DOCUMENT RESUME

ED 134 480 SO 009 522

AUTHOR Dugan, C. Lawrence
TITLE A Study of Attitude Change and Achievement in an Innovative Program at the Junior High School Level.
PUB DATE Jun 73
NOTE 53p.; Not available in hard copy due to poor legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Changing Attitudes; *Educational Innovation; Educational Research; Evaluation Criteria; *Evaluation Methods; Humanism; Humanistic Education; Junior High Schools; Measurement Instruments; *Pilot Projects; Program Administration; Program Descriptions; Program Effectiveness; *Program Evaluation; Schedule Modules; Secondary Education; Structural Analysis; Student Attitudes

ABSTRACT

The Small House program at Slauson Junior High School, Ann Arbor, Michigan, during the 1971-72 school year was evaluated. Two purposes of the evaluation study were to assess developments in the Small House program and to develop a prototype or model for evaluating innovative educational programs that delineates the evaluation, design, and analytic techniques. Small House was a pilot program that students could elect as an alternative to traditional education. It was designed to humanely alter the administrative and psychological environment of public junior high schools through flexible modular scheduling and team teaching. This study evaluated goals that Small House teachers set for themselves: normal student achievement in reading and mathematics, positive student attitudes toward school, and humane behavior in both students and teachers. A control group consisted of traditional mathematics classes at the junior high school. The California Reading Test, California Arithmetic Test, School Sentiment Index, and an attitude questionnaire were administered and student and teacher behaviors were observed. Evidence suggested that the program goals were achieved. However, variables such as administrative support, selection bias, team-cluster structure, and program newness confounded the evidence. Development of an evaluation model for innovative educational programs was not met due to methodological flaws and limitations in the study. However, a skeletal model was suggested that contains instructional, institutional, and behavioral dimensions. (ND)

A STUDY OF ATTITUDE CHANGE AND ACHIEVEMENT IN AN INNOVATIVE PROGRAM AT THE JUNIOR HIGH SCHOOL LEVEL

by

C. Lawrence Dugan

" . . . secondary schools tend to be even more authoritarian and repressive than elementary schools. . . . AND THE JUNIOR HIGH SCHOOL, BY ALMOST UNANIMOUS AGREEMENT, IS THE WASTELAND — ONE IS TEMPTED TO SAY CESSPOOL — OF AMERICAN EDUCATION (Silberman, Crisis in the Classroom)

June 1973

FOREWORD

Operation Small House was initiated in September, 1971 at Slauson Junior High School, Ann Arbor, Michigan. This experimental program was a direct result of the Humaneness in Education report submitted to the Ann Arbor Board of Education on January 13, 1971 and represented a concerted effort to alter the administrative and psychological habits of public junior high school education. Many of the goals and guidelines for Small House are set forth in the Humaneness Report.

This report constitutes an attempt to evaluate the Small House program conducted at Slauson between September 1971 and June 1972. Just as Small House was a pilot program, this was intended to be a pilot study with a two-fold purpose:

a. To assess developments in the on-going Small House program

b. To develop a prototype or model for evaluating educational programs, particularly innovative programs, which would delineate the essential components of evaluation, as well as design considerations and analytic techniques.

At the outset, it was hoped that we could develop, pre-test, and implement these assessment techniques all within one year. We either underestimated the complexity and enormity of the task confronting us or we overestimated our capabilities and resources, or both. In any event, while the results of this study might be interesting and in some cases, provocative, much of what was evolved remains to be finalized and validated in the future. However, I believe we were successful in developing the skeleton of a broad-based multi-method model for evaluation which would be applicable to other educational programs. This portotype is by no means considered a finished product, merely a guide for future research.

There are several other important considerations which should be mentioned at this time:

a. Parts of the research suffer from methodological flaws and limitations which effectively preclude reaching any definitive conclusions, promising though the results may seem. The nature of these limitations will be specified in the appropriate sections; I wish to stress that the results should be read with these limitations in mind.

b. There is another study of Small House being conducted by Marcia Radin and Mary Ann Morris at the University of Michigan which covers areas ignored or only touched on in this report. I feel

that both of these reports are critical to a full evaluation and understanding of Small House.

c. There is no adequate description of the experimental treatment (Small House program) available and I will not presume to present one here. There is a videotape made which attempts to summarize the program and which could be obtained by contacting Mr. Paul Siano at Slauson Junior High School. From the start, the Small House Team has stressed that this program was their own and, in that sense, unique. A brief sketch of Small House has been provided in the Introduction and attempts were made in other sections to identify some characteristics of the Small House program.

I owe a special thank you to Ken Heile who persevered with me from the very start of this undertaking to the bittersweet end and was essential to its completion. I also wish to acknowledge my sincerest appreciation to Anne Savishinsky for her help in shaping our observational schemes and to Professors Morse, Nadelman, and Dyer and the members of the Thursday morning seminar for their comments and suggestions. To the teachers and students at Slauson, I am particularly grateful; despite the constant demands made on them, they remained patient and cooperative throughout.

This report will be organized into seven major sections as follows:

Chapter I - Introduction - Includes a brief history and background of Small House, a very brief description of the program and goals that guided the research.

Chapter II - Psychometric Evaluation - Summarizes the results of the testing conducted. Each test is handled in a separate subsection.

Chapter III - Questionnaire - Describes the administration and results of a questionnaire specifically designed for the Small House evaluation.

Chapter IV - Systematic Observations - Summarizes the results of one month of observations at Slauson, in Small House and in the Control groups with respect to gross distinctions among teacher and student behaviors.

Chapter V - Nonsystematic Observations - Includes anecdotal materials from notes compiled over the year nonsystematically that bear on the evaluation.

Chapter VI - In Retrospect - Synthesizing the information and the development of the infrastructure of an evaluation model.

Chapter VII - Conclusions and comments.

CHAPTER I

INTRODUCTION

A. SMALL HOUSE: A BRIEF OVERVIEW

The impetus for the Small House concept stemmed from the Humaneness report submitted to the Ann Arbor Board of Education; at that time, or soon thereafter, seventh grade teachers in each junior high school were solicited for volunteers to begin a Small House the following September. Approximately 20-25 teachers from Slauson volunteered, from which four were selected — along with one counselor — to act as the Small House team. These five persons were given some planning time over the summer to develop a program consistent with the goals set forth in the Humaneness report; in addition, they established a set of goals for themselves. These individuals shaped the program that was to be implemented at Slauson in September.[1] There was no official Policy and Procedure statement nor any formal espoused philosophy for the Small House program, except for the general guidelines of the Humaneness report and the goals which the teachers set for themselves; these latter established the objectives of this report.

One of the direct results of this unspecified philosophy was that the details of the operation of Small House were worked out tentatively, and continued to evolve over the course of the year, within the framework provided. However, those directly concerned with the program made it perfectly clear that Small House was not intended to be an Open Classroom or a Free School. (As will be seen later, that distinction, like the distinction between graded and nongraded classes, can be a nebulous one.) This tentativeness (highly reminiscent of the espoused philosophy of Frazier in Walden Two) makes any but the most superficial description of Small House impossible; it does, however, reflect a great deal about the attitudes of those directly concerned with the operation of the program.

At the core of the program developed for Small House were two essential concepts: flexible modular scheduling and team planning. Instead of four 45-minute periods with five minutes between classes, Small House operated on ten 20-minute modules with no time between classes. (This was made possibly because of the physical proximity of the rooms to one another.) Of the 50 modules each week, approximately 50% were required, with the other 50% being spent in electives.[2] (Sample schedules are at Appendix I and sample elective sheets at Appendix II) Thus, in effect, students in Small House were only spending 50% as much time in traditional classroom settings as their seventh grade counterparts who were not in Small House.

---

1 The report of Morris and Radin mentioned in the Foreword would probably include greater detail of event during the developmental stages.

2 In general, the electives were changed every week or two, although some were retained throughout the year.

The flexible modular scheduling was the structural component that made it possible for the Small House to accomplish some of their goals including increased parental and community involvement, planned sessions for student feedback, and a broadening of the curriculum — it is estimated that some 50 different electives were offered over the course of the year. The Small House team also developed a new approach towards student evaluation (used in parallel with the traditional method since both students and parents in some cases expressed a desire for traditional grades) and, in some classes, implemented a different philosophy of testing, one which was essentially criterion-referenced.

Just as flexible modular scheduling was the skeleton on which Small House was built, the "planning team" provided the heart and brains of the system; the instructors used the planning sessions to discuss issues, develop electives, handle administrative details, and assist each other in professional growth. The teachers themselves stressed the contribution of the planning team to Small House successes.[3]

The students selected to participate in Small House represent a quasi-random sample of all those students whose parents indicated a willingness to have their children participate. Of the more than 300 letters sent out (one to every incoming 7th grader), some 200 replies indicated a willingness to participate; from these 200, 100 students were selected on a quasi-random basis, with special attempts being made to assure a racial, sexual, and behavior-problem balance.

In summary, Small House is identifiable administratively as a program which utilizes the flexible modular scheduling concept and team planning. The psychological environment and characteristics are directly related to the goals of Small House and will be discussed in a later section.

B. HISTORICAL PERSPECTIVE AND REVIEW

America's purported propensity for educational innovation and concomitant concern for evaluation notwithstanding, there is a paucity of systematic research on innovative educational programs (Carbone,1962; Hillson et al,1964; Westbury, 1970). Westbury (1970) reports that over a three-year period not one evaluation study appeared as a topic in any of eight major educational journals— despite the fact that evaluation continues to be one of the most frequent subjects of papers. This leads to the conclusion that evaluation is like the weather,i.e., everyone can talk about it but no one does anything about it!

---

[3] The reader is again referred to the paper by Radin and Morris for elaboration; also, some of this will be addressed in Chapter V.

The dearth of research is particularly acute with respect to the junior high schools; except for some descriptive studies (Stewart & Shark,1971; Marconnitt, G.R. , 1969) and sporadic investigations of team teaching (Zweibelson et al, 1965;Gamsky,1970; Boyles, 1968; Fraenkel,1968), the research focusing on the junior high school is virtually nonexistent! Therefore, it will be necessary to review research from a number of different areas and extrapolate from those results to the junior high school.

Most of the evaluation research and reports centering on innovation are directly related to the "nongraded vs the graded" classroom. Bockrath (1958), Ingram(1960) and Shapiski(1961) found significant differences in reading favoring the nongraded classes. Bockrath's research is of special interest since her report covered a period of three years and included n's of 5,000+ and 8,000+. Carbone (1961) reports that a 1952 comparison of 99 nongraded students with 123 controls showed reading achievement and personality adjustments were better for the nongraded students; moreover, Steere(1972) summarizes other studies from Appleton, Wisconsin and Bellevue, Washington, both of which support the contention that nongraded students demonstrate greater gains in reading. Buffe (1962) found that nongraded students made greater gains in math and mental health. Billson et al (1964) did a caerful study, controlling for many extraneous variables and found significant differences favoring the nongraded classes; a follow-up study, however, indicated that these differences disappeared over time (Jones, et al, 1967). More recently, Engel and Cooper(1971) and Morris et al (1971) found significant differences favoring the non-graded structure. Not all of the research results have been in favor of the nongraded classroom. Jones et al(1967) found no differences, as did Hopkins (1965); Carbone (1962) found that in all areas of achievement, graded pupils were significantly better than nongraded where the original scores were adjusted for intelligence; there were differences in the ways the two groups of students described their teachers, with the nongraded students using more favorable terms to describe their teachers.

There are several critical inter-related issues which must be kept in mind with respect to the cited research:

a. As Goodlad(1968) has pointed out, the distinction between the nongraded and "traditional" classroom is often a fuzzy, if not totally arbitrary, one. Labelling can hide the fact that some "traditional" classes have more in common with other classes designated as nongraded than other classes designated as graded. Engel and Cooper's (1971) recommendations constitute only a partial solution. This criti-

cism can be broadened to include distinctions among educational treatments (or curricula) in general; we assert differences a priori (by calling them different names) with little or no basis for doing so.

b. Methodological and design flaws characterize most of the research. Almost all of the research is restricted by a model which has gone little beyond the concept of paper and pencil tests; the failure to incorporate systematic observations as an integral part of a comprehensive model has been the most glaring defect of educational evaluation research.

c. In many cases, the research covers a brief period (1 year); those cases which have covered a longer period (Hopkins,1965; Jones,et al, 1967) have indicated no differences.

Like the research on the nongraded classroom, the research on team teaching produces different results. Zweibelson et al(1965) found no changes in achievement, but improved attitudes on those under the team teaching method. Gamsky's (1970) findings corroborate these results. Rhodes (1971), on the other hand, shows that the average reading gain was worse for team teaching, as were changes in pupil attitudes. The criticisms directed at research on the nongraded classroom are applicable to that focusing on team teaching.

C. RESEARCH EFFORT

Commencing in September of 1971, the author began meeting regularly with the teachers and spending time observing and taking notes in the classrooms. Observation periods were random or nonsystematic two or three times per week and continued through June of 1972. These notes form the basis of Chapter V and it was from these notes that the questionnaire (Chapter III) and the proposed observation schemes were derived. All of this, however, was directly related to the goals which the teachers had set for themselves (see page 1). We did not succeed in addressing ourselves to all of the goals set forth and developed by the teachers, at least on a thoroughly systematic basis; we emphasized those that were essentially observable and those that had been given higher priority by the Small House team. (The goals, as they were stated in their original formulation, are at Appendix III.) These goals could be rephrased to read:

a. Do Small House students show gains in performance on standardized reading and arithmetic achievement tests comparable to those of students in other classes?

b. Do Small House students show a more positive attitude towards school as compared to students in other classes?

c. Are there characteristic differences in teacher and student behavior between Small House and other classes? If so, what are they? How are they related to the goals of the humaneness report?

A severe limitation of this study was that all of the control classes were mathematics classes. This made it impossible to compare the observations of other Unified classes and Science classes with those of Small House.

CHAPTER II

PSYCHOMETRIC EVALUATION

A. CALIFORNIA READING TEST

Among the goals that the teachers had set for themselves were those related to academic achievement, particularly improved reading skills. The Small House team was particularly concerned that students in Small House show no decrement in performance in the basic skills as compared to the gains reflected by a control group, despite the fact that the former were spending less time in traditional classes. The California Reading Test (1963 edition, Form W) was administered to all the students in Small House and to a small sample of those in the control groups[3]. In order to fit the test into the 45-minute period allowed, some modifications of the normal testing procedure were required:

a. Only the Vocabulary Test and Interpretation of Materials subtests were administered.

b. Students were only given 30 minutes to respond to items in Section G of Test 2, Interpretation of Materials.

c. In scoring these tests, no correction was made for guessing. In addition, each correct response in Section G was given a weight of two (2) points to overcome the preponderance of vocabulary items (60 vs 45).

Table I presents the results of the groups' performance on both the pre- and posttests. It appears that there are no significant differences in the net gains made by either group. Because the control group was expected to be more homogeneous at the beginning of the year, it was decided to use Analysis of Covariance to compare the two groups. In order to achieve parity in group size, each student from Small House for whom pre and posttest reading scores was available was randomly assigned to one of four groups; thus, four separate analyses of covariance were performed, comparing each of the experimental groups with the control group. In effect, we have four tests of the hypothesis of no difference. Tables II - V present the results of these comparisons.

None of the F-ratios obtained is significant; nevertheless, there is one interesting difference that emerges from this analysis. The variance of the control group on the posttest was almost double that of the control group on the pretest. But in each of the four experimental groups, the variance on the post test was less than than the variance on the pretest!

[3] It was possible to only use a small sample due to the fact that these students were also taking other tests and it was necessary to minimize demands on class time.

TABLE I

PRETEST AND POSTTEST MEANS (RAW SCORES) ON CALIFORNIA READING TEST

GROUP	N	PRETEST MEAN	POSTTEST MEAN	NET GAIN
SMALL HOUSE	85	77.7	90.2	12.5
CONTROLS	20	79.2	92.6	13.4

TABLE II

ANALYSIS OF VARIANCE FOR FINAL READING SCORES BY COVARIATE ADJUSTMENT OF INITIAL READING SCORES

SOURCE	df	SUM OF SQUARES	MEAN SQUARE	F-RATIO
BETWEEN	1	60	60	
WITHIN	37	4307	117	.57
TOTAL	38	4367		

TABLE III

ANALYSIS OF VARIANCE FOR FINAL READING SCORES BY COVARIATE ADJUSTMENT OF INITIAL READING SCORES
(SMALL HOUSE SUBGROUP 2)

SOURCE	df	SUM OF SQUARES	MEAN SQUARE	F-RATIO
BETWEEN	1	281	281	
WITHIN	37	3995	106	2.76
TOTAL	38	4276		

TABLE IV

ANALYSIS OF VARIANCE FOR FINAL READING SCORES BY COVARIATE ADJUSTMENT OF INITIAL READING SCORES
(SMALL HOUSE SUBGROUP 3)

SOURCE	df	SUM OF SQUARES	MEAN SQUARE	F-RATIO
BETWEEN	1	27	27	
WITHIN	37	2616	72	.37
TOTAL	38	2643		

TABLE V

ANALYSIS OF VARIANCE FOR FINAL READING SCORES BY COVARIATE ADJUSTMENT OF INITIAL READING SCORES
(SMALL HOUSE SUBGROUP 4)

SOURCE	df	SUM OF SQUARES	MEAN SQUARE	F-RATIO
BETWEEN	1	14	14	
WITHIN	37	3320	89	.15
TOTAL	38	3334		

There are several viable explanations for this phenomenon:

a. It is simply a regression effect which is not evident in the control group since they are more homogeneous and, on the average, were among the more highly motivated and more able students—according to their placement.

b. That, in fact, the brighter students gained less—in an absolute sense—than those who scored lower on the pretest. This is a reasonable assertion (within limits) because of the intensive programs that Small House had for students whose basic skills were in need of remedy.

It is critical that certain limitations be kept in mind in reviewing the results of this portion of the research:

a. The number of students in the control group was smaller (n=20) and much more homogeneous than the students in Small House. This suggests that the groups were not comparable at the outset.

b. Testing conditions, particularly in Small House, were somewhat less than ideal, and not standard; in one case, due to interruptions, it was necessary to readminister the test.

c. Although tests were de-emphasized for all of the students taking them, students in Small House, especially on the pretest, expressed the belief that their performance on the tests would affect the future of Small House. It would be impossible to gauge the precise effect, if any, that this might have had on their performance.

Given the limitations of this portion of the study, the results must be considered tentative, at best. In general, they are consistent with the trends of the research on the nongraded elementary classroom. They share the common problem of research on the nongraded classroom in that a one-year period is insufficient time to adequately assess the impact of any broad-based social program. Certainly, some of the findings are provocative; if, for example, a follow-up study also discovered that a control group tended to show greater variance on the posttest than on a pretest while Small House students did the opposite, we could be far more confident of these results. It should also be kept in mind that the Small House teachers, while wishing to open up the educational experience and provide diversity for their students, specifically implemented programs designed at students with problems in the basic skills. From an evaluation context and in terms of understanding precisely what the treatment consisted of, this is a very cogent variable.[4]

---

[4] Other examples will be provided and discussed in Chapter V.

B. CALIFORNIA ARITHMETIC TEST

In addition to their concern with reading skills, the Small House team was concerned with improvement in basic mathematics skills as compared to students in other classes. Besides the fact that the students would be spending only about 60% as much time in math classes as their 7th grade counterparts, the Small House math program was constructed around individualized programmed instruction kits. Thus, the basic mode of instruction was radically different in conception from the "traditional" 7th grade mathematics class. In addition, though, special remedial instruction was offered for those who needed it.

The California Arithmetic Test (1963 edition, Form W) was administered to all of the students in Small House and to all students participating in the control groups. In order to fit the test into the 45-minute period of the normal classroom schedule, some modifications in administration procedures were required.

a. Only the computational portion of the test was administered.

b. The time limit for each section (Addition, Subtraction, Multiplication, and Division) was 10 minutes instead of the 10,10, 12, and 15 used in the standardization. This has a bearing only if we wish to examine the mean grade level placement or compare these students with the national norms, neither of which is a high priority goal of this research.

c. No correction was made for guessing.

Attention was restricted to computational skills because of time limitations and the fact that the teachers cooperating in the research indicated that these were the skills emphasized in 7th grade. As with the reading test, students expressed their belief that their performance would have significant bearing on the future of Small House; other than that, testing conditions could be considered standard for all groups.

The control group consisted of three different first-hour classes, taught by different instructors who also volunteered for observations and attitude testing of their students. One of the three control groups was designated a 372 (low level) class, while the others were designated 374 (average +). For this reason, the analysis includes a breakdown by class, in addition to combining the three classes for comparison with Small House. Table VI presents the pretest and posttest means for the groups involved.

TABLE VI

PRETEST AND POSTTEST MEANS (RAW SCORES) ON CALIFORNIA ARITHMETIC TEST

GROUP	N	PRETEST MEAN	PRETEST S.D.	POSTTEST MEAN	POSTTEST S.D.	NET GAIN
CONTROL A	19	25.4	8.25	37.9	14.1	12.5
CONTROL B	26	42.6	8.35	55.8	10.1	13.2
CONTROL C	24	39.3	10.34	53.0	9.4	13.7
COMBINED CONTROL GRPS	69	37.1	11.76	50.1	13.78	13.0
SMALL HOUSE	83	37.1	13.72	50.1	12.53	13.0

TABLE VII

PRETEST AND POSTTEST MEANS (RAW SCORES) ON CALIFORNIA ARITHMETIC TEST
— LOW ABILITY STUDENTS —

GROUP	N	PRETEST MEAN	PRETEST S.D.	POSTTEST MEAN	POSTTEST S.D.	NET GAIN
CONTROL A	19	25.4	8.25	37.9	14.1	12.5
SMALL HOUSE SUBGROUP	19	26.5	9.01	38.4	9.4	11.9

There are several noteworthy peculiarities of this table. The first, of course, is that when the control classes are combined the pretest and posttest means are identical to those of Small House students. The second is that regardless of which class was involved, all with different instructors, the mean gain for each class was approximately the same, irrespective of class mean on the pretest! Yet another is the tendency for the posttest variance to be greater than the pretest variance for the control group while the converse is true for the Small House students. This latter is particularly evident in Control Group A, the low ability class. It was decided, therefore, to follow the recommendations of Campbell and Stanley (1963, p 185) --albeit for the wrong reasons— and compare the performance of this class with a comparable sample drawn from Small House.

Table VII presents the results of this comparison. The subgroup drawn from Small House tends not to show the large increase in variance reflected by the control group. An additional analysis of the performance of the low ability students in the other control classes shows a tendency for them to fall further behind the mean over time. These findings were substantiated to some degree when the teachers identified those students whose scores showed the least improvement as those who had, in the eyes of the teachers, learned the least.

Like the previous test results (California Reading Test), these data should be read with a jaundiced eye. Even though they are based on a slightly larger N, there are still some questions as to the essential comparability of the groups on variables other than math. (N.B. All of the students in Small House had parents who volunteered to participate.) There may have been motivational differences between the two groups. Therefore, promising though they may seem, these results should be considered tentative at best.

C. SCHOOL SENTIMENT INDEX

Consistent with the humaneness report, one of the primary goals of Small House was to effect changes in the attitudes of students towards school—to "turn students on" to school. One of the ways selected to measure this change was the School Sentiment Index (SSI), a criterion-referenced measure developed at the Instructional Objectives Exchange at UCLA. Additionally, we set out to develop an observation schedule which would supplement the paper and pencil measure; hopefully, this would enable us to quasi-validate the SSI (i.e., students with low scores would demonstrate behaviors distinct from those with high scores). Moreover, it was also hoped that we could use some of the test items for inclusion in the observation schedule. Technical difficulties delayed the completion of the observation schedule to the extent that it was only used nonsystematically

and will be included in Chapter V.

One final goal of the use of the SSI was to identify some smaller number of items which would be more amenable to classroom administration, be sensitive to change, and yet still be reliable

1. Procedure. The SSI was administered to all students in Small House in September and the control group in October.[5] The posttest was administered to both groups in May,1972. The following instructions were read to the students:

> "We are interested in what 7th graders think and feel about school, their teachers, and their classes. Therefore, we would like you to read each statement carefully and indicate whether you agree or disagree with it. If you agree, circle yes on the answer sheet; if you do not agree circle no. Please make sure that you are circling the correct response—the number next to the answer you are circling should be the same as the number next to the statement. THIS IS NOT A TEST, THERE ARE NO RIGHT OR WRONG ANSWERS. None of the teachers or the principal will see what you answer!

The major difficulty with the test was interpreting the meaning of the word teachers. For the Small House students, it referred only to their cluster teachers and did not include electives. For the students in the control groups, it referred only to instructors in the basic subjects, Math,Science, Social Studies and English. Nevertheless, students in the control groups made it eminently clear that they considered it unreasonable to reply to the items dealing with teachers in any generic sense. This was not the case with Small House.

2. Results. Table VIII presents the pretest and posttest means on the SSI for both groups. These results support the proposition that Small House students show a significant gain in attitude between pretest and posttest while the control group does not. Table IX summarizes the four basic comparisons (t-tests) made. In order to highlight this comparison, it was decided to compare students in both groups whose scores changed significantly, a little, or not at all. These categories were arbitrarily defined as +11 to +30, +5 to +10, -4 to +4, -5 to -10, and -11 to -30. Table X presents the results of this breakdown; since there are different numbers of students in the two groups, entries will be expressed as percentages.

Significantly more students in Small House made gains of 11 points or more on their tests relating to attitudes towards school, as measured by the School Sentiment Index. This table also reveals, however, that approximately 10% of the students in each group showed poorer attitudes towards school at the end of the year than they had at the beginning. In some cases these changes were substantiated by observations which were conducted throughout

---

5 The reason for the difference in administration times was that the Small House students were going on a camping trip before October and it was considered necessary to minimize euphoric effects based on a single experience; however, the control groups could not be tested at the same time because of the need for parental permission.

TABLE VIII

PRETEST AND POSTTEST MEANS (RAW SCORE) ON SCHOOL SENTIMENT INDEX

GROUP	N	PRETEST MEAN	PRETEST S.D.	POSTTEST MEAN	POSTTEST S.D.	NET DIFFERENCE
SMALL HOUSE	83	71.1	14.4	76.7	12.6	5.6
CONTROL GRP	69	69.3	13.2	70.7	11.7	1.3

TABLE IX

COMPARISON	$M_2 - M_1$	t	
Small House-pretest vs Controls-pretest	1.8	.82	not significa
Small House-posttest vs Controls-posttest	6.0	3.[illegible]	significant - .01
Small House-posttest vs Small House-pretest	5.6	2.67	significant - .01 [6]
Controls - possttest vs Controls - pretest	1.3	.71	not significant [6]

TABLE X

ALGEBRAIC DIFFERENCES BETWEEN POSTTEST AND PRETEST SCORES ON SCHOOL SENTIMENT INDEX

NET CHANGE	SMALL HOUSE	CONTROL GROUP
+11 to +30	27.5%	8.3%
+5 to +10	19.8%	20.1%
-4 to + 4	42.9%	60.9%
-5 to -10	3.5%	6.5%
-11 to -30	5.5%	3.9%

6 The t-test for correlated means was used for this comparison

the school year on a nonsystematic basis, the nature of the change being consonant with the direction of the change of their attitude test score.

One additional analysis was conducted in an effort to identify the dimensions of difference between Small House and the Control groups—to identify the individual items on which there was a significant difference between the two groups on the posttest though no significant differences existed on the pretest. Table XI presents the results of this analysis; these results are misleading in a sense because the differences among the control groups on some items was as large—or larger—than the differences between Small House and the control group taken as a whole. Table XI presents those items that were significant at the .01 level; there were other items which -- though not significant—reflected a similar trend.

These results are probably the most important findings since they do identify dimensions where Small House has shown consistent significant differences which are related to the goals that the Small House team set for themselves and also to the broader issue of "humaneness." The one item which runs counter to the trend concerns peer relations (The other kids in my class are not friendly towards me.). It is also important to note that there were vast differences among the control classes along the same items; this simply verifies observations of and disucssions with students about certain combinations of instructors and the fact that those combinations of instructors will be viewed differently by different students.

3. Discussion. One of the major shortcomings of the School Sentiment Index is that it uses generalized referrents (our teachers are . . .; our classes are . . .) and students frequently expressed their feelings that it was not fair to ask them to lump teachers and/or classes together and judge them, that the only reasonable responses that they could make would be to assess a particular teacher or a particular class. This is a legitimate complaint, but it was necessary to work within the constraints of the system which included eliminating references to individual instructors.

Another complaint voiced by the students concerning the SSI was the simple "yes" or "no" answers required. The students felt that it would be more appropriate to have more categories of responses, "most of the time", "about ½ the time", Hardly ever", "never." Moreover, the students felt that the tests were too long (a problem we had anticipated) and recommended.

One of the major problems with this study is that the control groups represented a sampling of instructors which may have been biased by unknown factors; it is not safe to assume that at least one representative of all possible instructor combinations was present. Interpreting the results then becomes a matter of believing that those students in the control group represented a random sample of Slauson 7th graders and that their instructors for their basic courses represented a random sample of possible combinations. Otherwise, it is difficult to ascertain what it is that Small House is better than.

TABLE XI

PERCENTAGE OF STUDENTS RESPONDING AFFIRMATIVELY TO ITEMS ON THE SCHOOL SENTIMENT INDEX [7]

ITEM	SMALL HOUSE	CONTROL GROUP
I usually look forward to coming to school	76	58
This school is like a jail	12	30
There are many different activities from which I can choose what I would like to do	96	61
My teachers try to make school interesting	96	72
My teachers treat me fairly	92	66
My teachers try to make sure I understand what they want me to do	70	51
I get tired of hearing my teachers talk all the time	10	54
The other kids in my class are not friendly towards me	29	13
Our teachers usually listen to our ideas	90	51
We usually have fun in our classes as well as learn	88	66
Most teachers have no sense of humor	24	49

TABLE XII

PERCENTAGES OF STUDENTS IN EACH OF THE CONTROL GROUPS RESPONDING AFFIRMATIVELY TO ITEMS ON THE SSI [7]

ITEM	CONTROL A	CONTROL B	CONTROL C
This school is like a jail.	45	18	28
My teachers try to make school interesting	82	73	62
My teachers do not care about their students	31	13	24
My teachers treat me fairly	68	72	84
Most teachers try to help you	100	80	100
School is mostly a waste of time	25	00	12

[7] All differences significant at .01 level.

CHAPTER III

THE 24-HOUR QUESTIONNAIRE

Reference was made earlier to the abortive attempt to develop an observation schedule to be used in tandem with the SSI to assess attitudes towards school; this observation schedule was to focus on the behaviors of individual students to determine if they coincided with certain criteria which the instructors and research staff had agreed would reflectcertain kinds of attitudes. While this was not fully utilized, it did lead to using observations to construct a questionnaire for students to respond to with respect to specific events that had occurred within a 24-hour period. Thus, the students provided observations of what had happened to them. This was not intended to be another paper and pencil test of attitude although it is certainly a useful adjunct to the School Sentiment Index. But essentially it was hoped that this method would avoid the generality of the SSI.

1. Administration. The 24-hour questionnaire was administered twice to each participating class—once in the first week of April and again in the last week of April. It was always administered prior to the beginning of each class (each student's first class in the morning) so as to minimize differences and possible misinterpretation. The following instructions were read to the students:

> "We are interested in some of the things that happen to 7th graders during an average day. Listed below are a number of things that could possibly have happened to you or that you could have done in the past 24 hours (since you came to school yesterday morning). Please indicate whether these things did or didn't happen by simply circling yes or no behind each statement. For example:
>
> I went to see my counselor yes no
>
> If you saw your counselor anytime yesterday or this morning, you would circle yes, otherwise you would circle no. Remember, all of the items refer to things that happened between yesterday morning and this morning."

2. Results. Like the results of the School Sentiment Index, the results of this brief questionnaire are both provocative and valuable. To a large extent, they substantiate the results of the SSI—even though at least 6 months intervened between the two administrations. As with the School Sentiment Index, the Small House students reflected more negative peer interactions (itesm #4 and #9) than did students in the control group. Seven of the items yielded significant differences among the groups on both administrations. Table XIII summarizes the results of this questionnaire. Ironically, Small House students also responded significantly better to item 8 (A student did something nice for me) indicating an increase in positive social interaction. The other dimensions of difference appeared to be one of boredom (#2,#5, and #10) and positive teacher activities (#1 and #11).

TABLE XIII

PERCENTAGE OF STUDENTS IN SMALL HOUSE AND CONTROL GROUP RESPONDING AFFIRMATIVELY ITEMS ON THE 24-HOUR INVENTORY

	ITEM	WEEK 1		WEEK 2	
		SMALL HOUSE	CONTROL GROUP	SMALL HOUSE	CONTROL GROUP
1.	A teacher said something nice to me--something that made me feel good.	57	38*	51	30*
2.	I got very interested in something that a teacher was talking about and wanted to ask (or did ask) a question about it.	56	40*	57	35*
3.	I volunteered information based on my personal experience or something I had read.	32	20	36	18
4.	A student said something unkind to me or something that made me feel bad.	47	26*	43	19*
5.	I was very bored during a class and spent most of my time looking out a window, day-dreaming, or talking to someone.	48	73*	56	76*
6.	Another student said something nice to me or something that made me feel good.	64	66	61	63
7.	A teacher hollered at me or got mad at me for no good reason and wouldn't even listen to my explanation.	16	20	23	17
8.	A student did something nice for me.	60	43*	62	44*
9.	Another student or group of students made fun of me or picked on me.	28	11*	27	13*
10.	I couldn't answer when I was called on because I wasn't paying attention or I didn't do some reading I was supposed to do.	13	13	10	24*
11.	A teacher volunteered to help me with a problem I was having with my schoolwork.	45	23*	46	28*
12.	A teacher said something unkind to me or something that made me feel bad.	16	14	19	13

* = Statistically significant at .01 level.

3. Discussion. In addition to supporting—at least in part—the results of the School Sentiment Index, the results of this questionnaire are interesting in their own right. For example, approximately 75% of the students admit to being bored at least once during the course of a school day (50% in Small House); by contrast, we see that only 37% of the students admit to being interested in anything that was going on to the extent that they wanted to ask a question about it (55% in Small House). This should come as no great surprise to persons familiar with the traditional fare offered by junior high schools. Nevertheless, it is a thought-provoking statistic with respect to the breadth of the problem. (What portion of the day or how often the students are bored is an unanswered question) Similar comparisons can be made with other items —#1 vs #12 for example—to highlight other points of interest.

As with the School Sentiment Index, there were differences among the control groups on some of the items (# 2,5, and 7) which exceded those that existed between Small House and the control groups taken as a whole. This again illustrates the importance of the combination of teachers that kids have as well as the weaknesses of using items with nonspecific referrents.

The value of the 24-hour questionnaire is that it is highly specific, and the items themselves could be validated through observations. It is short (requires only 10 minutes to administer) and could be administered several times throughout the course of a year.

CHAPTER IV

SYSTEMATIC OBSERVATIONS

1. Introduction. Throughout the course of the school year, classes in Small House and the control groups were observed in an attempt to develop classifications of characteristic teacher and student behaviors that were both comprehensive and precise. At first, these observations were non-systematic in nature; only later were they formalized into an observation schedule. This schedule was finalized and tested through the month of April.

Beginning in May, all of the classes participating in the study were observed regularly by two observers on three separate occasions for a total of six 45-minute observation periods per teacher. Each observation period included two 10-minute blocks of data collection; for the control groups this meant one block at the start of the class and the second block starting at the middle of the class. For the Small House classes, this meant two consecutive modules of observing the same teacher.

Each 10-minute block consisted of two parts. The first part was a coding of teacher-student interactions (coding the behavior on every 15th second—20 codings per 5 minutes); the second portion included two scans of the classroom, noting what each student was doing. Student-teacher interactions were coded on the basis of who initiated them and what the content of the interaction was. Student behaviors were classified as traditional learning, nontraditional learning, socializing, observing, or daydreaming/withdrawal. (For a complete description of the Coding Scheme, see Appendix IV.)

One of the serious limitations of this part of the study is that all of the participating classes in the control group were math classes[8] and that comparisions therefore are not directly possible between Small House and another total treatment combination, or even between Small House Unified classes and traditional Unified classes. It is for this reason that the original comparisons are simply between the Small House math class[9] and the other three math classes.

2. Procedure. As the number of naturalistic observations we had made increased over time, we attempted to break them down into well-defined, mutually exclusive categories and then use these to guide observations. The major stumbling block confronted was a band-width—fidelity dilemma. In order to achieve consistently high agreement, we had to widen the parameters of our categories. The system selected was considered the best alternative available from the perspective of clarity, distinction, and reliability. Subtler discriminations and specificity were temporarily sacrificed due particularly to time constraints.

---

8 Several classes which initially agreed to participate in the study dropped out, leaving only the math classes.

9 Only required math classes in Small House were observed, not electives.

Starting in April, the two main observers began observing classes in order to establish fundamental reliability. Consistent with the recommendations of Medley and Mitzel (1963), Westbury (1967), and McGaw et al (1972) reliability was conceptualized as having three distinct components or facets:

a. Inter-observer agreement. Two observers in the same class at the same time will make essentially the same statements about the process.

b. Intra-observer agreement. One observer in the same class at two different times will not comprise a source of error, i.e., will be consistent in his use of categories.

c. Inter-observer reliability. Two observers in the same class at different times.

Medley and Mitzel (1963) summarize the distinction thus:

> ". . . The coefficient of observer agreement tells us something about the objectivity of an observational technique; the coefficient of stability tells us something about the consistency of the behavior from time to time.[10] But only the reliability coefficient tells us how accurate our measurements are." (p 254)

a. The coefficient of inter-observer agreement (C_1) was calculated using the formula

$$C_1 = \frac{\left(\sum A\right)^2}{\left(\sum O\right)^2}$$

$\sum A =$ the number of agreements between observers and $\sum O =$ the total number of observations. In this case, O was always 20 since 5-minute blocks were used. Agreements were based on the number of matched pairs — both observers had identical notations on the 15th second, the 30th second, the 45th second, etc. The C_1's obtained for 24 training trials ranged from .41 to 1.00 with a mean coefficient of .77.

b. Consistency Coefficient (C_2). It would be illogical and impossible to calculate a Consistency Coefficient in the manner in which the C_1 was computed because while we expect the general trends to be the same, it would be presumptive to expect identical sequencing across occasions. Consequently, a different computational procedure was used to determine C_2 than that which was employed for C_1.

---

10 It is here that I would part company with Medley and Mitzel; they assume or at least imply that the observer is a flawless recording device, an unwarranted assumption from my perspective. I prefer to regard b as a consistency coefficient— to what extent does the observer report the same behaviors on different occasions, even though both observers and behavior may vary.

$$C_2 = \frac{\left[\left(\sum O\right) - \frac{\sum_{i}^{K}\left(O_{i1} - O_{i2}\right)}{K-1}\right]^2}{\left(\sum O\right)^2}$$

$\sum O$ = the number of paired observations, O_{i1} = the frequency of category i on trial 1 and O_{i2} = the frequency for category i on trial 2. The C_2's obtained across 24 training trials for each observer ranged from .51 to 1.00 with a mean coefficient of .84.

c. Reliability Coefficient (C_3). Computationally, the reliability coefficient is identical to the consistency coefficient or stability coefficient; however, rather than having the same observer on two different occasions, the protocols of different observers are utilized. The C_3's obtained during training ranged from .33 to .94 with a mean C_3 of .72. The trials used to compute the C_3 were not the same trials used to obtain C_1 and C_2 data.

Beginning in May, each teacher participating in the study was observed on three separate occasions by each observer. Each teacher knew ahead of time when to expect the observer. Thus, for each teacher there were a grand total of 240 observations of student-teacher interactions and 24 n (where n=the number of students—approximately 22— in the class) observations of student behaviors.

3. Results. Tables XIV and XV present the results of these observations for the teacher-student interactions and student behaviors respectively. As Table XIV indicates, there are pronounced differences among instructors with respect to the interaction patterns. There are significant differences between the Small House instructor and the control groups taken as a whole along four of the five basic dimensions or combinations (1-5 singly, 1&2, 3 & 4, 2 & 4). Again, however, it is evident that the differences among the control group teachers equal or exceed some of the differences between Small House and the control teachers taken as a unit. The singular exception to this is the amount of teacher-initiated academic interactions; however, as indicated earlier, the particular program implemented in Small House was an individualized program.

It is also interesting to note that only 80% of the teacher-student interactions in Small House were academic in nature while in the control classes the percentages ranged from 91.5% to 95.9%. This difference is significant and constant, and tended to be relatively evenly distributed between social interactions and control-oriented interactions. The difference in percent of control statements, while not significant, is in and of itself a phenomena worth considering. It has been axiomatic that if you offer new and unusual opportunities to students they will respond by behaving appropriately; these results would not support that particular contention. No record was made of the kinds of control

TABLE XIV

PERCENTAGES OF STUDENT-TEACHER INTERACTION MODES FOR FOUR MATHEMATICS CLASSES

TYPE OF INTERACTION	TEACHER A	TEACHER B	TEACHER C	TEACHER D
Teacher-initiated academic	40.7	67.2	85.0	55.2
Student-initiated academic	39.5	24.2	7.0	40.7
Teacher-initiated social	2.0	3.2	1.0	0.0
Student-initiated social	5.2	0.0	0.0	1.2
Teacher-initiated control	11.1	5.4	7.0	2.7

TABLE XV

PERCENTAGES OF STUDENT BEHAVIORAL PATTERNS FOR FOUR MATHEMATICS CLASSES

TYPE OF BEHAVIOR	TEACHER A	TEACHER B	TEACHER C	TEACHER D
Traditional learning	58.4	57.5	59.2	78.2
Nontraditional learning	9.4	2.1	2.2	2.1
Socializing	22.4	24.8	19.1	9.1
Observing/passive participation	1.8	5.8	.72	1.4
Daydreaming/withdrawal	8.0	9.8	18.8	8.5

(positive or negative) used.

Something that is not evident from the tables is that the characteristic interaction patterns were generally highly stable from day to day and across observers. As will be shown later, this too appears to be a function of the particular teacher. Table XV suggests that student behaviors might be independent of interaction patterns. With a few notable exceptions, the percentages of students falling into each of the categories were virtually identical for all classes. Again, within teachers, these percentages were stable over time and across observers.

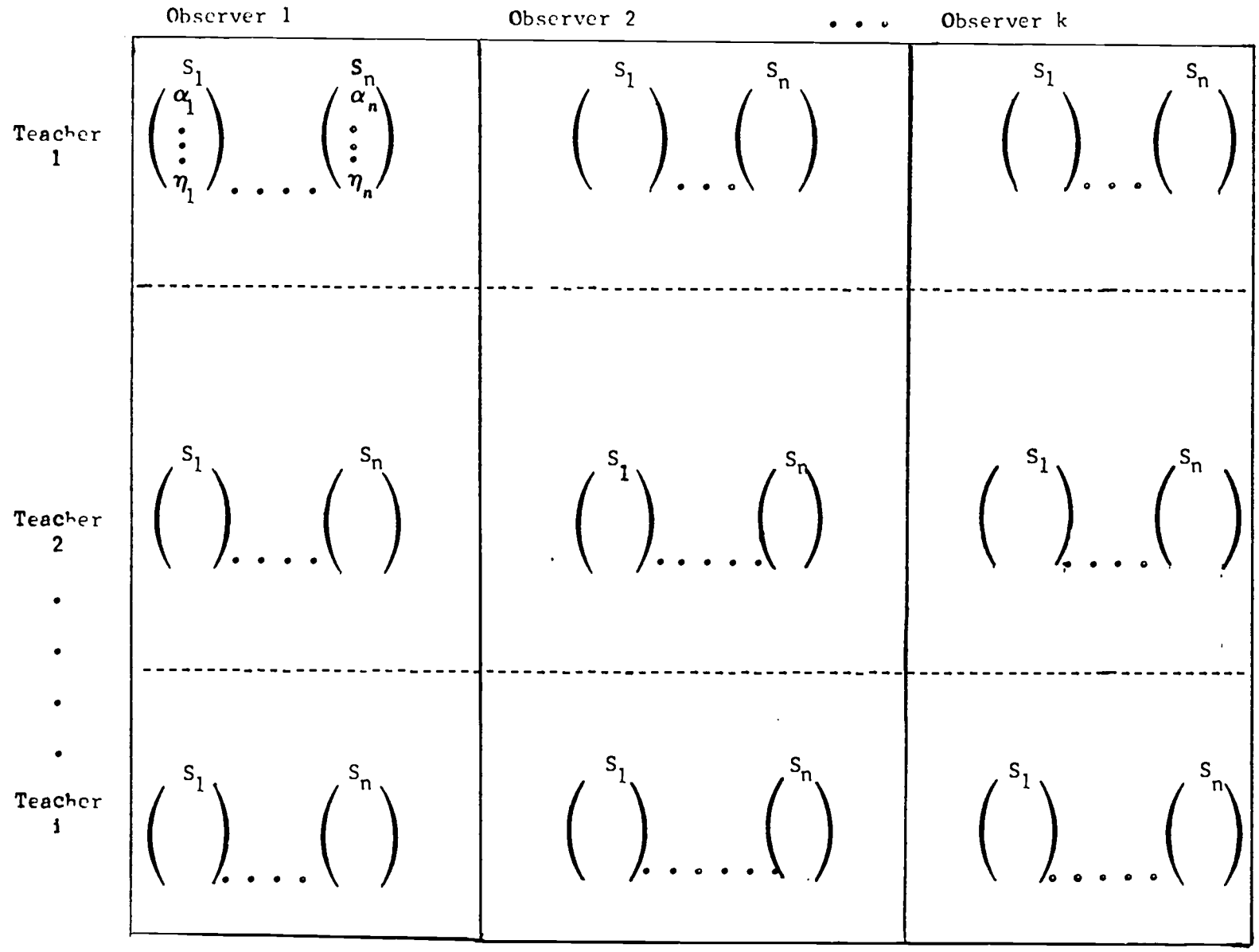
The comparable data for the other three Small House teachers is presented in Tables XVI and XVII. On the average, there are more student-initiated interactions and fewer teacher-initiated ones for these teachers than for the others. Also, there tends to be significantly more teacher-student social (as opposed to academic) contacts in Small House. Conversely, contrary to expectation, Small House instructors made more frequent use of control statements and the students' activities indicate far more non traditional learning instances than occur in the control groups. The regrettable lack of control classes in Science and Unified Studies makes it impossible to determine whether these differences are attributable to differences in subject matter or teacher role.

It is important to note that these results are based on observations of required classes in Small House. Thus, they do not take into account the other very broad activities and opportunities offered the students. This tact was absolutely essential to minimize the differences between subject matter areas since most of the electives offered were offered through Unified Studies. Also, the observations of nonrequired classes were far more difficult to code and less reliable. Nevertheless, many observations relating to these classes will be included in Chapter V.

Following the recommendations of McGaw et al (1972) we can identify the variance components of the observations by creating a breakdown similar to that of Figure 1 (page 25). Using such a model it will be possible to identify the extent to which teacher/student behavior patterns vary from day to day, how much observers tend to differ, and how much interaction there might be between observers and teachers. Applied to this particular problem, each vector represents the frequencies appearing for each observation period. It is then possible to do an Analysis of Variance, using vectors as entries (see Figure 2), and ascertain the variance due to teachers, observers and teacher-observer interactions by categories.

FIGURE 1

MODEL FOR IDENTIFYING VARIANCE COMPONENTS OF OBSERVATION SCALES
(GENERAL)



S_1 = Situation 1, S_n = Situation n, α and η are observations or frequency counts (for category scales) made by Observer k while observing Teacher i in Situation n. Thus, for example, if you had 3 observers, 6 teachers, and each observer observed each teacher 3 times, you would have 54 matrices, i.e., [3 x 6 x 3] values for $\alpha \ldots \eta$.

FIGURE 2

SAMPLE ANALYSIS OF VARIANCE — SMALL HOUSE OBSERVATIONS

SOURCE	SUMS OF SQUARES	df	MEAN SQUARE	F-RATIOS
Observers	10.6 3.3 • • 5.1	1	10.6 3.3 • • 5.1	.33 .31 • • 1.4
Teachers	2576.4 1112.3 • • 70.5	6	429.4 185.3 • • 11.7	13.0 * 19.1 * • • 2.0
Teachers x Obs	124.2 231.6 • • 12.8	6	20.7 38.5 • • 2.14	.6 4.0 * • • .4
Error	944. 266. • • 154.	28	33.8 9.5 • • 5.5	
Total	3654.8 1612. • • 242.	41		

* Significant at .01 level

The results presented in Figure 2 indicate that there are no main effects due to observer differences and that there are significant differences among teachers for all categories with the exception of teacher-initiated control. But, there is also a significant difference in the interaction between teachers and observers for the use of one category (student-initiated academic interactions). This suggests either that for one or two of the instructors certain behaviors were not stable over time or that in fact one of the observers was biased in his coding for one or two teachers. Since the coding was a must coding system, this latter seems a more plausible explanation.

CHAPTER V

NONSYSTEMATIC OBSERVATIONS

Allusions were made earlier to the difficulties inherent in attempting to supply an adequate description of Small House from anything more than a superficial structural perspective. The intent of this chapter is to examine in detail data which are relevant to this research but which would not fit neatly into any of the predefined categories, some of which could be subsumed under the rubric of "confounding variables" and some of which are simply anecdotes or generalizations concerning some area of concern. Later (in the following chapter) an effort will be made to integrate all of the data from the project into some kind of coherent whole, discuss both the methodological flaws and the ramifications of this research, and present the skeleton of an evaluation model.

All of the data in this chapter are based on voluminous notes taken between September 1971 and June of 1972 by the observers; they are in no way intended to be presented as systematic or representative observations and should not be construed as such. In some instances a single issue or inference made has been derived from a large number of observations; at the same time, other occurences may be mentioned only once, though they happened twenty times.

A. SOME CONFOUNDING VARIABLES

As mentioned earlier there were several cogenc variables which may have had a significant impact on the Small House program; some of these are directly related to the team's goals, others were functions of the team structure itself, and still others were functions of outside influences. Regardless, they are issues that must be addressed. While it might be defensible to consider these variables an integral part of Small House (Scriven, 1967,pp 50-51), to do so without specifying what they are could inhibit replicability and generalizability.

1. Administrative Support. That Small House was first implemented at Slauson was not entirely accidental. The school's administration was particularly committed to the concepts espoused in the Humaneness Report. The principal actively supported and encouraged the Small House team from its inception throughout the year,without attempting to assume responsibility for directing the program.

One of the essential components of this support was the granting to the teachers a great deal of freedom in designing their curriculum, handling their discipline problems, and working through a lot of technical details. This "freedom to experiment" accorded the teachers in turn seemed to generate

within them a tremendous amount of involvement. They appeared willing to work twice as hard as they previously had—and they admitted to other instructors that it was more work—in order to keep the program running. At times, the teachers were spending their own money for some of their projects. The teachers acknowledge that two of the reasons they were willing to put forth the effort were the psychological atmosphere (at least from the supervisory level) that was conducive to their ideas and the fact that this was their program, not something imposed from without.

2. Selection Bias. One other consideration that might attenuate the results is the selection biases involved—for both the students and the teachers. On the one hand, all of the students in the program were there because their parents—for whatever reasons—had decided that an alternative educational experience was preferrable to the more traditional offerring available. Consequently, there is at least one, and possibly more, psychological dimensions which would distinguish Small House from any easily obtained control group.

On the other hand, the teachers themselves were volunteers (p 1) who had been hand-picked by the principal from the larger group of volunteers. The fact that they had volunteered again suggests that they--along with the rest of the teachers who volunteered--were different in some respects, along some dimension. The fact that they were selected by the principal suggests that they fulfilled whatever criteria he was using more so than the other instructors; this only further complicates the issue.

3. Team-Cluster Structure: Side Effects. Although the concept of the planning team was an integral part of Small House as designed by the team, it also generated a number of side-effects which could not be considered as fundamentally generalizable. The planning team and the flexible modular scheduling made these things possible; however, the fact that they were conceived and executed was a function of a number of other variables.

For example, several different groups were involved in operating a number of different programs for Small House students that the teachers themselves did not design or implement. These experiences,nevertheless, were offered as electives for students in Small House. One of the outcomes was that class size was often diminished considerably for students taking courses offered by Small House teachers. This extra help itself is a confounding variable.

The cluster structure also seemed to have a significant impact on relations with parents. The teachers could plan events such as a picnic or pot-luck supper without having to do it alone. During the course of the year, several such events were sponsored. One of the effects was to obtain active support from parents of Small House students, many of whom proved to be resources for Small House. This would have been impossible (or virtually so) without the team-cluster strucutre; neverthelss, the team-structure itself provided no guarantee that such would occur.

4. Operational Modification. Since this was the first year during which Small House was in operation, there was a great deal of experimenting and changing going on over the course of the year. In many respects the program at the end of the year was radically different from what it was at the beginning of the year. Consequently, the program might be conceived as a series of treatments with continuity of personnel, structure, and programs, but differing over time with respect to one or more dimensions. The degree to which these differences affected individual students would be a function of the particular student. For those students with special needs, for example, the "required electives" concept may have radically altered their psychological environment whereas other students would have been unaffected by it.

5. The Novelty Syndrome. The evolution of any deviation from established practice—especially in education—tends to act as a magnet for persons with abiding interest in educational reform. Small House was no exception. At one point in time, there were eighteen (18) different adults participating in some capacity in the operation of Small House. This included teachers, student teachers, C-390 students, parents, and others. This in and of itself would not necessarily be a particularly cogent variable except that the philosophy of the Small House team was to involve everyone to the maximum degree possible; people were permitted to come in and just observe, but were encouraged to do more than just watch. From their first day on the job, student teachers were pushed to get to work, encouraged to develop an elective or electives in accordance with their interests and to teach. C-390 students more often than not ended up by getting involved in some special project. It was very interesting to note that the teachers allowed their student-teachers and others the same freedom which had been granted to the team.

This is not considered an exhaustive list of the possible confounding variables, simply an identification of the major ones. While it would be impossible to isolate the specific contribution of these factors in any statistical fashion (let alone their interaction), it is necessary to recognize that they constituted part of the treatment included in the term Small House. For example, if the freedom and support given the teachers provided the impetus for much of the teacher involvement, then this freedom and support would be a necessary ingredient for other programs in the future—even though it is highly possibly, indeed probable, that a different set of four teachers would design a program that was radically different structurally and/or psychologically. This creates a tremendous paradox with respect to essential reproducibility, i.e., you cannot guarantee both the freedom and the structure (psychological and physical dimensions) of the program in any attempt to replicate with a different set of instructors. By imposing the structure, you remove the choice on the part of the instructors: by granting the freedom, you leave open the possibly selection of an alternate structure.

B. SOME GENERAL OBSERVATIONS

It would be easy to develop a rigorous program for studying some

phenomenon and execute it without every stopping to consider what one is not measuring. There was a great deal more to Small House than what has been presented by examining test scores or systematic observations; the constraints of the system plus the limited resources of the evaluators combined to reduce the capability to produce as thorough a study as was desired. The data that follow are fundamentally anecdotal materials which highlight aspects or dimensions of Small House not evident in the cut-and-dried type of data presented in the first four chapters. Besides being of value for understanding some important ideas about Small House, they identify topics which might be addressed by future evaluation studies.

1. Observations of Individual Students. One source of data ignored thus far is that which was generated by observing individual students and changes in these students over the course of the year. Nonsystematic observations of these students indicate radical changes in behavior between the beginning of the school year and the end of the year. The difficulty with interpreting such changes is that confounding variables must be considered. There were other significant events (psychotherapy, behavior modification intervention projects, etc) impinging on the lives of these students which could have effected these changes in and of themselves. All such cases were excluded from data presented below.

a. There were two students in Small House who were designated as Class A Mentally Retarded individuals. In the early part of the year, there were some cruel remarks directed at these two; no such remarks were heard or observed after January and at no point was teacher intervention required. By the middle of the year (December-Janaury) these individuals were well integrated into group activities. This was in sharp contrast to the two Class A students observed in other classes who remained virtual isolates throughout the school year.

b. There were three or four students who had had extremely poor attendance records throughout elementary school whose attendance at class was regular. Record keeping being what it is, precise figures are not available to compare the figures among years. In the judgments of administrators and teachers, and according to their peers, there were significant differences in the attendance of these students.

c. On at least one occasion, a student enticed to cut classes refused to do so.

d. Six or seven students with poor academic records were observed on more than one occasion working on their own during break or during one of their electives on some area with which they were having difficulty. On some of those occasions they were soliciting assistance from a teacher or a peer.

e. Even some of those students whose academic, attendance, and behavior records were particularly good demonstrated some marked changes—becoming

somewhat more open and friendly over time, getting involved in different projects and generally smiling more. These observations were substantiated by parental reports.

f. There were a small number of students (4 or 5) who demonstrated absolutely no changes between September and June and whose behavior warranted change. One of these students dropped out of the program (and the public schools) after one semsester; the other four stayed for the entire school year. The crucial consideration here is whether these kids would have had any chance at all in a normal school program—which is both an unfair and an unanswerable question.

2. Rap Sessions with Students. Much of our data came from informal discussions with the kids, both in and out of Small House. These discussions provide additional insight into the process of Small House as a treatment milieu. The basic findings of these discussions could be summarized as follows:

a. Ninty percent of the students in Small House viewed it as an incredibly good thing during the early stages. It was different and exciting and as a group they were getting a great deal of attention.

b. Beginning in December, many of the Small House students began expressing the sentiment that they were bored (more so than their counterparts in traditional classes). A significant percentage began expressing a sentiment voiced elsewhere that they were not learning anything. Behavior problems became more common than they had been. This precipitated a change in the polides which the Small House Team had towards discipline, an event which alienated some of the students even more. I feel that at this point in time the halo effect had worn off for both the students and the teachers. The general flatness and disenchantment that seemed to occur in Decmeber and January appeared to be a function of a letdown on the part of the teachers (who had been running themselves ragged), modification of the Small House policies, and an incredible emotional upheaval concerning the future of Small House.

There was a noticeable change in the program beginning in late January or early February which was reflected in student behaviors, teacher behaviors and discussions with the students. According to the students, this renewed interest was a function of the fact that they were having more input into the selection of alternatives for electives and other changes that they saw the teachers making. Changes in the discipline policy were also cited. In any event, between February and June, the program developed a high degree of consistency even though the teachers continued to experiment with ideas.

c. Discussions with the non Small House students elicited mixed reactions. Most admitted that they were envious of the opportunitites open to the Small House students, particularly such things as the camping trip. However, approximately 40% of those interviewed claimed that they would not want to be in Small House for several reasons. One was that the kids in Small House were

not learning anything; another was that many of the kids in Small House were "losers." It is impossible to ascertain how much of this is sour grapes, how much of it reflects the ideas of significant others or how much of it is a function of peer group pressure.

d. There were a small number of students in Small House who maintained from the srart that they were not particularly enthralled with the concept of Small House and who disliked it throughout. These students included individuals from all spectra—academically advanced as well as academically retarded, black and white, male and female. The one characteristic which they shared was a total lack of involvement in Small House activities. Most of them did not go on the camping trip or become involved in any of the other projects open to them even when encouraged to do so.

e. In addition to the aforementioned students, there was another small group who admitted that while it had been a lark for one year they would prefer a more normal eighth grade experience. Some of them admitted bluntly that they were afraid that they had not learned anything, other indications to the contrary.

3. General Classroom Atmosphere. There was a qualitatively different atmosphere involved in the two settings. However— as in other areas — the difference among some of the control classes were as great as the differences between Small House and the control classes considered as a unit. Basically, the Small House classes were much more lively — including both positive and negative kinds of interactions. There seemed to be more laughing and smiling on the part of the Small House kids than generally was evident in the control classes. A word of caution is in order. Our experiences at Slauson indicate that the psychological environment of a classroom is a highly variable factor, a function of many other factors ultimately dependent on the perceptions of the individual student and probably independent of the teacher's behavior in the classroom as viewed by an adult observer!

Perhaps the most interesting observation about classroom climate was with respect to one of the control classes. The first term it was a highly routinized setup with minimum student participation and a high degree of apathy and hostility — in the judgment of observers. That same class in the second term was radically different; there was a particularly strong-willed and hard-working student teacher who was granted permission to try out her ideas — with remarkable success. The instructor himself admitted that he could see and feel the difference in the class. Essentially, they had gone to totally individualized instruction. There were fewer behavior problems, the kids were still learning, the noise level was higher but even the instructor was happier. However, after the student teacher's departure, he returned to the former mode of operation since the alternative was "too much trouble" to run by himself.

C. ESTIMATION OF GOAL ACHIEVEMENT

Scriven (1967) has made the crucial distinction between evaluation and estimation of goal achievement. And, while one may not concur whole-heartedly with his granting of seeming omnipotence to the evaluator, this specification of evaluation as more than estimation of goal achievement is essential for clarity.

The Small House team had set themselves a series of goals prior to implementing their program. As part of the evaluation process, it is legitimate to attempt to determine the extent to which they approximated those goals. Not all of these goals will be addressed; some of them were stated in such ambiguous terms that attempts to assess them even peripherally were abandoned. It is also important to remember that these judgments are subjective estimations based on data which were sometimes contradictory.

1. Goal: Improved Reading Skills. Given the limitations of the control group in this particular instance and the limited reliability of the test, any answer to this particular question must be qualified. The Small House students seem to keep pace (using mean gain scores) with the control group; the gains demonstrated by the Small House students over time (mean gain in raw score) is equivalent to that of the national sample. Unfortunately, the reliability limitations and the entire "change score" problem prohibit making any effective judgments about individuals. Nevertheless, there were several other indications that for several students who needed it the most this goal was effectively achieved.

2. Goal: Improved Math Skills. The larger control group involved in this case provides some measure of comfort and a little more faith in these results. Again, the gains demonstrated by the Small House students were comparable to those of the control group and equal to those expected on the basis of the national sample's mean gain. Again, the reliability question and the "change score" problem qualify the affirmation of the achievement of this goal.

3. Goal: Turn Students On. Some of the measures used to estimate this particular goal indluced absenteeism and tardiness—both in relation to the control group and when using each student as his own control. This latter was done particularly for those students whose records indicated a high degree of absenteeism or tardiness in the past. Other measures included the School Sentiment Index, discussions with students, and observation of pupil activities. (See Appendix). The degree of success which the Small House team achieved for most of the students (80% or so) was high. The frequency of positive instances or examples of this type of behavior seemed to increase over time for most of the students. There were a few students who literally never seemed to be interested in what was happening.

4. Goal: Interpersonal Growth: In addition to those items on the School Sentiment Index which related to this, the criteria used included specific behaviors (see Appendix) demonstrated by the students. There was some discrepancy between those questionnaires and some of the events observed.

One example of a significant event is that of a girl who approached one of the teachers asking what she could do about one of her friends. Her friend was extremely biggoted; the student wanted to remain friends with this person, yet had very strong feelings about bigotry. The two things worthy of note here is that the issue itself was enough of a concern to the student to cause her to raise the issue and secondly that she chose to do so with one of the teachers.

Another striking example in this domain centers around one of the black students in Small House. Originally, his only contacts with his peers were with two of the other black students in Small House. He was also very hostile towards almost all of the adults with whom he came in contact. His participation in any group activity was virtually nil. Although it took almost six months to alter this even minimally, by June there was a high degree of interaction centering around this particular student, he had begun to activiely participate in many of his classes and was even observed smiling (something that had not happened at all in the first six months of observation) on several occasions.

These are but two — and the most dramatic two — of a number of events within Small House which reflect the attainment of this goal. Again, there was some small number of students for whom this would not necessarily hold true.

5. Goal: Intrapersonal Growth/Improved Self Image: Originally, it was hoped to assess this particular variable using a combination of psychometric devices and behavioral observations. The latter were foregone due to the difficulty of arriving at specific behaviors that could be agreed upon as measures of self-concept; the former led to one of the more bizarre, but educational, experiences of the study.

The Self Appraisal Inventory (developed at the Instructional Objectives Exchange at UCLA) was administered to both the experimental and control groups in September and again in May. The original administration involved little or no difficulty, although 2-3 students ineach of the groups either refused to take the test or responded in such a fashion as to invalidate the results. When we attempted to bive the posttest to Small House, however, more than 50% of the students simply refused to respond, either in part of in whole. Many of them wrote notes on the answer sheets, after having answered none of the questions or responded only to those that they chose to answer. Some of the comments written are included below:

"You are too nosy" (from four students)

"Some of these questions are none of your business!"

"I do not believe ehat any of this is anyone's business but my own and it is my choice to telly anyone, only if I want to."

"I do not believe I have to answer that."

"I do not have to tell you any of this."

" I don't think it's any of your business to ask us these questions. My life is my business and nobody else's. You can look at my grades if you want to know whether I am a good student but I feel that I am under no obligation whatsoever to answer these questions."

"Unless you are going to give each person who took this test some sort of help, you have no right to this information . . . It makes me feel like a guinea pig."

"You are infringing on my right to privacy."

After the testing situation, we had an opportunity to sit down and discuss all of the parts of the study with a number of students (30-40% of them), but we wanted particularly to talk about the Self Appraisal Inventory. I was concerned about why they had acceded to readily on the first administration and only later objected so strenusously. Almost to the man, the students admitted that they had felt pressured into taking the tests—all of the tests—because of the emphasis being placed on evaluating Small House, i.e., they did it to help the teachers. They also admitted that in fact they virtually randomly on the first test, or answered always in a positive direction, or developed some other response strategy. Fortunatly, they admitted that for the most part they had taken almost all of the other tests seriously, including the attitude scale. One of the reasons that they made the distinction between the attitude scale and the self-concept scale was that they could understand the justification for the inclusion of the attitude scale as being an integral part of the study.

This entire experience has raised several issues concerning classroom research, some of which will be addressed in the following sections. However, it also provides an index of intrapersonal growth which is far more relevant than a gain of 7 points on some arbitrary scale.

6. Goal: Greater Student Invovlvement in Learning and Educational Decision-Making. Generally, this goal was meant to be implemented at a number of different levels. On the one hand, teachers wanted to encourage each student to assume responsibility for personal growth in the academic disciplines. Therefore, there were a limited number of required courses for each student. Even within these, students were frequently offered options to do something else during that particular period, if the student felt that he had sufficient mastery of the topic being covered. Over time, the team did decide to impose one restriction. For those students who were doing especially poor work, some courses were required until they showed some improvement (the courses required included Math Lab, Study Skills, remedial work, etc.). Discussions with the students revealed that very few of them objected to this policy though they all admitted that they would on occasion skip this additional required class if something particularly interesting was happening in one of the other electives.

On a broader level, the students were invited to have input into the planning and scheduling of electives, or even their construction and design. As mentioned earlier, even though this had been the policy from the start, very few of the students seemed to take it seriously until the December-January letdown. After that, the kids began generating things that interested them, and in many cases took the lead away from the teacher in design. For example, some of the students suggested offering a drama elective, which a member of the team agreed to do, thinking that the students wanted to simply read or walk through a simple drama. The students, however, insisted on putting on a full production of West Side Story and later, Oliver. The students assumed responsibility for initiating other electives including one in American History, one in Geography, and other areas as well as the less academic offerings.

7. Goal: Self-Initiated Learning. One of the major thrusts of the Small House experience was to induce students to take a far more active role in the educational process. Thie invovlement in decision-making was one part of this; another was to encourage students to explore on their own topics and areas and ideas which were of interest to them and not depend on the teacher to undertake a learning experience even in the traditional domains. There were many instances observed where individual students would enter a class and set about their work with no instructor present. Additionally, on a number of accasions, different students were observed during break or an elective period working on some area with which they were having difficulty and even soliciting the assistance of one of their peers. Students did such things as write their own radio broadcasts, develop a multisense-nonverbal communication process, and invent machines. They also developed issues for debate and then debated them in small groups. Much of this was accomplished with minimal guidance from the teachers.

The evidence available would suggest that each of these goals--to varying degrees--was in fact achieved. This indtoruces another relevant consideration. These goals were an integral part of Small House and as such, and to the extent that they were achieved, constitute additional "confounding variables" (pp 28-30) affecting replicability/generalizability."

D. THE NONREQUIRED CLASSES

The list of electives ~~at Appendix~~ provides some gross indication of the alternatives available to a student at any one point in time, and this is only a partial list. The nonrequired classes are of interest for more reason than simply the diversity of content options which they provided the Small House students. While they fulfilled this major function, they also provided a forum for interchange between teachers and students on a level radically different than that even in the required Small House classes.

The well-defined distinction between student and teacher— retained even in the Small House required classes — was considerably diminished in most of the electives; the atmosphere was more like an open classroom model. In short, the differences between nonrequired classes in Small House and required classes may have been greater than the differences between Small House required classes and the control classes along a number of dimensions.

There are two substantial considerations focusing around the nonrequired classes that generate some concern and also have implications for future such endeavors. The first is the question of required electives. From the first, the teachers had offered among the electives courses designed to aid individuals in remediation of their basic skill deficits. After the firts eight weeks or so of operating, the teachers felt that there were certain students who were making no progress and were in need of extra help who were not taking advantage of these options. Their solution was to begin designating these students to take these electives (but never totally eliminating the options for this group) until they showed some improvement. The fact that they had to require students to take these electives went against the grain of some of the team members — they considered it philosophically inconsistent. For the most part the students involved accepted the decision as a reasonable one and some of them seemed grateful that the teachers had taken this action.

The issue at stake is delineating the true philosophical tenets underlying the program. If you are committed to some extent to granting students freedom of choice and are simultaneously concerned about his lack of preparation for the future and if you have a student unable to tolerate too many degrees of freedom, what takes priority? Is there some course of action available whereby you can meet what at first glance appear to be contradictory goals? No answer will be attempted here. The important point is that such things must be addressed explicityly. From its inception the major stumbling block of Small House was that though they had a set of goals, they had not established any priorities and could not generally conceive of a situation where these goals might conflict. As a result, the development of Small House often floundered while they worked through these things operationally.

The second major issue centers on the dichotomy between the required classes and the nonrequired classes. First, how can you classify an educational treatment that subsumes the gamut of treatments from laissez-faire to authoritarian, depending on the day and the hour, as well as the teacher? Second, is it fair to ask students to accept this environmental schizophrenia, or are there developmental prerequisities necessary for them to be capable of functioning in such a setting? For some students this seemed to be relatively easy; for others, impossible.

This latter group of students tended to view the teachers as being inconsistent — a valid conclusion from a certain perspective. Why was it permissible to move freely around the room talking at one time and not another? If the teachers wanted them to enjoy school, why were

they still expected to complete projects, demonstrate skills, and pass tests on occasions? Why did they have to attend an elective just because they signed up for it? Why couldn't they just go to the library or change their mind and go to a different elective or cut altogether? This group of students was small but significant. The fact that unlike most of the students who seemed able to grasp the nuances of the situation these students could not suggests that the ambiguity might be a little too much for some students.

As stated previously, the nonrequired classes were more than just non-traditional content taught in a traditional fashion. They were the heart of the Small House operation and constituted an exceedingly complex set of factors relative to the "treatment" as defined by that term. Replication of the Small House experiment would be largely contingent on the ability to reconstruct the flavor and atmosphere of the nonrequired classes.

E. ONE FINAL OBSERVATION

It would be possible to continue generating additional anecdotes, observations, and other material relevant to this research. The utility of so doing would be questionable. Nevertheless, there is one other piece of data which is tremendously significant. When the teachers announced that they were running short of operating funds in the Spring and would not be able to fulfill their plans, the parents of the students contributed between $1500 and $2000 to subsidize the Small House operation. This in a city which has defeated two millage elections in the last three years. In and of itself, this constitutes a penultimate evaluation of Small House.

CHAPTER VI

It might be highly relevant at this point in time to ask " So what?" What is the utility of collecting all of this information? How can it be synthesized and integrated? What does it mean? To attempt to answer these questions, it is necessary to reexamine the historical context of educational research, including the previous conceptions of curricular evaluation, the inadequacies of these models, and current developments in philosophical orientation.

The systematic study of education is not a new concept. Stevens (1967) summarizes 50 years of educational studies; the eight-year study, one of the most comprehensive examinations of education ever undertaken, was initiated more than 35 years ago. However,most of these studies were predicated upon the traditional psychological paradigm (with its underlying assumptions and methodologies) which, though historically venerated, was technically insufficient. Schutz (1969) puts it this way

> "Curriculum researchers have danced to the tune played by psychological research methodology. The lyrics of psychological research methodology have in turn been drawn for statistical analyses growing out of agronomy and biology. The disparity among the characteristics of agronomy, biology, psychology, and education was clearly noted by Glass (1968)."

Other authors have also addressed this same issue, though often from different perspectives:

> "The stability and predictability of rows of corn have seduced us into demanding (read assuming) stability and predictability in rows of children before we can examine and compare them properly. . . . We know next to nothing about the relationships between teachers attitudes, personality characteristics, and behavior. We have hardly touched on the relationships between teacher behaviors and learning in students. . . . These are difficult problems—complicated, hard to approach with traditional research methodologies, and well beyond the kind of problem we think we can deal with in research right now—but they are crucial issues in education.") (Allen, 1969)

> " . . . Too often, the pr occupation with satisfying the requirements of design and statistical models violates the instructional treatment and reduces the utility of the research to zero. Coversely, the lack of specificity of treatment has often made the application of elegant procedures a waste of time and money at best and a smokescreen at worst. The requirements for both treatment specification and design must be considered carefully to ensure useful research results." (Baker, p 340).

"Curriculum developers have tended to assume an isomorphism between their interpretation of curriculum guides and instructional materials and that of teachers and students. . . . However, word is leaking out that a number of cherished curriculum beliefs are figments of the researcher's theoretical imagination. Nongraded classrooms appear to be indistinguishable from graded classrooms (Goodlad,1968); self-instruction in schools appears to be indistinguishable from teacher-instruction in schools, and endemic gaps are being identified "between emotionally-toned accounts of the ideology and the day-to-day reality in the life of innovative schools and classrooms" (Smith and Keith, 1968, p 5). (Schutz, 359).

" . . . The data would suggest that there really is no such thing as a BSCS curriculum presentation in the schools. . . . Each teacher filters the materials through his own perceptions and to say that a student has been through the BSCS curriculum probably does not give as much specific information as the curriculum innovators might have hoped. (Gallagher,1966,p 33)

Traditional experimental psychology can justify the continued use of the agronomy-type of paradigm because, in their case, they can assume that the varaibles affecting the variable of interest are small in number, are being directly manipulated, or are controlled by random assignment of subjects. Additionally, they can specify the treatment so that it can be duplicated. This is generally not true foreducational research. Although we can specify variability across classrooms, among students, among teachers, and across socioeconomic levels (or any other dimension), we have no way at present of distinguishing relevant behaviors from irrelevant ones. In most cases, with educational research, we are dealing with nonrandom assignment of subjects, noncontrol (and often nonspecification) of treatment, and a large number of unknown variables, possibly having more unknown variables than subjects.

What is needed,then, is a viable alternative strategy for guiding and channeling educational research, particularly those activities subsumed under the rubrics curriculum evaluation and evaluation research.* We can no longer make simplisitic assumptions about treatments being equal except for the variable of interest; we can not make assumptions about random assignment of subjects or equality of treatment populations a priori; we cannot justify labelling two instances of a treatment program with the same label without concomitantly ascertaining whether they are in fact identical from every conceivable perspective! In a sense, we have to begin examining the schools as ecosystems with many components and evaluate results and programs with those specific components in mind. Our attitude here parallels that of Johnson (1970):

" . . . The question here is not just what a program accomplished, but what caused it to do so. At a minimum we need a detailed description of what the conditions were that yielded the given results. The practical problem is that of trying to describe what went into

---

* Cf Baker (1969) and Johnson (1970) for discussion of terminology and meaning.

the program in sufficient detail that the program can be effectively reproduced. True, some social action programs may not be conceived of as something to be "reproduced," but as single, one-shot applications. But for most purposes when we speak of a new program of reading instruction, or any pilot program of education, or social service, we are surely concerned vitally with the question of whether the program will have the same effect when reintroduced in a new setting. The reproducibility of the program, and the extent to which such reproducibility is inherently built into it, is therefore of crucial importance in interpreting the results of evaluative data on the outcomes produced by any program. (p 20)

The issues addressed on pages 28-30 specifically focus on this problem of reproducing the program in toto.

EVALUATION MODELS: THE SECOND GENERATION

Traditional models of evaluation are essentially obsolete. The newer models (systems models) are far more ecumenical about what should be considered data. Stake (1967) classified data into three broad classes which Glass (1972) elaborates on. Stake refers to antecedents, transactions, and outcomes. Transecting these three classifications are intents (the planned-for environment student behaviors, etc);observations (actual environment, student behaviors, etc); standards and judgments. A typical innovative program might involve observation of each of the following: *

Antecedents--

1. Pupil entry behavior and biographical data
2. Biographical and other data on the project personnel
3. Pupil and staff expectations
4. Parents expectations and wishes.
5. Description of the instructional environment.
6. Description of the instructional materials.
7. Record of financial resources.
8. The intended schedule and activities of the program
9. Comparative data on the pupils and teachers of non-innovative classrooms.

Transactions--

1. Actual schedule of activities
2. Time sampling studies of how typical classroom hours are spent.
3. The running record of the classroom (disturbances, significant deviations from intended activities, absences of staff and pupils, etc).

Outcomes

1. Pupil perofrmance and satisfactions both immediate and long-range.
2. Staff satisfactions and complaints both immediate and long-range

* Glass, 1972, p 107.

3. Parent satisfactions and complaints.
4. Pupil performance on cross-curricular obj ctives.
5. Staff mobility.
6. Financial costs.
7. Side effects of the program on nonparticipating pupils and staff.

Now, Glass continues to say "Anything that could conceivably be related to the success of the program becomes data for the evaluation" and that the only limitations under a systems model would be time and money (cf Scriven,1967, p 83). As Glass points up, Stakes model remains incomplete--there are a myriad of seemingly crucial dimensions ignored.

A far more elaborate model, and perhaps the most comprehensive put forth yet, was suggested by Hammond (1972). He attends to the anthropological and sociological (Sindell, 1969; Harp and Richer, 1969) aspects of education in much greater detail. His structure for evaluation (Hammond, p 232) is highly reminiscent of Guilford's structure-of-intellect model. Hammond's 3-dimensional block includes an instructional dimension, institutional dimension, and a behavioral dimension. Each of these is further broken down, as indicated below:

INSTRUCTIONAL DIMENSION

Organization
    Time
    Space--Vertical organization , Horizontal organization
Content
Methodology
    Teaching activities (16 categories)
    Types of interaction
    Learning theory (
Facilities
    Space, equipment, etc
Cost

INSTITUTIONAL DIMENSION

Student (age, grade, sex, family variables, socio-economic input, physical health, mental health,achievement, ability, interests, relationship to innovation)

Teacher (age sex, personaltiy characteristics, physical health, educational background and work experience, professional affiliations, nonprofessional affiliations, socioeconomic status of residence, degree of involvement in program, etc.)

Administrators (same as teacher)

Specialists (same as teacher)

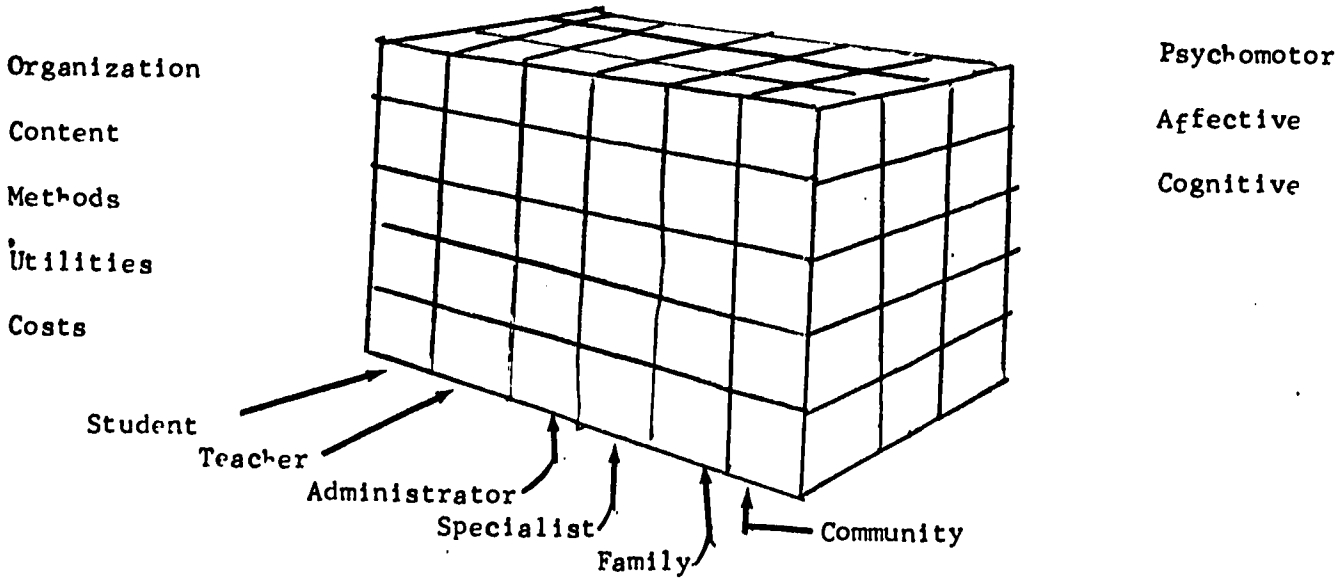


Figure 3 Hammonds Structure for Evaluation

Family (degree of involvement with innovation, degree of planning, general characteristics, size of family, socioeconomic status, pattern--nuclear or extended--income, residence, education of parents, siblings, mobility, social and professional affiliations,

Community (geographical setting, historical development, population characteristics, populations size, density, change patterns, growth patterns, economic characteristics, sources and ranges of individual incomes and tax bases, social characteristics, power structure, socioeconomic stratification

BEHAVIORAL DIMENSION

Psychomotor

Affective

Cognitive

In addition to this structure, Hammond (p 235) goes on to elaborate a process, taking the necessary step of establishing that this structure is only a means of permitting an analysis of the interaction of the forces and is independent of a process of evaluating.

Hammond's model makes explicit some potentially significant factors which could affect any attempt at educational innovation and which are ignored (or unassumed unimportant) by other strategies or models. I am not suggesting, nor is Hammond, that we change one inappropriate model for another; both must be a part of any concerted effort to evaluate the changes affecting and effected by an educational treatment or innovation.

There are some issues which are left implicit in the model of Hammond's which have been addressed by other authors. For example, Barth (1971) cautions against attempting to adopt an "open classroom" orientation unless the ideologies of those involved in the change are consistent with the change. It is my personal feeling that the failure of many attempts to innovate are a direct function of a nonphilosophical or aphilosophical orientation on the part of those who initiate those attempts. I have witnessed several cases where the "open classroom" process was simply legislated on to unsuspecting and often unwilling participants. This report itself falls short on failing to achieve some index of philosophical orientations of the different instructors.

From this on point (philosophical beliefs of teachers) we could generate an entire series of additional factors and interactions which could affect the innovative process. For example:

What happens if there is dissonnance between the philosophies of the teachers and those of administrators, particularly the principal?

What happens if there is dissonance between the philosophies of practitioners (teachers and principals) and thos of the school board (and/or community)?

Is a teacher's philosophy as measured on paper consistent with her behavior?

These questions can be added to those that are already present in the model developed by Hammond.

Is change possible in a highly transient community? If instituted, will it be maintained?

What is the relationship between the psychological environment of the home and a student's capacity to adapt in a particular school environment?

Is continuing effective change dependent upon a "strong man"?"

What are the relationships among teachers' personalities, administrators' personalities and effective continued change?

What kind of psychological environment are the teacher's working in" Were they instrumental in planning the change? Were they given the freedom (each and every one of them) to design their onw programs? Do they all feel that the principal and/or superintendent supports them regardeless of whether he agrees with them or not?

There is one far-reaching implication to all of this and to the question of an effective model for educational evaluation and research. Much of what was done in the past may, in fact, be invalid — useful, but invalid. Still another, and one of far greater import concerns the future. Specifically, the cost of adhering to such a model. On that issue, I can only echo Scriven (1967, p 83):

" . . . The educational profession is suffering from a completely inappropriate conception of the cost-scale for educational research. To develop a new rocket or automobile is a very expensive piece of business, despite the extreme confidence of the properties of physical materials. When we are dealing with a teaching instrument, such as a new curriculum or classroom procedure, with its extreme dependence upon extremely variable operators and recipients, we must expect considerably more expense. The social payoff is enormously more important . . . The educational profession as a whole has a primary obligation to recognize the difficulty of good curriculum development with its essential concomitant, evaluation, and to begin a unified

attack on the problem of financing the kind of improvement that may help us towards the goal of a few million enlightened citizens on the earth's surface, even at the expense of one on the surface of Mars.

In addition to the financial implications, Hammond's model has tremendous implications for methodologies. The simplistic, alleged objective type of evaluation utilizing paper and pencil only, is gone. Such a model demands a multimethod approach, particularly one which utilizes naturalistic and systematic observation and other clinical techniques.

CHAPTER VII

What we essentially set out to do was evaluate one program initiated at a junior high school, covering a period of one academic year; additionally, it was hoped that the inadequacy of existing approaches to evaluation would be demonstrated by this technique. There are several key points concerning educational evaluation which must be addressed; some of these we have touched on briefly in the preceding pages, others will be mentioned for the first time.

a. If it is at all possible, it is reasonable to attempt to achieve the same degree of control as psychological experiments via random assignment, etc. If the groups are already established, then the "best" that we can possibly do is attempt to demonstrate that--as best as our tests (observations, etc) can measure--the groups are equal along certain dimensions. This is what happened in this particular research. However, if the groups differ significantly in any respect, this must be included. For example, in this research, all of the Small House students had parents who were willing to volunteer; there is no way of equating for this "home climate" in any reasonable control group.

b. Evaluation studies should cover a period longer than one year. It is impossible to rely on a determination based on one-year trial simply because it virtually takes one year to establish a program, iron out the bugs, and get rid of the halo effects.

c. There are at least two units of analysis in educational studies--the class, and the individuals. To ignore either one is to jeopardize the value of the study. Particular attention should be paid to aptitude-treatment interactions (where aptitude is broadly defined) (Messick,1972; Bracht, 1969).

d. The kind of data gathered and the data-gathering processes must both be expanded far from what they have been in the past.

e. Perhaps the greatest task confronting educational research and evaluation is to determine which variables are irrelevant to the educational process and which are relevant and what interactions exist among the relevant ones.

f. Evaluation studies may require combining criteria, judgments, and costs in a complex fashion. For example, the distinction was made in this paper between estimation of goal achievement and evaluation, the former being merely one part of the latter. Or, suppose that the students had demonstrated a significantly more positive attitude but their gains in reading and math had been significantly less than the control group? These are the kinds of issues which must be integrated into evaluation research.

Medicine as a science was capable of developing only after it had broken free of the influence of Galen. While he admittedly did much to advance the cause of medicine, Galen eventually became a stumbling block to its continued progress. Educational research and evaluation is in a position similar to that of Medicine; the benefits that have accrued from the Psychological model have wedded that model to education and precluded further development. Unless there is a break with that simplistic model, educational research and evaluation will become fruitless excercises and hypertrophied ritualism.