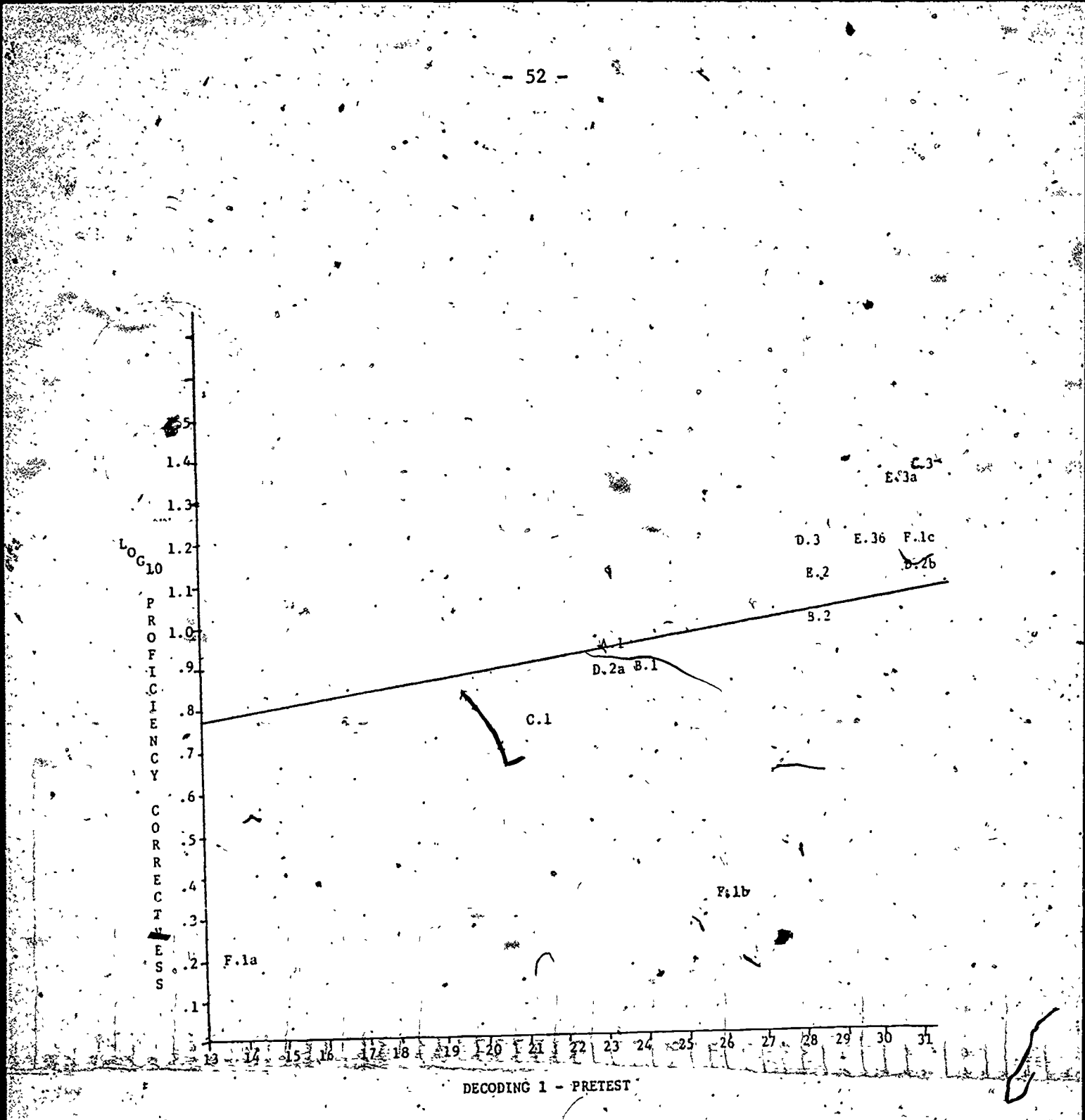

FIGURE 3.5

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)

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 3.4):

1. Along the horizontal axis are units of scores on the John Test. Assume that a class 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.
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 3.3 and 3.4, 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 3.4 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 3.4 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

The Literacy Test scores could not be accounted for solely by the four reliable classroom interaction contrasts. From Figure 3.3 we see that the variation in these scores is probably due to the exceptionally good posttest performance of the level 1 students in classroom C.1. 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 B.1, 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 3.1.

Differences in Achievement of Decoding Skills

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 3.2 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 relatively superior, then other aspects of performance are emphasized.

Differences in Achievement of Oral Proficiency

As for the Oral Proficiency Test correctness scores, little is revealed by inspection of Figures 3,4 and 3.5 except that the achievement of level 1 students in classes F.1a and F.1b is lower than might be expected. Notice, however, that neither the John Test nor the Decoding 1 Test is an optimum predictor for oral proficiency.

The Structure of Between Classroom Variation in the Adult Learning Center

There was evidence of rather strong relations between student background/performance characteristics and teacher/student interaction factors at the classroom level. We asked whether we might be able to find these relations at the level of individual students' background characteristics and test scores, and the classroom interaction factors. The answer is yes.

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 was 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 asked to read or encouraged to ask questions under direction of the teacher; on the contrary, the principal pattern in these classrooms is the question-answer interaction pattern and the "free response" format of instruction.

Some superior classrooms contained students who had been in this country for a relatively long while and who were established in higher level jobs. They also had a history of English study in the former country. This second

type of class had exceptional achievement of literacy, along with an above-average achievement of proficiency. One classroom interaction pattern characteristic of these classes was the "direct-read and/or ask questions" interaction pattern which is highly associated with posttest literacy. Other characteristic classroom interaction patterns were "Other" and "Free," along with the "question-answer-corrective feedback-prompt-answer" sequence.

A third group of superior classrooms contained well-employed women who had been relatively well educated. Individuals in these classrooms, who may have arrived from Cuba some time ago, 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 predominated in these classes: the "Other" pattern and the "teacher-direct-student read and/or ask quetion" pattern. Other classroom interaction patterns are deemphasized, including "model-practice" as well as "question-answer-corrective feedback-prompt-answer."

Those classroom in which a more supportive and individualized pattern of teacher-student interaction prevailed contained students who were 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 2) contained 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 were very deficient in Decoding 2 skills (phoneme-grapheme correspondences) in the fall but became exceptionally proficient by the end of the school year.

This analysis reveals the complexity and richness of the ESL training in terms of student background characteristics, initial test perormance, 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 relative achievement and not absolute 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 education 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 the questions and answers. 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. It is important to note that it is not the initial level of proficiency that alone determines how much additional proficiency is achieved. 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, 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 form 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 purpose 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.