DOCUMENT RESUME

ED 082 834 PS 006 849

AUTHOR Lukas, Carol Van Deusen; Wohlleb, Cynthia
TITLE Implementation of Head Start Planned Variation: 1970-71. Part I.
INSTITUTION Huron Inst., Cambridge, Mass.
SPONS AGENCY Office of Child Development (DHEW), Washington, D.C.
REPORT NO OCD-H-1926
PUB DATE Aug 73
NOTE 283p.; PS 006 850 is Part II (Appendixes) of this report

EDRS PRICE MF-$0.65 HC-$9.87
DESCRIPTORS Classroom Observation Techniques; Data Analysis; Measurement; *Models; *Preschool Children; *Preschool Programs; *Program Design; Program Effectiveness; *Program Evaluation; Site Analysis; Staff Role; Training
IDENTIFIERS Planned Variation; *Project Head Start

ABSTRACT

This report is Part I of a study of program implementation in 12 models of Head Start Planned Variation. Chapters examine (1) how well models are implemented, using sponsor ratings of teachers as the data source; (2) the factors which may influence the extent to which models are implemented, such as sponsor input, staff reaction and input, and the context in which implementation is undertaken; (3) the correlations between the factors which explain variations in levels of implementation; and (4) what the model classrooms are like in practice, using classroom observation data. Included are a summary of findings, recommendations for future implementation studies, and data tables. (SET)

IMPLEMENTATION OF HEAD START PLANNED VARIATION: 1970-71

CAROL VAN DEUSEN LUKAS
CYNTHIA WOHLLEB

PART I

August 1973

HURON INSTITUTE
CAMBRIDGE, MASSACHUSETTS

This document was prepared for Grant # H 1926 from the Office of Child Development, Department of Health, Education and Welfare, U. S. Government. The conclusions and recommendations in this report are those of the grantee and do not necessarily reflect the views of any federal agency.

The project director was Marshall S. Smith.

Project staff included:

Mary Jo Bane
Barbara Behrendt
Anthony Bryk
John Butler
Tom Cerva
David Cohen
Jane David
Richard Elmore
Helen Featherstone
Nathan Fox
David Gordon
Deborah Gordon
Sharon Hauck
Gregory Jackson
Robert McMeekin
Anne Monaghan
David Napior
Ann Taylor
Herbert Weisberg
Jack Wiggins
Cicero Wilson
Joy Wolfe
Stanley Yutkins
Diane Zipperman

PREFACE

This report summarizes a study of program implementation in the second year of Head Start Planned Variation. The primary questions to be answered are: to what extent are the experimental treatments actually present in the Head Start classrooms, and what factors affect the success of treatment implementation? The report not only presents the results of our data analyses, but also makes recommendations for future implementation studies. The body of the report is contained in Part I. Part II is a volume of appendices which present the details of the analyses used in Part I, copies of data collection instruments, and some additional analyses and statistics.

We are indebted to the people at the Stanford Research Institute who designed the instruments and collected the data which formed the basis for this report. We would also like to thank the OCD staff, particularly Thelma Zener and Lois-ellin Datta, for their assistance.

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

Chapter 1

OVERVIEW

GOALS AND DESIGN OF PLANNED VARIATION

Head Start Planned Variation is conceived as an experiment undertaken to investigate "the impact of various well-defined educational environments and learning situations on the Head Start child."[1]

The design of the study involves implementing preschool curricula, or models as they will be called here, in a number of existing Head Start sites across the country. In 1970-71, there were 12 models and 37 sites in Planned Variation. For 15 sites in 8 models, this was the second year of participation; for the other sites, it was the first year. Table 1 displays this information in detail.

Ten of the twelve models that participated in 1970-71 are pre-developed programs -- the nature and structure of the program are determined before it is brought to a community -- which focus primarily on educating children in the classroom. The models vary on a number of dimensions such as the amount of structure in the classroom or the importance of academic activities. The eleventh model, the Enablers, has a classroom

[1] Office of Child Development. Head Start Planned Variation Study, November, 1971. Washington: USHEW, OCD, 1971 p.1.

TABLE 1

Head Start Planned Variation Sites: 1969-72

	1969-70	1970-71	1971-72
Far West	Duluth	Duluth Fresno Tacoma Salt Lake Buffalo	Duluth Salt Lake Tacoma Buffalo
Arizona	LaFayette Lakewood	LaFayette Lakewood Lincoln	LaFayette Lakewood Lincoln
Bank Street	Tuskegee Wilmington	Tuskegee Wilmington Boulder Elmira	Tuskegee Wilmington Boulder Elmira
Oregon	Tupelo E. St. Louis	Tupelo E. St. Louis E. Las Vegas	Tupelo E. Las Vegas
Kansas	Oraibi Portageville	Oraibi Portageville Mounds	Oraibi Portageville Mounds
High Scope	Ft. Walton Central Oz.	Ft. Walton Central Oz. Greeley Seattle	Ft. Walton Central Oz. Greeley Seattle
Florida	Jacksonville Chattanooga	Jacksonville Chattanooga Jonesboro Houston	Jacksonville Chattanooga Jonesboro Houston
EDC	Washington Johnston Co.	Washington Paterson Johnston Co.	Washington Paterson Johnston Co.
Pittsburgh		Lock Haven	Lock Haven Montevideo
REC		Kansas City	Kansas City
NYU		St. Thomas	St. Thomas
Enablers		Billings Colorado Sp. Bellows Falls Newburgh Puerto Rico	Billings Colorado Sp. Bellows Falls Newburgh Puerto Rico

focus but is not pre-developed. It emphasizes the community's development of its own goals and curricula with the help of an early childhood expert who visits regularly. The twelfth model in Planned Variation differs from the others in that it is based on the premise that the parent is the primary educator of the child. Although all Planned Variation models include a parent component in some way, this program, the Florida model, is the only one which is actually a parent educator model and does not specify what should go on in the child's school environment.[2] For simplicity of presentation, we often refer to all models in terms of classroom programs, recognizing that this is not appropriate for the Florida model.

The use of models in Planned Variation is associated with the principle of sponsorship. The originator, (or sponsor) of a model is responsible for delivering his program to a site. The sponsor and his staff not only provide manuals and materials, but also introduce the model to the local people and provide training and continuing support in using it. In this report we distinguish between 'model' and 'sponsor', using model to

---

[2] A brief description of the Planned Variation models is found in a seperate report entitled: Smith, M., "Some Short Term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation -- 1970-71". The Huron Institute, Cambridge, Massachusetts. 1973.

refer to the curriculum or educational program[3] and sponsor to mean the person or people who deliver that model.

Each of the Planned Variation models works with one to five sites. A site is a Head Start center or group of centers usually located in the same community and administered by a single agency. Within each site the number of classes working with a model varies, and in most cases it is less than the total number of classes in that site. In 14 sites, the classes not working with the model are designated as comparison classes which are to be used as controls in testing the effects of the model treatments.[4] All Planned Variation classes within a site work with the same model. The site is formally responsible to the sponsor for only the areas included in his model. The site is usually independent in administrative practices and operation of non-Planned Variation classes.

In each Planned Variation class the staff is expected to organize its teaching to conform with model prescriptions

---

[3] There is some question, now, as to what should be included in the term 'model'. Originally, it seemed to be limited to the program for the children; from more current descriptions it appears that some models -- EDC, for example -- also consider their system of advisors and training procedure as part of their model. This changing definition will be discussed later in this chapter.

[4] In addition to the comparison classes which are housed in the same sites as the Planned Variation classes, there are comparison classes in 6 other communities. These were selected when a Planned Variation site did not have enough classes to form a comparison group. They are intended to match the PV sites on relevant demographic dimensions.

or principles. For the great majority of teachers, aides, and volunteers, the model requirements differ from their usual modes of teaching. In varying degrees for different teachers, then, working with a model requires that a Head Start staff acquire new skills and techniques in the class, new ways of relating to children, and in some cases, new values.

IMPLEMENTATION ISSUES

The planned variation approach is characterized by its use of experimental methodology. This methodology, which can be described in terms of conventions or requirements, enables the researcher to create a situation in which the effects of specified variables, or treatments, are tested while the operation of other possibly confounding factors is minimized. The advantage of the experimental approach is that it provides the basis for making inferences about the causes of the effects that are found.

In this report, we will not examine all requirements necessary to a valid experiment, but only two related to treatments. First, the treatments should be well-specified. In the case of HSPV, this ideally means that models should be clearly defined; we should have a reasonably detailed knowledge of what should go on in a classroom during the day with each model. It also means that the treatments designated provide complete descriptions of factors which may affect

outcomes. Second, the treatments should be fully implemented. If implementation is defined as the extent to which classroom practice matches model theory, then full implementation can be taken to mean that all or most aspects of the model are present.

These requirements are important for both the internal and the external validity of the experiment. Internal validity, as Campbell and Stanley put it, is a question of: "Did in fact the experimental treatments make a difference in this specific experimental instance?"[5] If the treatments are not present, we cannot answer this question affirmatively and as a consequence, cannot interpret the results of the study.[6] The treatment requirements can also be seen as issues of external validity. Since a social experiment is intended to inform policy, it is important to be able to generalize its findings. Bracht and Glass state that one aspect of external validity is explicitly described treatments: "Generalization and replication of the experimental results presuppose a complete knowledge of all aspects of the treatment

---

[5] Campbell, S. and Stanley, J. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally and Company, 1963, p. 5.

[6] The presence of the treatments is not the only requirement for internal validity. Random selection and assignment of subjects, and the minimizing of possibly confounding variables are also important but are not within the scope of this report.

and experimental setting."[7] Adherence to the treatment requirements are necessary from a research perspective because they provide the basis for a valid study and thereby enable us to interpret and generalize the stud 's findings.

Meeting Experimental Requirements:

After two years, however, it is apparent that the treatment requirements are not met in Planned Variation.

The specification of treatments: The requirement of specified treatments has two dimensions: the program models should be clearly defined in operational terms, and the treatments as designated should include all factors which might affect outcomes. With respect to the first dimension, the HSPV treatments deviate for several reasons. First, not all sponsors felt that their models were fully developed at the beginning of the study. With the exception of two models, sponsors were unable to provide written descriptions of what happens in the classrooms. As a result, the treatments in most cases evolved and changed during the course of the experiment. While such evolution is educationally desirable if the models are incomplete or inappropriate, it complicates the treatment from the researcher's point of view because it is more difficult to define a treatment if it is constantly changing, and because a model may evolve to different forms in different sites, in which case sites may become separate

[7] Bracht, G. and Glass, G. The external validity of experiments. American Educational Research Journal, 1968, 5, 437-474, p. 438.

treatments in themselves rather than replications of the same treatment.

Second, the requirement of clearly-defined models is not met because some models were not complete and well-defined even at the end of the study. During the third year, a member of the Huron Institute staff developed checklists intended to detail the components of all models.[8] Though refined through interactions with the sponsors, many of these checklists are still only partial descriptions of a classroom program. This is a function of the state of the art of educational theory, and does not result from poor planning or a failure to take the experiment seriously. The models included in HSPV reportedly were the best available approaches to early childhood education when the study began. Developmental theory, however, does not yet provide an adequate basis for determining what factors affect a child's growth and learning. Although HSPV models take their focus from the different schools of thought in the psychology of learning -- behaviorist theory, Piagetian cognitive growth, discovery learning, open education, total child development -- these theories (with the possible exception of those from the behaviorist tradition) do not seem to be adequate yet for fully specifying and integrating an educational program. Even if these theories were adequately reflected in educational models, it is not known whether the differences among models would be differences which are important to children's learning and

[8]Monaghan, A. "Patterns of Implementation in Head Start Planned Variation," The Huron Institute, Cambridge, Massachusetts, forthcoming.

and behavior. Further evidence of the incomplete state of this field is that there is no agreed upon scheme for classifying the variation in the approaches included in this experiment. It is not clear that the models in HSPV represent a systematic variation of all relevant dimensions.

Third, some models are less easily dealt with in an experimental situation than others. We argued above that some models do not meet the criterion of a well-defined treatment because of an inadequate theoretical base. But even if all models were complete in the educational sense of having a clear and integrated theory of how to teach children, some of the sponsors still would not be able to provide detailed, operational descriptions of day-to-day classroom activity because their philosophy runs counter to this approach. These models, which set out principles and encourage individual adaptation in implementing them, are less amenable to conventional experimental requirements than are models which are defined in detailed operational terms.

We conclude, then, that the models in HSPV are not experimentally well-defined because they have changed during the course of the experiment, because the theories on which they are based are not always fully adequate for describing an educational program, and because not all the models are defined in operational terms.

A second aspect of the experimental requirement that treatments be well-specified is that all components of the

treatment are identified. In HSPV, previous descriptions of the treatments have been limited to descriptions of curricular programs for children. In practice, however, the treatment associated with a model is more than this. One additional dimension is training: a sponsor is responsible for introducing the program to the local staff and training them in its use. Analyses in Chapter 3 (Tables 8 through 19) indicate that there are systematic differences among both models and sites within models in the extent and types of training received by the teachers. Moreover, anecdotal evidence indicates that sponsors often participate in a broad range of activities which cannot be subsumed under either staff training or the program for children. In one case, for example, a sponsor sent three staff members to a site for three days to hold a retreat dealing with staff problems stemming from a controversy over the ethnicity and competence of a newly appointed teacher. In another site, a sponsor was called on by the teacher aides to assume the role of a child advocate when a teacher struck a child. These are important aspects of the treatments which may affect outcomes and therefore, cannot be ignored. Their omission from the model descriptions indicates that the requirement of well-specified treatments is not met.

Thus, from the experimental design point of view, the treatments in HSPV are not well-specified, either in terms of the curriculum models being well-developed or in terms of all relevant factors being included in the treatment description.

The implementation of treatments: From the earlier definition of implementation as the extent to which classroom practice matches model theory, a high level of, or full, implementation can be taken to mean that all or most aspects of the model are present. Although it was expected that full implementation would be reached by all models in HSPV,[9] available data suggest that this expectation was not met.

Tables 4, 5, and 6 (pp. 38, 40, 44) in Chapter 2 show the analyses of our only direct source of information on how well the models are being implemented: the sponsors' ratings of teachers' performance in the models. The findings, in those analyses, of variation in the extent of implementation demonstrate that the treatments are not fully present in all classes and that classes within models cannot necessarily be taken as replications of a single treatment. Moreover, the level of implementation does not improve consistently over time as was expected at the beginning of the study. These analyses are examined in detail in Chapter 2. The point to be made here is that the requirement of fully implemented treatments is not met.

Thus, after two years, it is apparent that the treatments in Planned Variation deviate from conventional experimental treatment requirements. Although the lack of treatment specification and the failure to reach full implementation are understandable given the practical constraints under which the study has been operating -- e.g., beginning without fully developed models, attempting to transfer models

[9] See, for example, Office of Child Development, p. 3.

to locations distant from the sponsors, working with operating sites which face a variety of demands other than Planned Variation -- they are problematic from a research point of view. Because these deviations confuse our interpretation of the causes of treatment effects, we must deal with them if we want to draw meaningful conclusions from HSPV. Our study of implementation, then, is an attempt to do this.

The Study of Implementation

In the early stages of Planned Variation, the study of implementation was directed toward two questions: Are the treatments present, and what are the most effective delivery systems for establishing the models in the sites? Since it was assumed that all models would eventually reach full implementation, these were seen as essentially straightforward issues.

As we have demonstrated, however, it has become clear as Planned Variation has progressed, that the notion of implementing comprehensive program models is extremely complex. Because of the deviations from experimental requirements which result from this complexity, the two original questions must be rephrased. The first shifts from a simple issue of checking on the presence of treatments to the question of what are the treatments if they are not the models as originally described? The second question is not simply an issue of describing delivery systems -- although that is still important -- but must be broadened to also ask why is there variation within models in the extent to which they are implemented?

What are the treatments? The question of what the treatments are has two parts: what is actually happening in the classroom, and to what extent does classroom practice match model theory?

The first part, the determination of what is happening in the classroom, is important for learning what "treatments" the children are in fact receiving. This issue is explored in Chapter 5 of this report using data obtained with a classroom observation instrument developed by the Stanford Research Institute. This instrument is designed to be used with all models and, therefore, provides general information. Model-specific observation checklists were developed by the Huron Institute for use during the 1971-72 school year. The data from these checklists are not available for the 1970-71 analysis, but are presented in a separate report for 1971-72.[10]

The question of what is going on must also be directed to an examination of variables outside the classroom. We have suggested that the non-model activities by the sponsor and other people may affect the treatment. Both of these areas are considered -- albeit inadequately -- in Chapter 3.

The second question, how well the models are implemented, involves a comparison of "what is happening" with a standard. A crucial issue here is defining the standard -- defining full implementation. We all assumed that we knew what we meant in using this term, but no definition was ever made explicit. In a conventional experiment we would define it.

---

[10] See Monaghan

in terms of replication: we would expect the treatment to be identical in all situations in which it was tested. In Planned Variation, where we are working with comprehensive treatments in "real life" situations, this standard seems unrealistic. The important factor in defining an alternative standard is allowing some variation among classes: we should accomodate the fact that a model may be adapted, by the sponsor and the teaching staff, to the unique situation in a site, or even to a class within a site, and to changes in a situation over time. Even in those models which do not emphasize adaptation, the classes are not expected to be identical. We need, however, to set limits on the variation: we must determine how much variation can exist while still considering a class to be a fully implemented program. We must be able to decide at which point the variation is so great that full implementation becomes partial, and the point at which partial implementation becomes nothing -- no evidence of the model at all. As yet no one has explicitly set out the limits of variation for all models, and thus there is no specific operational definition of full implementation. This is not to say that sponsors cannot identify a well-implemented class, only that they have not systematized, or at least communicated, their criteria. In the absence of agreement on an alternative definition of full implementation, then we continue in this report to define it in terms of replication of the model although we recognize the inadequacy of this approach.

Other problems arise in attempting to measure the extent to which a model is implemented. Examination of the sponsor ratings of implementation, to be discussed in Chapter 2, illustrate the difficulty, for example, of comparing levels of partial implementation. Comparison across models is difficult because the scales are not anchored to any common standard and because models have different components and are at different levels of operationalization. As a result, we do not know whether a rating of, say, 5 (on a 9-point scale) has the same meaning in different models. The problem of comparison also exists to some extent within models: one class which is rated 5 may not look like another with the same rating because different parts of the model may be present in each. When this is true, it is not clear that we have replications of a single treatment.

We raise these issues in an attempt to demonstrate that the question of how well an innovation is implemented is difficult as well as important. This is the case not simply because of inadequate measures but also because of conceptual ambiguities. We have not resolved these difficulties in this report but point them out for consideration in future studies.

Why is there variation?

Because of variation in levels of implementation, we are interested in identifying the factors which affect implementation not simply from the point of view of describing

the process, but from the perspective of explaining outcomes: we want to explore the question of why implementation is more successful in one classroom than in another.

One group of variables which no doubt affects implementation is sponsor input. We expect that this category includes not only factors such as type and amount of training in the model, or feedback and continuing support, but also the sponsor's staffing patterns and the role he assumes. More important, we propose that the process of implementation depends on more than the sponsors' delivery systems: implementation is not determined by sponsor input alone, but by the sponsor in interaction with the local staff working with the model, and with the context in which the implementation takes place. We contend that the involved staff members are not passive recipients of the model: they are not equally capable and equally desirous of working with the models. We believe that implementation is influenced by a staff member's reaction to the model, to the sponsor's staff, and by his or her past experience and training. Moreover, we expect that sponsor and local staff inputs are influenced by the context in which they interact. We propose that the site administration and the priorities of the administrators, as well as characteristics of the site, such as its size and location, affect implementation. These categories will be presented in more detail in Chapter 3.

Inputs from people other than the sponsor's staff may

also influence implementation. This issue is discussed both in relation to the sponsor's input and within the framework of administrative priorities. We predict that inputs which affect the classroom program, but do not reflect the model, occur when the Planned Variation experiment has low priority among an administrator's concerns.[11]

Summary of Implementation Issues

Traditionally, two treatment requirements are important to a valid experiment: the treatments should be well-specified and they should be fully implemented. In Planned Variation, the findings of incompletely specified treatments and variation in levels of implementation indicate that the

---

[11] Another assumption of the original conceptions of Planned Variation as an experiment was that the comparison classes do not work with the model. The comparison classes are intended to provide a basis for conclusions about model effectiveness by showing how well similar children do without the model. There is evidence, however, that in some sites, comparison classes, as well as Planned Variation classes, work with the model. Examples of this problem are numerous. One OCD consultant reports, for example, that when "asked if the $350 allocated for materials for next year would only be for the 8 classes in PV, (the local sponsor representative) said it's physically impossible and all classes will use the materials." And another consultant states that "the conscientious teachers are trying to get as much as possible of (the model) for their children and themselves, whether they are technically included in the special program or not. 'Contamination' in the program is great." It is clear that, because of such contamination, comparison classes cannot provide appropriate tests for model effectiveness in some sites. Moreover, providing resources and training for all classes in a site rather than for only Planned Variation suggests that model inputs are being diluted.

requirements are not met. Our study of implementation is an attempt to understand and to deal with these deviations. This means, first, that we cannot simply check on the presence of the treatments, but must ask what the treatments are. This question involves both the determination of what is happening and the comparison of "what is happening" with a standard. Important to the latter category are the problems of establishing a definition of full implementation which allows for and yet sets limits on variation among classes, and of comparing levels of partial implementation. The need to anchor rating scales is relevant to both issues. Although we do not resolve these concerns in this report, they should be dealt with in future studies.

The task of describing delivery systems similarly should be expanded to include the more general issue of identifying factors which influence implementation. The goal of this undertaking is to attempt to learn why the level of implementation is a complex process dependent not only on the sponsor's input, but also on the staff's reaction, and the context in which implementation occurs. If this proposal is proven true, then we can conclude that the variation in levels of implementation results not from a poorly planned experiment but from the nature of this type of study. While this report contains analyses directed to this point, data are not available for all factors we think important.

DATA ANALYSES

Data Sources:

The data used in this report are drawn from a number of sources. Six instruments designed and administered by the Stanford Research Institute are included in the analyses:

Sponsor Ratings of Teachers
Teacher Questionnaire
Teacher Aide/Assistant Questionnaire
Sponsor Implementation Report
Final Consultant Report
Classroom Observation Instrument

Copies of these instruments are included in Appendix B.[12] The first three instruments above provide information which can be related to individual classrooms, while the Sponsor Implementation Report and the Final Consultant Report provide data for the site as a whole. The Classroom Observation Instrument is the only source of data which systematically deals with people and activities within the classroom. In addition to the SRI instruments, we have obtained anecdotal information about implementation from conversations with sponsors and from the narrative reports written by OCD consultants. The OCD consultants are early childhood education experts who were hired by the Office of Child Development to monitor model implementation by visiting Planned Variation sites regularly The same consultant always visits the same site.

Although a great deal of data was collected through these instruments, not all of it is included in this report. The primary reason for excluding items was their

[12] For descriptions of the administration of these instruments, the reader should consult "Implementation of Head Start Planned Variation Testing and Data Collection Effort, Final Report", by the Stanford Research Institute, Menlo Park, California.

lack of relevance to the implementation questions which have been raised here. Another important reason for excluding items was that some questions were poorly conceived. Consequently, the answers are ambiguous and cannot be interpreted. For example, question 19 of the Teacher Questionnaire asks teachers to outline the techniques they would use when pupils have difficulty in seven types of situations. The question is open-ended. The situations are vague and difficult to deal with in the abstract. Not surprisingly, the answers are also vague and varied, to such an extent that it would be extremely difficult to summarize and interpret the responses from approximately 300 teachers. This question was not included in our analysis. There are other ambiguous items which were included, however, because of the lack of alternatives. For the most part, problems with individual items will be dealt with as they are presented in the text. One example serves to illustrate the point here: all questions on training in the Teacher Questionnaire refer only to training in general and do not specifically ask about training in the model. If, as we suspect, training other than model training occurs in Planned Variation sites, the failure to distinguish between them severely limits the conclusions which can be drawn in relation to implementation. Thus, not all the data collected are included in this report, and there are limitations to some data which are included.

Moreover, in other areas which we believe to be important to the study of implementation there are no data.

Analysis Strategies:

The major portion of the analysis in this report involves examining a large number of individual variables or single items taken from the questionnaires. In most cases, the class is the basic unit of analysis, although in some instances the site is the basic unit. For each variable, the means and standard deviations are presented to give a picture of the content of the information gathered on implementation. In addition, analyses of variance were performed for each variable to learn if there are significant differences among models, among sites within models, or between PV and NPV classes. Since all the analyses except sponsor ratings of teachers (Chapter 2) are intended to explain differences in models, only variables for which there are significant differences are included. Variables which show no significant differences on any factors are not included.

In selecting the most appropriate analyses and in using the results, a number of factors must be considered. First, the analyses require a hierarchical design because sites are nested within models in the sense that a site works with only one model, or treatment, rather than all of them. Nesting influences the analyses which are done and the conclusions which can be drawn. In this case the effects of a site cannot

be separated from the effects of the site in interaction with the model. As a result, we must always speak of "sites within models".

Second, models and sites are both considered to be fixed rather than random factors.[13] Although models are unquestionably fixed factors in this study, the nature of the site factor is not clear. Initially, sites were considered to be random. This is logical in view of the goal of assessing the effectiveness of models which implies that findings will be generalized beyond the Planned Variation sample. If sites are fixed, no generalizations can be made and the study is better conceived as a series of descriptive case studies than as an experiment. There are problems with sites as a random factor, however. First, sites were not randomly selected and assigned to models. Instead, the Head Start programs in certain communities, chosen on the basis of participation in Follow Through, geographical location, and recommendations from a variety of sources, were given the choice

---

[13] A random factor is one in which the intervals or groups included in the experiment are taken to be representative of a larger population; conclusions drawn about a random factor in the experiment are generalizable to the total population from which it is drawn. Subjects in an experiment is an example of a random factor. In contrast, a fixed factor is one from which conclusions are not generalizable. The levels or divisions of a fixed factor are of interest in themselves and are not considered to be samples of a larger population. Treatments in psychological experiments are good examples of fixed factors.

of working with the same model with which the Follow Through program in the community was working, or not working with a model at all. On the basis of selection, then, sites are not random. At the same time, it could be argued that the sites in Planned Variation do represent the range of Head Start centers found in the country, and on this basis can be considered random in spite of the non-random selection. A second, more compelling argument against using sites as a random factor is a statistical one. In this study, when sites are random, the F ratio for model effects is formed by comparing the mean square for models with the mean square for sites. With only one to five sites per model, however, there is not enough statistical power in this ratio to find significant model effects. Tests on model effects then, may be overly conservative and we may tend to overlook real differences (Type II error). With sites as a fixed factor, the classroom rather than the site is used to test model effects. The results in this case are probably too liberal: significant effects may be recorded for differences which can really be attributed to chance. On the basis of this analytical difference, we have chosen to use sites as a fixed factor because our strategy is to report as many effects as possible. It must be remembered, however, that these tests may report chance differences. It should also be recognized that the decision to use sites as a fixed factor was post hoc and was

made for statistical rather than experimental reasons. The decision should not be interpreted as strong conviction that sites are in fact fixed factors.

The third factor which must be considered in analyzing the implementation data is that we must deal with an unbalanced design. There are both unequal numbers of classes within sites and unequal numbers of sites within models. This is unimportant as far as the basic statistics, correlations, and regressions included in this report are concerned. With the analysis of variance, however, it must be considered. An unbalanced design is problematic because the unequal cell sizes cause the effects to be confounded. As a result, the F ratios are biased and the results are difficult to interpret. The problem of unequal sites within models has been dealt with here by artificially creating a balanced design after the data was collected. Although this excludes some sites and models from the analysis, the loss of the data is seen to be less of a problem than the confounding effects. Table 2 shows the sites included in balanced designs for three sets of analysis. The criteria for establishing each of these designs can be found in Appendix A. The design for the analysis of the Teacher and Aide Questionnaires will be referred to as the "standard design" because it includes the greatest number of sites and because it is used most often.

TABLE 2

Models and Sites to be Included in Implementation Analyses of:

Model	Site	Sponsor Ratings	Teacher and Aide Questionnaires: PV Only*	Teacher and Aide Questionnaires: PV-NPV Compar.
Far West	Buffalo	0	0	
	Duluth	0	0	
	Fresno			
	Salt Lake			
	Tacoma	0	0	
Arizona	LaFayette		0	
	Lakewood		0	
	Lincoln		0	
Bank Street	Boulder	0	0	
	Tuskegee			
	Wilmington	0	0	0
	Elmira	0	0	0
Oregon	E. St. Louis	0	0	0
	Tupelo	0	0	0
	E. Las Vegas	0	0	
Kansas	Oraibi		0	
	Portageville		0	0
	Mounds		0	0
High Scope	Ft. Walton B.	0	0	0
	Central Ozarks			
	Greeley	0	0	0
	Seattle	0	0	
Florida	Jacksonville			
	Jonesboro	0	0	0
	Chattanooga	0	0	0
	Houston	0	0	
EDC	Washington		0	0
	Paterson		0	
	Johnston Co.		0	0
Pittsburgh	Lock Haven			
REC	Kansas City			
N.Y.U.	St. Thomas			
Enablers	Billings	0	0	
	Colorado Sp.	0	0	
	Bellows Falls	0	0	
	Newburgh			
	Puerto Rico			
# of Models		6	9	6
# of Sites		18	27	12

*The models and sites in this column will be referred to as the "standard design."

The problem of unequal numbers of classes within sites is dealt with by using an approximate unbalanced analysis: an unweighted means analysis.[14] The unweighted means analysis uses the harmonic mean of the cell sizes to approximate equal n's. A criterion for using this analysis is that cell sizes be approximately the same. The results can be interpreted as if cell sizes were equal and the effects were independent.[15]

Hence the analysis of the implementation data is first determined by the original design of the study where sites are nested within models. The analysis is further determined by decisions made after the data was collected: sites are considered to be a fixed factor; balanced designs are created where necessary by excluding some sites from the analyses; and the analysis of unbalanced data (within sites) is approximated by an unweighted means analysis of variance.

---

[14] Searle, S.R. Linear Models New York: John Wiley and Sons. 1971, p. 365.

[15] An alternative to the unweighted means analysis is the least squares analysis. This is an exact rather than an approximate analysis of unbalanced data. The results of the least squares solution, however, are difficult to interpret because the effects are not independent. Consequently, the order in which the factors, in this case sites and models, are entered, changes the sums of squares which in turn may change the findings about the significance of the effects. Since we have no hypothesis about order, we will rely primarily on the unweighted means analysis as a good approximation of the unbalanced data analysis. Some least squares analyses are used in Chapter 7. They will be more fully discussed there.

Analyses will be fully described as they are presented in this report.

ORGANIZATION OF THE REPORT

The body of the report is organized within the framework laid out in this introductory chapter. Chapter 2 deals with the question of how well the models are implemented, using the sponsor ratings of teachers as the data source. Chapter 3 explores the factors which may influence the extent to which models are implemented. The variables are discussed both as descriptors of what is involved in working with a model and as potential explanators of variations in levels of implementation. The chapter is divided into three groups of variables: sponsor input, staff reaction and input, and the context in which implementation is undertaken. Within each of these sections, we first discuss the relevant data from the teacher, consultant, and sponsor questionnaries, and second, evaluate the data in terms of identifying factors which account for variation in implementation. Chapter 4 presents the correlations between the variables introduced in Chapter 3 and enters the variables into regression equations in an attempt to explain levels of implementation (sponsor's ratings of teachers). In Chapter 5, the question of what the model classes are like in practice is explored through the classroom observation data. Finally, Chapter 6 summarizes the report and draws conclusions.

Chapter 2

LEVELS OF IMPLEMENTATION

One of the primary questions to which a discussion of implementation must be addressed is the extent to which classroom practice matches model theory, or in other words, how well a model is implemented. If all aspects of a model are present in a class in the manner prescribed, then the model can be said to be fully implemented. If some, but not all, aspects of the model are present, then the model is partially implemented. Similarly, if no components are present, then the model is not implemented. These degrees of match between theory and practice will be called "levels of implementation". Thus, a high level of implementation, for example, means that all or most aspects of the model are present.[1]

The determination of the levels of implementation in this study is important for two reasons. First, the requirement of testing the effects of specified treatments, or in

---

[1]No explicit definitions of implementation were given in any of the instruments to be used in this report. We have derived the definitions used here from our knowledge of what was assumed about the nature of the study and about how the models would be implemented. We spoke earlier of the difficulty of defining "full" implementation, particularly in relation to allowing and yet setting limits on variation among classrooms. These difficulties have not been resolved and we will use the definition of full implementation as duplicating the model here, however, because it is the one used most commonly.

this case, well-defined preschool models is basic to the conception of Planned Variation as an experiment. Unless we know whether the experimental classes do in fact approximate the models, we cannot draw useful conclusions about the effects of those models. In this sense, the question of levels of implementation is one of checking for the presence of the specified treatments. Second, implementation is a question of service delivery: can pre-established programs be distributed and instituted under a wide range of conditions? In general, the models included in Planned Variation were developed and tested in controlled, closely supervised situations, usually laboratory schools. In Planned Variation, however, the sponsors must establish their models in the complex real world. Sponsors cannot supervise implementation as closely as when developing the models, because the sites are often far from their home bases. Thus, the question of whether these models can be "delivered" is a crucial one.

The only direct measure of levels of implementation from the 1970-71 data is the sponsors' ratings of the teachers.[2] As part of the Planned Variation evaluation, each sponsor was asked to rate individually every head teacher with a Planned Variation class on a 0-9 scale ("not acceptable" to "outstanding")

[2] A sample rating sheet is included in Appendix B.

on the basis of how well he or she performed in the sponsor's model.[3] The ratings are impressionistic measures, in that judgments are based on the sponsor's personal conception of the total model, rather than on the basis of explicit, predetermined criteria. In most cases, the ratings were actually completed by the sponsor's field staff, rather than by the sponsor himself. Thus, within one model, ratings were often done by more than one person. The ratings were recorded for five points in time: October, February, prediction of May, (actual) May, and prediction of the following fall. The forms, however, were filled out only twice. The first set of ratings was completed in February when the sponsors were asked to judge the teachers' performance for the previous October and the following May, as well as for February. The second set, done in May, included predictions for the following fall as well as judgments for May.

Although this instrument appears to be straightforward, there are a number of issues which must be considered in interpreting the findings drawn from it. First, the ratings only focus on one aspect of the model: the classroom program. Many of the models have other components, such as parent involvement and take-home tasks, which are not reflected in this measure.

[3]Only one teacher per class was rated; aides were not included. If a person taught more than one class, she was still only rated once. In our analyses, however, we duplicated the ratings for each PV class a teacher was assigned to.

With the exception of the Gordon model, however, the classroom program seems to be the most important part of the model, so that ratings of the class can be taken as a reasonable, if not perfect, assessment of the model. In this same vein, many models emphasize the role of aides as classroom teachers, so that implementation must be viewed as the result of a team effort, rather than that of the head teacher alone. It is possible the head teacher is doing an excellent job with the model, but the aides are not, or vice-versa. In either case, a rating of the head teacher alone does not adequately represent the extent to which the model is being implemented. It could be argued, however, that in the majority of classes, the performance of the head teacher is indicative of the team contribution to model implementation. So again, we conclude that the sponsor ratings provide a reasonable but not perfect representation of levels of implementation.

A second factor to be kept in mind in using the sponsor ratings is the time at which the ratings were made. One problem is that not all sponsors completed the second set of ratings in May as had been planned. In fact, the dates of the second ratings vary considerably, with some not being completed until the following fall.[4] This undoubtedly lowers the reliability of the measure. A second problem stems from the fact that the

[4]Some ratings were received by SRI as early as June 9, while others were as late as October 18.

ratings were actually made only in February and May; the October ratings are not independent judgments. The importance of considering this procedure is illustrated by Table 3 which shows that apparent improvement in implementation between October and February is larger than the improvement between February and May. If this relationship is valid, it provides important information about the process of implementation. We suspect, however, that the increase between October and February is an artifact of the instrument, rather than an accurate reflection of changes in teacher performance. Since the October ratings were actually made in February, it seems probable that they are influenced by the sponsor's knowledge of the teachers' performance in February: because he expects teacher performance to improve over the year, and because he cannot remember the teachers' actual performance in October, the sponsor may unknowingly arrive at the October rating by simply subtracting a reasonable amount from the February rating. In an attempt to make a decision about the nature of these ratings, we examined the correlations between them. Since the elapsed time between October and February is 4 months and that between February and May, 3 months, we would expect the February-May correlation to be at least as large as the October-February correlation. On the contrary, we find that the October-February correlation is .86, while the February-May correlation is .65. This supports

---

[5]The predictions for May and the fall have been excluded because they are not appropriate for determining actual levels of implementation.

TABLE 3

Sponsor Ratings
Changes in Ratings Between Points in Time

Model	Oct.		Feb.		May
Far West	3.7	(1.3)	5.0	(0.8)	5.8
Bank Street	4.0	(0.8)	4.8	(0.3)	5.1
Oregon	4.6	(1.0)	5.6	(-0.7)	4.9
High Scope	4.1	(0.8)	4.9	(0.1)	5.0
Florida	4.5	(0.7)	5.2	(0.1)	5.3
Pittsburgh	3.7	(1.3)	5.0	(-0.1)	4.9
REC	4.2	(2.8)	7.0	(-0.5)	6.5
Enablers	4.3	(0.9)	5.2	(0.4)	5.6
Total	4.1	(1.2)	5.3	(-0.1)	5.2

Note: The numbers in parentheses give the size of the change in ratings between the two points in time. Negative numbers indicate that the average rating goes down in that interval.

the argument that October ratings are largely a function of February ratings. Therefore, in the remainder of the analyses, we will include only the February and May ratings.

Third, it must be recognized that the sponsors' ratings cannot be used for comparisons between models. On the one hand, impressionistic measures are good because they convey the sponsor's global evaluation of how the model is going, and this can be seen as a valid judgment. On the other hand, this approach is limited because there are no shared standards for making the ratings. This is particularly true in judging different models: in order to make comparisons among models, we would have to assume that sponsors were using the rating scale in the same way. Since we cannot make this assumption, we cannot use this instrument to compare models as to their level of implementation in Planned Variation.[6] The lack of shared, or at least explicit, criteria is also problematic in terms of interpreting the results of the ratings. In many cases, it is difficult, if not impossible to determine what a sponsor means in making his ratings. For

---

[6] Even within models there is some question of the comparability of implementation ratings, since one person rarely rated all sites in a model. We assume, however, that in working together and working with the same model, a sponsor's staff will have fairly comparable standards for judging implementation. At the same time, we must recognize that this is only an assumption, and we expect that it may be more accurate for some models than for others.

example, we do not know whether a sponsor has the same standards for first and second year teachers, nor do we know how he interprets the category "average".

A final factor which must be considered in using the instrument is missing data. This is not a problem of the instrument itself, but of the data which results from it. For two models (Kansas and EDC), there are no data on this measure; these sponsors refused to use the instrument. For two additional models (Arizona and NYU), there is only information from the second rating form; the first was not completed for any of the sponsors' sites. In the remaining models, some data are missing because changes in individual staff members were not accurately documented. As a result, the data are very confused and cannot be used. Because of missing data, the analyses have been done on less than a total sample of Planned Variation classes.

Thus, in using the sponsor ratings of teachers we must be aware that they focus only on the head teacher, and primarily on the classroom; that only the ratings for February and May provide independent judgments; and that the ratings cannot be used to make comparisons among models. Within the limits of these factors, however, the sponsor ratings do provide overall quantitative judgments of teacher performance which may be useful in assessing levels of implementation.

In the remainder of this chapter, we examine the findings from these ratings. We first compare levels of implementation in sites which have participated in Planned Variation for two years with sites which have participated for one year. Second, we discuss the mean levels of implementation at each site. Third, we examine the amount of variation between levels of implementation within models and over time.

EFFECTS OF NUMBER OF YEARS OF PARTICIPATION ON IMPLEMENTATION

Using the sponsor ratings of teachers, the first question to be addressed is what differences are there between first and second year sites in their average level of implementation? Although 1970-71 was the second year of the Planned Variation experiment, for some sites in all models (as well as some entire models), it was the first year of participation. The question of differences between first and second year sites is important because we would predict that a site which had worked with a model for two years would have a higher level of implementation than a site which had worked with it for one year. If this is true, we must separate year 1 from year 2 sites in subsequent analyses. To test our hypotheses we performed a repeated measures unweighted means analysis of variance on the ratings for February and May for sites in models which had been in

the experiment for two years.[7] We did not include models as a variable in this analysis because there was an insufficient number of comparison points for all but two models. Table 4 shows the means for each year at each point in time and the results of the analysis of variance. Although the means for sites with two years of participation in Planned Variation are slightly higher than those with one year at each point in time, the differences are not statistically significant ($p = .178$). The interaction between year of participation and time of rating is also insignificant. We cannot say that the differences at either time are due to anything other than chance. Thus, sponsor ratings of teachers cannot be used to distinguish between levels of implementation during the first and second years. This does not necessarily mean that there are no differences in the first and second year of implementation. It is possible that differences exist but that the sponsor ratings are not sensitive to them. If a sponsor has higher standards for teachers in second year sites than in first year sites, a second year teacher could perform at a higher level than a first year teacher but receive

---

[7] Four models, REC, NYU, Pittsburgh and the Enablers, were not part of the Study during the first year and therefore, have no second year sites. Since we are interested in a comparison between years, there is no reason to include the sites from these models in this analysis. For a list of which sites were included in this analysis, see Appendix A, Table 1.

TABLE 4

Sponsor Ratings
Comparison of First and Second Year Sites

Means

Years of Participation:	Feb.	May
One	4.8	4.9
Two	5.3	5.3

Analysis of Variance

Source	df	Mean Square	F-test	% of Total Sum of Squares
Year of Participation	1	7.680	1.849	1.27
Site	14	9.783	2.356**	22.67
+Class within site	89	4.153	NOT TESTED	61.18
Rating-time	1	0.047	0.056	0.01
Year X Rating-time	1	0.142	0.170	0.02
Sites X Rating-time	14	1.100	1.318	2.55
+Class X Rating-time	89	0.835	NOT TESTED	12.30
Total	209	2.891		100.00

Note: An unweighted means analysis of variance is used to approximate a solution for unequal cell sizes. A balanced design was created by selecting an equal number of first and second year sites from the total sample. Models are not included in this analysis because there are an insufficient number of comparison points to maintain a balanced design. Ratings from both February and May are included as a repeated measure.

+: indicates the effect used in testing the preceding effects; class-within-site is used to test both sites and models because sites are considered as a fixed factor.

**: $p < .01$

the same rating. If criteria are relative to previous levels of performance or to sponsors' expectation, then real differences in implementation may not be revealed. Regardless of whether real differences do or do not exist, they are not reflected in this measure.[9] Therefore, year of participation will be ignored in further analyses in this chapter.

MEAN LEVELS OF IMPLEMENTATION

The second question to be examined through the sponsor ratings of teachers is: to what extent does the Planned Variation teacher's performance match the sponsor's conception of his model? Table 5 presents the means and standard deviations for sites and models. The primary conclusion to be drawn from this table is that in May all models are considered to be moderately well-implemented, but none are outstandingly well-implemented. The mean ratings for all models range from 3.7

[9]It is possible that the site level is too gross to reveal the effects of the number of years of participation on implementation because there are first year teachers even at second year sites. To test the hypothesis that teachers who have been working with the model for two years have higher implementation ratings than teachers who have been working with the model for one year, we performed a repeated measures unweighted means analysis. Using only second year sites (there is no variation in first year sites), we found no differences between first and second year teachers in levels of implementation. See Appendix C Table. 1.

TABLE 5

LEVEL OF IMPLEMENTATION: BY SITE AND BY MODEL
MEAN AND STANDARD DEVIATIONS

	Feb.		May		
	$\bar{x}$	SD	$\bar{x}$	SD	N
Far West	5.0	2.30	5.8	1.41	27
Buffalo	4.2	2.99	6.5	1.64	6
Duluth	4.4	2.45	5.5	1.51	8
Fresno	4.0	1.41	---		4
Salt Lake	6.2	1.50	4.8	.96	4
Tacoma	6.8	1.10	6.2	.84	5
Arizona	---		5.1	1.29	29
Lafayette	---		4.7	1.36	17
Lakewood	---		6.0	.89	6
Lincoln	---		5.3	1.03	6
Bank Street	4.8	1.81	5.1	1.50	32
Boulder	4.5	2.08	5.2	2.06	4
Tuskegee	4.0	1.96	4.7	1.03	13
Wilmington	5.5	.93	5.6	1.19	8
Elmira	5.4	1.90	5.0	2.37	7
Oregon	5.6	1.50	4.9	1.91	16
E. St. Louis	6.0	1.63	5.0	1.16	7
Tupelo	6.3	1.53	5.7	2.52	3
E. Las Vegas	4.8	1.17	4.5	2.51	6
Kansas: No data					
High Scope	4.9	1.39	5.0	1.37	31
Ft. Walton Beach	5.0	1.58	4.4	2.07	5
Central Oz	5.0	1.41	5.2	1.42	16
Greeley	5.2	.96	5.5	1.00	4
Seattle	4.5	1.64	4.7	.52	6
Florida	5.2	2.13	5.3	1.89	18
Jacksonville	4.0	0.0	---		1
Jonesboro	5.0	0.0	5.3	.58	3
Chattanooga	6.4	.53	6.2	.83	9
Houston	2.8	3.20	3.2	2.75	5
EDC: No data					
Pittsburgh	5.0	.58	4.9	1.34	7
REC	7.0	.76	6.5	.54	8
NYU	---		3.7	1.95	11
Enablers	5.2	1.53	5.6	1.40	29
Billings	4.6	1.34	5.8	.84	5
Colorado Springs	5.8	2.14	5.8	1.60	6
Bellows Falls	6.2	1.17	6.2	1.60	6
Newburgh	4.8	1.39	4.8	1.39	8
Puerto Rico	4.6	.58	6.0	1.56	4

Note: These means and standard deviations are based on all available data.

to 6.5, indicating that for 9 out of 10 models, the means fall within the category designated as "average" on the original rating form (ratings 4, 5 and 6 are included in the average category). The tenth model (NYU, 3.7) is only slightly below this category.[10] The problem with these relatively straightforward findings, however, is that the category labels are confusing: the scale is not anchored by an explicit definition of what constitutes an "outstanding" performance, and the meaning of "average" is not clear. Because of these ambiguities, our interpretation of the data must be cautious. It does seem fair to state that all sponsors seem to be moderately pleased with the implementation of their model on the average, but that none are raving about unqualified success.

A second conclusion to be drawn from Table 5 is that there is a great deal of variation within sites in levels of implementation, as is reflected in the large standard deviations.

---

[10] An examination of the frequency distribution for these models confirms these findings. The frequency distributions for May show that for the same 9 out of 10 models, at least 69% of the teachers in each model are judged to be performing at average or better levels; again, the tenth model is only slightly lower, with 64% of the teachers being rated average or better.

Examination of the frequency distributions for this measure show that in many sites, the ratings range from 3 to 7 or 8.

From Table 5 it is also apparent that there is variation among sites in the same model in the mean ratings for both points in time. In addition, there are differences between the February and the May ratings for both sites and models. Before we can draw conclusions about the importance of either of these observations, however, we must perform further analyses to see if the differences are statistically significant.

VARIATION WITHIN MODELS AND OVER TIME

The major questions to be addressed in this section, then, are: how much variation in levels of implementation exists within models, and how much change in ratings occurs between February and May? To answer these questions we ran a repeated measures unweighted means analysis of variance. The analysis takes the three relevant variables -- models, sites within models[11], and ratings times (February and May) -- into account at the same time.

[11] It is important to remember that sites are nested within models and that as a result, the effects of a site cannot be separated from the effects of the site in interaction with the model; we must always speak of "sites within models."

The sites and models included in this analysis differ from those in the "standard" balanced design described in Chapter 1 because of missing data. In addition to the models previously excluded from the standard design, EDC and Kansas were excluded from this analysis because they did not complete the ratings. The Arizona model was also excluded because teachers were only rated in May, making comparisons over time impossible.[12] The remaining six models were run with three sites per model in order to generate a balanced design.[13] Table 6 shows the results of the analysis of variance.

A number of conclusions can be drawn from this analysis. The strongest finding is a significant effect for sites ($p < .01$). This means that there are significant differences between sites within the same model in the ratings they receive. Such a finding suggests that sponsors can distinguish between sites as to how well they are implementing the sponsor's model, and this in turn suggests that models do not represent uniform treatments in this experiment.

Another finding illustrated by Table 6 is that approximately 58% of the variance in ratings lies within sites. This

---

[12] An analysis done for May alone with Arizona included does not alter the conclusions drawn from the major analyses without it.

[13] See Table 2 for a list of sites and models included in this analysis.

TABLE 6

Sponsor Ratings
Repeated Measure Analysis of Variance

Source	df	Mean Square	F-test	% of Total Sum of Squares
Model	5	3.540	0.834	3.05
Site Within Model	12	10.385	2.446**	21.44
+ Class within Site	79	4.246	NOT TESTED	57.71
Rating-time	1	1.811	2.041	0.31
Model X Rating-time	5	2.490	2.806*	2.14
Site X Rating-time	12	1.596	1.798	3.29
+ Class X Rating-time	79	0.888	NOT TESTED	12.06
Total	193	3.012		100.00

Notes: The analysis is based on a balanced design with 6 models and 18 sites; the balanced design was created by eliminating models with only one site and by eliminating level I sites (minimal testing) where necessary. An unweighted means analysis of variance is used to approximate the solution with an unequal number of classes within the sites. Since sites are nested within models, the effects of a site cannot be separated from the effects of the site in interaction with the model.

+: indicates the effect used in testing the preceding effect; sites are considered as a fixed factor.

*: $p < .05$
**: $p < .01$

indicates that most of the differences in levels of implementation are between teachers in the same site; in general, all sites have both good and bad teachers. This finding supports the assertion made above that experimental treatments vary greatly across the classrooms representing each model.

The third finding of the analysis is a moderately significant model by rating-time interaction ($p < .05$). This means that the relationship between February and May ratings is dependent on the model to be considered. Examination of Table 7 shows mean ratings for the six models included in the analysis. From this table we can see that in some models the ratings improve between February and May, while in others they decrease or remain essentially the same. The largest changes are in Far West and Oregon: the Far West ratings increase by 1.1 points, and the Oregon ratings decrease by .7. The interpretation of the scores which improve or stay the same is fairly clear: the sponsor is as satisfied or more satisfied with the teachers' performance in his model in May as he was in February. The scores which decrease are more difficult to interpret. It is possible that the change in ratings indicates an actual decrease in level of implementation; perhaps the teachers were tired or discouraged with the model at the end of the year. It is also

TABLE 7

Sponsor Ratings
Model Means from the Analysis of Variance

Model	Feb.		May
Far West	4.7	(1.1)	5.8
Bank Street	5.2	(0.1)	5.3
Oregon	5.7	(-0.6)	5.1
High Scope	4.9	(0.0)	4.9
Florida	4.5	(0.2)	4.7
Enablers	5.3	(0.5)	5.8
Total	5.1	(0.2)	5.3

Note: The model means differ from those given in Table 5 because they are based only on data from 3 sites in each model; Table 5 is based on all available data.

The numbers in parentheses indicate the size of the change in ratings between February and May.

possible, however, that the decrease reflects a change in the sponsor's standards more than in performance; the sponsor may have higher expectations of implementation in May than in February, such that performance that was essentially the same at the two points in time would be given a lower rating in May. The information available does not enable us to choose the correct interpretation. The absence of an overall effect for time of rating is somewhat surprising because the original assumption of the Planned Variation study was that all classes move consistently toward full implementation.

Finally, the analysis reveals no significant differences between models. In fact, the F ratio for this effect is extremely small ($p > .50$). Even empirically, sponsor ratings cannot be used to determine whether one model is implemented to a greater extent than another. This reinforces the conceptual argument expressed earlier against using these ratings for between-sponsor comparisons because the ratings do not necessarily have the same meanings to different sponsors. Thus, even if we wanted to use this measure to rank sponsors, we could not. There is no evidence that the differences between models are due to anything other than chance.

SUMMARY

In using the sponsor's ratings of teachers as indicators of levels of implementation, we have reached a number of conclusions.

1. Twenty-eight of the thirty-one sites for which there is data on this measure, on the average, are reported by the sponsors to be moderately well-implemented, but none are outstandingly well-implemented.
2. Contrary to our expectations, analyses reveal no significant differences in levels of implementation between first and second year sites and no overall effect of time of ratings. There is a model by rating time interaction which shows that implementation in some models improves between February and May, while in other models it declines. The conclusions drawn from each of these findings must be tentative because of a lack of knowledge of how the ratings were being used. It is possible that the findings are artifacts of the instrument, rather than indications of actual levels of implementation.
3. Perhaps more important, there are no significant differences between models in levels of implementation, but there are significant differences between sites within the models. Moreover, most of the variation in this measure is within sites. Although, as

we have said, there is a question about how to interpret the ratings, the finding of differences indicates that sponsors do distinguish between people and sites within their model, judging some of them to be doing a better job of implementing the model than others.

These findings, then, do suggest that implementation is not proceeding as had been expected. They raise the possibility that implementation of a curricular model involves more than a technical process of transferring a well-developed model. Thus, the answer to the question raised at the beginning of the chapter -- can models be delivered? -- does not appear to be yes in all cases. Perhaps the sponsors cannot duplicate their models in all classes.

The large variation in sponsor ratings within sites also indicates that models do not represent a uniform treatment, as would be expected in an experiment testing model effectiveness. By simply indicating with which model a class or site is working, then, we cannot describe the classroom program there. Thus, the answer to the second question raised at the beginning of the chapter -- whether the treatments are fully implemented -- appears to be no.

These possibilities that implementation is not proceeding as expected require that additional questions be asked. First, if the treatment is not fully implemented, then we must ask which parts of the model <u>are</u> being implemented

before we can draw conclusions about effectiveness. This question is dealt with only in a preliminary fashion in this report, but is being studied more fully during the 1971-72 year. Second, since models are not being implemented in all classes, we must ask why there is variation. What factors affect implementation such that all classes are not at the same level? The next two chapters are devoted to this question.

Chapter 3

FACTORS WHICH INFLUENCE IMPLEMENTATION

In Chapter 1, we asserted that the question of which factors affect implementation can be interpreted in two ways. First, the factors, or variables, can be used descriptively. In this context, the relevant implementation questions are, what does participation in Planned Variation involve for the Head Start staff and community, and how does that participation differ from participation in regular Head Start? To answer these questions, in part, we will examine specific aspects of implementation, such as the kinds of training received by the teachers, and the characteristics of the teachers who are working in Planned Variation. We are interested in both comparing experiences in different models and in comparing Planned Variation with non-Planned Variation. The descriptive framework is one in which the variables to be presented were originally conceived.

Second, the factors which affect implementation can be used to explain levels of implementation. In the last chapter we found large variation among classes in the extent to which they were implementing the models. This leads to the question of why such variation exists: why is there a higher level of implementation in one class or site than in

another? The same variables which are presented descriptively can also be used to try to explain this variation. From this perspective, it appears that the process of implementation is more complex than originally anticipated. While we would clearly expect the sponsor's input to be an important determinant of implementation, we think that this conception of implementation should be expanded. First, we propose that an examination of sponsor input should not be limited to training, as it has been previously[1], but should include other dimensions, such as staffing patterns and involvement. Second, and more importantly, we propose that factors other than sponsor input influence model implementation. In exploring these additional factor, we have broken them into two general categories: staff input and the context in which the staff operates.

The important of the first category of variables rests on the assumption that sponsor input will not be effective in achieving implementation unless the model is adopted and used by the staff. The success of such adoption

---

[1]Support for this definition is found in the fact that data was collected only on this aspect of the sponsor's input.

(or implementation) depends both on how well a teacher likes the model and is willing to make an effort to implement it, and on the level of her skills and previous teaching methods.

The second category, the operational context of the staff, is based on the assumption that the sponsor and staff inputs are influenced by the situation around them. More specifically, we propose that implementation is affected by characteristics of a site (such as its size or funding agency), by the efficiency with which a site is managed, and by the support which administrators give to a model.

Thus, we see implementation as an interactive process which depends not only on the input of the sponsor, but also on the reaction and input of the staff with whom the sponsor is working and the context in which they operate.

The data to be presented are taken from a number of sources: the Teacher and Aide Questionnaires, the Sponsor Implementation Reports, the Final Consultant Reports and several items from miscellaneous sources. Our primary analytic tool is the unweighted means analysis of variance as described in Chapter 1 (see p. 26). Four sets of analyses were performed. The first, based on Planned Variation classes only, examines differences among sites and models on specified variables in the standard balanced design. The second contrasts Planned Variation with non-Planned Variation (PV-NPV)

and uses a balanced design with a smaller number of sites and models.[2] The third and fourth, based on the sponsors' reports and the consultants' reports[3], are simple site within model analyses. Since there is only one observation per site, the site is the basic unit, and a balanced design is not necessary.

One analysis issue which has not been discussed previously is the appropriateness of the analysis of variance strategy with dichotomous response variables. This is an important consideration since a sizable proportion of the variables which follow are dichotomous. Although it may appear questionable to use an analysis of variance in this situation, Lunney[4] has recently presented evidence that this departure from the usual assumptions of the ANOVA model may not be serious. A theoretically more appropriate way to look at the data is a 2-way contingency table (sites by presence of characteristics). Suppose there are s sites and

---

[2] See Table 2 for the specific sites included in these designs.

[3] Sponsor reports were completed for only 20 sites. There is no data for Tuscon, Weikart or EDC. Moreover, they were not done until at least 2 or 3 months after the school year: the earliest report is dated August 10; the latest, December 21. For 9 sites, 3 entire models, the reports were filled out on October 29 or later. Consultant reports are available for 34 sites.

[4] Lunney, G. "Using analysis of variance with a dichotomous dependent variable: an emperical study" J. of Educational Measurement, 1970, 7: 263-269.

m models. As shown by Cochran[5], the total χ^2 with S-1 degrees of freedom can be partitioned into components with m-1 degrees of freedom for testing model differences and s-m degrees of freedom for testing differences among sites within models. This alternative analysis was carried out as a check on the ANOVA. The resultant findings show no serious discrepancies from the findings obtained with the analyses of variance model. The analyses are summarized in Table 11, Appendix C. The analyses of variance are included in the body of the report instead of the χ^2's for simplicity of presentation.

Three additional issues must be kept in mind in using these data: lack of independence among variables, testing large numbers of variables, and using sites as fixed factors. First, the number of significant effects is probably inflated since many of the variables included here are not independent of each other both because they measure dimensions (such as behavior and characteristics) that are correlated, and because they are structurally related: several variables may be taken from one question. Second, testing large

---

[5] Cochran, W. "Some methods of strengthening the common χ^2 tests." Biometrics, 1954, 10: 417-51.

numbers of variables at the same time increases the probability of finding significant effects by chance. Third, the use of sites as fixed factors, results, in this analysis, in liberal F tests. A † symbol on analyses of PV only tables indicates model effects which would have been insignificant if sites had been considered random[6]. There are many of them. Each of these issues leads to the conclusion that the significant effects presented in this chapter should be interpreted cautiously. We will use significance levels heuristically, then: as rough estimates of differences and as indicators of interesting findings, but not as the basis for grand inferences.

Hence, Chapter 3 is intended to examine a large number of variables relevant to model implementation in Planned Variation. The body of the chapter is divided into three sections, corresponding to the categories of factors presented earlier: sponsor input, staff reaction and input, and the operational context of the staff. Within each section, the variables first are presented and the findings discussed. Then, the variables are considered as predictors of levels

---

[6] Using sites as random factors results in overly conservative estimates of significance because of the small number of sites within each model.

of implementation, and additional areas of study are suggested. The second step is more speculative than the first, but is necessary because the variables with which we are working do not adequately tap all dimensions which might affect model implementation.

Because there are more data specific to the study available for sponsor input, the data presentation in that section is much longer and more detailed than in the others. Conversely, the discussion of the variables as predictors is fairly brief, both because there is less need to elaborate on the anticipated relationships between training and implementation since these relationships are more commonly accepted than others we suggest, and because there are fewer aspects of sponsor input for which there are no data. In the sections on staff input and on context, the primary emphasis is placed on exploring ways in which variables might be used to explain variations in levels of implementation, and in suggesting additional variables which should be considered. The data presentation in these sections is relatively limited.

SPONSOR INPUT

The sponsor's primary responsibility in Planned Variation is to transfer the models from the sponsor's home base to the Head Start sites. Since the modes of teaching required by a model may, for many teachers, be different from their usual modes, implementation requires that a Head Start staff member acquire new skills and techniques in the class and new ways of relating to the children. The sponsor bears the responsibility for training the staff in these new models of teaching. In this section we will present data on training from the Teacher Questionnaire and the Sponsor Implementation Report. Several aspects of training are dealt with. We will discuss data on the length, frequency, types, and sources of both pre-service and in-service training. In addition, we will examine data on the continuing support and feedback provided by the sponsor.

While training may constitute a major portion of the sponsor's input, it does not define it completely. Therefore, after presenting the available data, we will discuss the the adequacy of the data within the framework of the expanded conception of implementation presented in the introduction, and will suggest other aspects of the sponsor's inputs which should be considered in attempting to deal with

implementation more fully. These include the sponsor's staffing patterns and the non-model areas with which he becomes involved in the community.

Pre-service Training:

Tables 8, 9 and 10 summarize the results of the analyses carried out on variables relevant to pre-service training.[7] Table 8 presents variables for which there are significant differences among models or among sites within models. In addition to displaying the significance levels, the table shows the partitioning of variance, giving the percent of variance explained by a knowledge of models and sites, and the percent of variance remaining among classes within the sites. The analyses are based on the "standard design" (as specified in Table 2) and include only Planned Variation classes.

Table 9 presents the results of the comparisons between Planned Variation and non-Planned Variation classes (PV-NPV) for the same variables that were analyzed in Table 8. Again, both significance levels and the partitioning of variance are given. The important effects in this table are the overall

---

[7] One general problem with most of the questions on training is that they are not specifically limited to model training. Using them in the context of implementation, then, requires the assumption that most, if not all training in a site, is related to the model. Since there is evidence that this is not always the case (e.g. people other than the sponsor representative are reported to give training), the findings presented here must be used with caution in drawing conclusions about model implementation.

TABLE 8

Pre-Service Training

Analyses of Variance for Planned Variation Classes

	Variable Name	Sources of Variance: Model	Site Within Model	Class Within Site	df Class
TQ#4	Did you receive pre-service (summer) training?	18.13%† ***	26.61% ***	55.26%	132
TQ#5	How long was the pre-service training period: # days	16.16† ***	23.79 ***	60.04	124
TQ#6	Kind of pre-service training: Demonstration lessons	20.31 ***	12.76 NS	66.93	135
	Lectures	23.11† ***	25.65 ***	51.24	135
	Individual meetings with leader	10.76† **	21.24 **	68.00	135
	Group discussions	15.36† ***	22.51 ***	62.13	135
	Discussion of video taped lessons	23.47 ***	13.37 NS	63.16	135
	Observations	22.15 ***	13.35 NS	64.50	135
	Role playing	18.80 ***	16.71 *	64.49	135
TQ#6	Who gave training (person who was checked one or more times) Sponsor	18.91 ***	17.52 *	63.57	135
	Consultant	23.23 ***	13.07 NS	63.70	135
	Local Head Start	14.18† ***	22.83 ***	62.99	135
	Other	9.61† *	14.56 NS	75.83	135

See attached page for notes.

Notes for Table 8:

The letter and number code in the left-hand column indicates the questionnaire and item from which the variable is taken: TQ is the Teacher Questionnaire.

The top row of numbers for each variable gives the percent of variance explained by models and sites and the percentage found within sites. The second row of numbers for each variable indicates the significance of the differences between models and sites. Conventional notation is used:

*** = $p < .001$
** = $p < .01$
* = $p < .05$
NS = not significant

Df class refers to the degrees of freedom for classes within sites (with the Teacher Questionnaire data, "class" is equivalent to "teacher"--there is one teacher per class; with Teacher Aide data, "class" is equivalent to "aide."). Since there are always 18 df for sites and 8 df for models in the analyses shown in this table, the information is not repeated in the body of the table.

Sites are considered as fixed factors; therefore, the mean square for both models and sites is tested against the mean square for classes within sites. A + symbol indicates that model effects would be insignificant if models were tested against sites. There is no site by model interaction term because sites are nested within models.

The analyses in this table are based on the standard design with nine models and 27 sites. It should be noted that the analyses in Tables 9 and 10 are based on a different set of sites and models.

The analyses of questions 5 and 6 are based on responses from all teachers, including those who reported no preservice training (question 4).

TABLE 9

Pre-Service Training

PV-NPV Comparisons

	Variable Name:	Sources of Variance: PV-NPV	(X)	Model	Site	Model X PV-NPV	Site X PV-NPV	Class	df Class
TQ#4	Did you receive pre-service (summer) training?	1.40 NS		26.66 ***	4.72 NS	3.16 NS	5.68 NS	58.39	78
TQ#5	How long was the pre-service training period: # days	9.81 ***	PV=7.2 NPV=3.2	14.62 **	12.21 *	4.39 NS	10.98 *	48.00	72
TQ#6	Kind of pre-service training:								
	Demonstration lessons	6.23 NS		26.13 ***	4.01 **	4.96 NS	5.79 NS	52.87	79
	Lectures	0.17 NS		42.81 ***	0.93 NS	8.02 *	2.32 NS	45.73	79
	Individual meetings with leader	0.04 NS		16.38 **	14.83 **	5.25 NS	6.16 NS	57.34	79
	Group Discussions	9.12 ***	PV=.77 NPV=.46	14.96 **	9.14 NS	3.13 NS	6.09 NS	57.56	79
	Discussion of video taped lessons	7.25 ***	PV=.48 NPV=.22	37.32 ***	5.61 NS	5.43 NS	2.28 NS	42.11	79
	Observations	11.39 ***	PV=.60 NPV=.25	24.86 ***	2.56 NS	1.16 NS	3.29 NS	56.73	79
	Role playing	8.18	PV=.63 NPV=.33	28.46 ***	11.01 **	2.71 NS	4.55 NS	45.09	79
TQ#6	Who gave training (person who was checked one or more times): Sponsor	7.48 **	PV=.62 NPV=.33	15.09 **	6.23 NS	3.81 NS	5.19 NS	62.19	79
	Consultant	2.74 *	PV=.65 NPV=.48	35.13 ***	2.35 NS	4.27 NS	2.61 NS	52.90	79
	Local Head Start	0.75 NS		29.31 ***	15.08 ***	7.58 *	2.58 NS	44.70	79
	Other	0.30 NS		19.76 ***	4.69 NS	1.49 NS	10.66 *	63.10	79

See the attached page for an explanation of this table.

Notes for Table 9:

The top row of numbers for each variable gives the percent of variance explained by PV-NPV, models, sites, interactions and classes within sites. The second row for each variable indicates the significance of the effects. Conventional notation is used:

*** = $p < .001$
** = $p < .01$
* = $p < .05$
NS = Not Significant

The analyses are unweighted means analysis of variance, based on a balanced design with 6 models and 12 sites. All effects are tested against class within model. df Class refers to the degrees of freedom for classes, or teachers, within sites. The degrees of freedom for the other effects are constant and are therefore not included in the body of the table; models have 5 df in this analysis, sites have 6 df, PV-NPV have 1 df, PV-NPV by model have 5 df and PV-NPV by class have 6.

Model and site effects are shown only to present the total analyses. They are difficult to interpret because they are based on PJ and NPV responses combined.

When PV-NPV effects are significant, the overall means for each group are shown.

TABLE 10

Pre-Service Training

Analyses of Variance from Sponsors' Reports

	Variable Name:	Sources of Variance: Model	Sites Within Model	df sites
SI#4a	Pre-service training (1) Total no. of hours			
	For Teachers	56.4% *	43.6%	12
	For Aides	56.1% *	43.9%	11
	(3) How many days did the pre-service training cover?			
	For Teachers	21.8% NS	78.2%	12
	For Aides	53.5% NS	46.5%	10

Note: SI: Items are taken from the Sponsor Implementation Report.

These analyses are based only on five models and 17 sites because of missing data and the omission of models with only one site. Since there is only one observation per site, there is no within-site variation, and the site effects cannot be tested. The top row of numbers for each variable gives the percent of variance explained by models and by sites within models. The second row for each variable indicates the significance of differences between models. Conventional notation is used:

*** = $p < .001$
** = $p < .01$
* = $p < .05$
NS = not significant

df sites refers to the degrees of freedom for sites; df models is not shown because it is always 4.

PV-NPV effects and the interactions of PV-NPV with models and sites. A significant PV-NPV effect means that differences between the two groups exist when all models and sites are considered together. If PV is greater than NPV on a given variable, this indicates that on the average, the scores for PV teachers tend to be higher than the scores for NPV teachers. When there are significant PV-NPV effects, the means for each group are also shown in the Table 9. An interaction between PV-NPV and sites (or models) means that the relationship between PV and NPV is different in different sites (or models). The model and site factors are shown only to display the entire analysis. They are otherwise confusing because they include both PV and NPV responses. Thus, when we find significant differences among models in this table, it means that differences exist even when PV and NPV are combined. The interpretation of findings of model differences which hold for both experimental and control groups is not clear. The PV-NPV analyses are not based on the same sites and models as the previous analyses on only PV classes. The PV-NPV analyses include 6 models and 12 sites, again in a balanced design. This number is smaller than the standard design because data for non-Planned Variation teachers and aides was collected in fewer sites, (see Table 2).

Table 10 contains analyses of variance for pre-service

training items taken from the sponsor implementation report. In these analyses there is only a model factor and a site factor because for each variable, there is only one observation per site. Consequently, the test of significance for model differences uses the mean square for sites as the denominator, and no test of significance of the differences between sites is possible. The analyses are based on 5 models with 2 to 4 sites in each.[8]

The means and standard deviations for all sites for each of these variables are given in Appendix D. The means for the models included in the analyses presented here are shown in Table 11. These means are shown because they are based only on sites in the balanced design and consequently are different from the model means shown in the appendix. The site means are not given because they do not change.

The first variable presented in Table 9 is the question "Did you receive pre-service (summer) training?"[9] Overall, 70% of the PV Teachers answer yes to this question. However, both model and site differences are significant. The model means range from 100% of the teachers in the Oregon model

---

[8] The number of sites can vary here because they are the basic unit and not a factor to be analyzed.

[9] For each variable discussed, the source will be indicated by an abbreviation in parentheses. This question is TQ#4: Teacher Questionnaire item number 4. An explanation of each abbreviation is given as a footnote in the accompanying tables and the complete questionnaires can be found in Appendix B.

TABLE 11a

Pre-Service Training
Means for Models in Standard Design

		Far West	Arizona	Bank St.	Oregon	Kansas	Hi-Scope	Florida	EDC	Enablers
TQ#4	Did you receive pre-service training?* (1=yes, 2-no)	1.36	1.49	1.46	1.00	1.07	1.17	1.32	1.22	1.72
TQ#5	How long was the pre-service training period?* (# days)	2.3	4.4	3.3	10.1	8.4	5.8	7.5	4.6	1.6
TQ#6	Kind of pre-service training received. (1=check, 0=no check)									
	Demonstration lessons*	.39	.45	.37	.95	.87	.66	.61	.56	.12
	Lectures*	.61	.43	.39	.90	.93	.63	.68	.56	.06
	Individual meetings*	.44	.38	.37	.56	.73	.43	.57	.50	.06
	Group Discussions*	.61	.45	.66	.90	.93	.83	.68	.78	.29
	Discussion of video taped lessons*	.24	.33	.14	.45	.80	.78	.46	.22	.06
	Observations*	.17	.33	.42	.82	.93	.45	.61	.61	.17
	Role playing*	.25	.30	.49	.77	.85	.72	.68	.44	.17
TQ#6	Who gave training? (1=check, 0=no check)									
	Sponsor Representative*	.41	.26	.57	.79	.87	.59	.47	.72	.11
	Consultant*	.17	.36	.40	.77	.87	.78	.47	.22	.17
	Local HS office*	.57	.16	.28	.66	.87	.48	.43	.44	.30
	Other*	.26	.04	.08	.38	.23	.19	.07	.00	.06

Notes: The means in this table are based only on data included in the standard design.
* indicates significant differences between models; extreme means are underlined: single underlining indicates high means and double underlining indicates low means.

TQ: Teacher Questionnaire

TABLE 11b

Pre-Service Training
Model Means from Sponsor Reports

	Far West	Bank St.	Oregon	Kansas	Florida
SI#4a Pre-service training					
(1) Number of hours for teachers*	29	26	60	24	40
Number of hours for aides*	29	23	60	24	40
(3) Number of days for teachers	3.5	7.2	7.5	4.3	5.0
Number of days for aides	3.5	4.0	7.5	4.3	5.0

Notes: SI#4a: The items are taken from the Sponsor Implementation Report; only means for models included in the analysis of variance are given.

*: indicates variables with significant model effect; on those variables, extreme means are underlined.

reporting that they received pre-service training while only 28% of the teachers in the Enabler model received it.[10] It is not surprising that the Enablers are lowest on this variable because there is no provision for pre-service training in this model. In addition to differences between models on this question, there are also large differences between the sites within a model. Four models have particularly large variations between sites in the percent of teachers who report that they have received pre-service training: Far West (Tacoma, 100% yes; Buffalo, 18% yes), Arizona (LaFayette, 81% yes; Lakewood, 0%), Bank Street (Boulder, 100%; Wilmington, 30%), Enablers (Bellow Falls, 83%; Colorado Springs and Billings, 0%). There are no differences between PV and NPV teachers on this variable.

---

[10] A number of the variables in this chapter have dichotomous responses: the answers are either yes/no or check/no check. With such variables we report the proportion of teachers responding in the text. In the tables of means, however, we report mean values. For check/no check responses where 1=check and 0= no check, the means are directly interpretable as proportions. With yes/no responses, there is a middle step because 1=yes and 2=no. In the data above, the mean for Oregon is 1.72, which means that 72% of the teachers did not receive pre-service training. The percentage who did receive training is obtained by subtracting the mean from 2.0. This procedure holds for all items where yes is coded as 1 and no as 2.

Because the significance levels in this report are primarily heuristic, we did not perform subsequent analyses after finding significant F ratios. Instead, extremes are reported. For differences between sites within models, then, we simply report the models with the widest range among sites on a particular variable, and give the means of the 2 sites with the extreme values.

There are also significant differences among both sites and models in response to the question of "How long was the pre-service training period?" (TQ#5).[11] Teachers in the Oregon model report the highest number of pre-service training days ($\bar{X}$=10.1 days), with Kansas (8.4) and Florida (7.5) following. At the other extreme, the Enabler teachers report the fewest days (1.6). and Far West is also low (2.3). Again, there is a great deal of variation among sites in the same model. The largest variation is in Arizona (LaFayette, 10 days; Lakewood, 0), Kansas (Mounds, 18.5 days; Portageville, 3.0), and High Scope (Fort Walton Beach, 10 days; Greeley, 2.2).

Data on the length of pre-service training are also available from the Sponsor Implementation Report (See Table 11). The sponsor reports, however, do not correspond with the teachers' responses. The correlation between the two accounts of the number of days of pre-service for teachers

---

[11]The analysis of this variable as reported here is based on a response classification slightly altered from the one obtained in the actual data. The original data showed a few sites with extremely high averages. From the large standard deviations which accompany those means, it appears that the question may have been misunderstood by some teachers. In an attempt to minimize this problem, we recoded all responses of 26 days or more to 26.

is only .02.[12] From the sponsor's accounts, the differences between models in the number of days of pre-service training given are not significant.[13] Oregon has the highest mean (7.5 days), but the mean is lower than that obtained from the teacher reports (10.1). In the case of Kansas and Bank Street, the differences are more serious because contradictory conclusions can be drawn from the two sets of data. From the teachers' data, the mean for Kansas is among the highest (8.4 days) but from the sponsor's account it is at the lower end of the distribution (4.3). The high mean reported by the teachers in Mounds may explain this discrepancy. There are also differences, however, between the sponsor and teacher reports in Bank Street. In this case, the sponsor's account of the number of days of pre-service trainings is high in relation to other models (7.2), while according to the teachers' account it tends to be low (3.3). Such discrepancies suggest that this question has different

---

[12]Surprisingly, the number of days of pre-service training reported by the teachers is more strongly related to the sponsors' reports of how much training was given to the aides (.413). It is not clear what this means. The correlations of teachers' reports of the length of their pre-service training also has low correlations with measures of pre-service in days. All correlations are displayed in Table 2, Appendix C.

[13]The Sponor Implementation Reports require a different analysis than do the Teacher Questionnaires. Because they have only one observation per site, we can test the significance of differences between models, but not between sites. Only five models are included in these analyses: REC, NYU, and Pittsburgh were omitted because each has only one site, and there are no data for Arizona, High Scope, or EDC.

meanings to different people. As a result, it is difficult to interpret the findings. There are significant differences in the total number of hours of pre-service reported by the sponsors with Oregon having the highest number (60) and Florida the second highest (40).

PV-NPV comparisons on length of pre-service (TQ#5) reveal an overall effect in which PV teachers report more pre-service training (7.2 days) than NPV teachers (3.2). There is also a moderate site by PV-NPV interaction; examination of site means shows that in contrast to the overall trend, NPV teachers in Wilmington, East St. Louis and Greeley, report more pre-service training than PV teachers.[14] Thus, it appears that models and sites differ in the amount of pre-service training they receive--although teachers and sponsors do not agree on what those amounts are--and that PV gets more pre-service than NPV.

The teachers were also asked, "What kind of pre-service (summer) training have you received and by whom?" (TQ#6).[15] Table 8 shows that there are significant model differences on all types of training as to whether or not the training

---

[14]No PV-NPV comparisons are possible from the Sponsor Implementation Reports.

[15]We are limited in interpreting the findings about types and sources of training by the generality of the labels used to identify some of these categories. Although it is impossible from the heading "observations" to know whether

was given; site differences are significant for four types (lectures, individual meetings, group discussion, and role playing) and are insignificant for the others. Examination of model means reveals that in all cases, either Oregon or Kansas has the highest proportion of teachers who report they had the specified type of training (See Table 11).

There are also significant differences among models on all four categories of response to the question of who gave the training, and differences among sites within models on two categories (sponsor representative and local Head Start office) (Table 8). It should be noted that these variables indicate only that a category was checked as

---

15 (cont.) it refers to a trainer observing the teacher, or to the teacher observing an exemplary classroom, the labels for the sources of training are more problematic than the headings for types of training. Each category has problems: one major difficulty is the question of where the sponsor's local representative is considered (this is the person who lives in the community, but is trained, and in some cases paid, by the sponsor to be responsible for model implementation in that site). Is she included as the sponsor representative or under the local Head Start office? We suspect that different teachers answered the question in different ways. The label "consultant" is even more confusing because a number of people can be included: the OCD consultant who is sent to monitor implementation (and should not be giving any training), the Regional Training Officer who works from the regional Head Start office, or people from local colleges. These people are all legitimately known as consultants, but are very different from each other. With "other" there is no space for specifying to whom it refers. As a result, we have no clue of how teachers used this category. Thus, the category labels contain ambiguities which limit the interpretations which can be drawn from these data.

being the source of at least one type of training; the variables do not tell how much or the number of types of training given. Again, Oregon and Kansas have consistently high proportions of teachers who report receiving some training from these people: (Sponsor: Kansas, 87%; Oregon, 79%; HS office: Kansas, 87%; Oregon, 66%) (Table 11).

Since these two models also have the highest number of teachers reporting that they received pre-service training, it is possible that the high means for types or sources of training are a reflection of that previous finding rather than real differences in the types of training offered. To examine this possibility, the analyses of variance on the pre-service training variables were re-run with only the responses of teachers who reported receiving some pre-service training included.[16] The results of these analyses (See Table 12)[17] show partial support of the hypothesis explaining

---

[16]Our assumption is that teachers who did not receive pre-service training will not be able to report on the length, type or source of that training. These zero responses may be extending the range of responses in the latter questions beyond what it would have been if only teachers who had received training were included. This results in skewing the means in the direction of those models and sites in which more teachers reported receiving pre-service training--toward those sites and models with fewer zero responses.

[17]The Arizona and Enabler models were excluded from this analysis because in one Arizona site (Lakewood) and three Enabler sites (Newburgh, Billings, and Colorado Springs) all teachers reported that they did not receive pre-service training. In the conditional analyses, these models would unbalance the design by having empty cells. To maintain a balanced design then, the models were omitted from the analysis leaving 7 models and 21 sites. Far West also had one site with no pre-service training (Fresno) but this does not affect the analysis because the site had been omitted previously.

TABLE 12

Pre-Service Training
Analyses of Variance Conditional upon Receiving Training
PV Classes Only

	Variable Name	Sources of Variance: Model	Site Within Model	Class Within Site	df Class
TQ#5	How long was the pre-service training	18.54% ***	42.25% ***	39.12%	60
TQ#6	Kind of pre-service training:				
	Demonstration lessons	13.04 NS	14.02 NS	72.94	64
	Lectures	26.92 ***	35.51 ***	37.56	64
	Individual meetings with leader	4.37 NS	23.01 NS	72.62	64
	Group discussions	9.30 NS	25.33 NS	65.37	64
	Discussion of video taped lessons	22.92 **	18.51 NS	58.57	64
	Observations	19.19 **	15.82 NS	64.98	64
	Role Playing	17.42 *	21.03 NS	61.55	64
TQ#6	Who gave training (person who was checked one or more times)				
	Sponsor Representative	7.46 NS	18.08 NS	74.46	64
	Consultant	21.41 **	11.21 NS	67.38	64
	Local HS Office	11.13 NS	29.41 *	59.46	64
	Other	9.95 NS	11.73 NS	78.32	64
TQ#6	Number of kinds of training given by:				
	Sponsor Representative	6.50 NS	34.53 **	58.97	62
	Consultant	10.02 NS	27.49 *	62.50	62
	Local HS Office	17.60 *	22.61 NS	59.79	62
	Other	17.78 **	36.99 ***	45.24	62

Note: The analyses in this table are similar to those presented in Table 8 with two exceptions: only teachers who reported that they received pre-service training are included, and as a result, the analyses are based on 7 models and 21 sites (Arizona and the Enablers are excluded).

TQ: Teacher Questionnaire

the high means for Kansas and Oregon. For types of training, Kansas has the first or second highest mean on all four variables for which there are significant differences between models, and Oregon is relatively high on two variables, but Florida and High Scope are also high.[18] Similarly, for consultant training, the one source of training with significant model differences, Kansas has the second highest mean (92% of the teachers indicate they received some training from a consultant) and High Scope (94% received training from a consultant), not Oregon, has the highest. The most striking finding of the re-analysis, however, is the reduction in the number of significant differences among sites and models. With the responses of teachers who received no pre-service training removed, the analyses of variance reveal significant differences among models on only five types and sources of training (as opposed to eleven significant variables with all teachers included) and only two significant differences among sites. Moreover, the level of significance was reduced for the significant variables also. Thus, the variation between models and among sites within models

---

[18] Lecture: Kansas, 1.0; Oregon, .91; Weikart, .80.
Discussion of video taped lessons: Weikart, .94, Kansas; .83
Observations: Kansas, 1.0, Florida, .91; Oregon, .82
Role play: Florida, 1.0, Kansas, .92; Weikart, .89;
Bank Street, .83

on the types and sources of pre-service training is reduced when only the responses of teachers who received pre-serv training are included.[19]

The analyses of PV-NPV differences on the types and sources of pre-service training variables, as originally defined (with all teachers included), show significant overall PV-NPV effects for half of the variables but very few interactions with sites or models (See Table 9). Thus, Planned Variation teachers report more group discussions, discussion of video taped lessons, observations, and role playing than do non-Planned Variation teachers as well as tending more often to receive some training from the sponsor representative and the consultant. For the variables with interactions, examination of site and means indicates that in some models more PV teachers than NPV teachers report receiving lectures and training from the local HS office,

---

[19]With people who received no training removed, it is possible to analyze the source of training variables in terms of the number of types of training given (this still is not an indication of how much training was given). Table 11 shows that under this condition, there are no significant differences between models in the number of types of training given by either the sponsor representative or by consultants (the sponsor representative gives 3.9 types of training on the average, and the consultant gives 2.9 types) but there is variation between sites within models on both. There are model-to-model differences in the number of types of training, given by the local HS office and by "other." Far West (4.8 types of training) and Kansas (3.6) have the highest means on the local HS office giving the training; and Far West (2.4) has the high mean for "other."

while in other models, more NPV than PV teachers make this report; and that in some sites PV teachers receive more training from "other" than do NPV teachers, while in other sites the finding is reversed. From these results, it appears that Planned Variation teachers tend to report a wider variety of training types and some training from sponsors and consultants than non-Planned Variation teachers do.

Summary of pre-service training: From Table 8, it appears that the major finding of this section is the large amount of variation among models and among sites within models in the length, types, and sources of pre-service training the teachers report. It cannot be concluded, however, that these differences are due entirely to the sponsors because the question on who gave training indicates that people other than the sponsor were involved. Table 9 shows overall PV-NPV differences on most variables. Since the PV mean is always higher than the NPV mean, this finding can be taken to show that Planned Variation teachers have more and a wider variety of pre-service training than do non-Planned Variation teachers. The means in Table 11, considering only PV teachers again, show that more teachers in the Oregon model than in the other models report that they received pre-service training. The teachers in Oregon and Kansas have the highest means on length of training (teachers

in Florida also reported long training) and on all types and sources of training. This suggests that the teachers in these models received a wider variety of training than teachers in other models. Of the eight variables on which there are significant differences between sites, Bank Street had large variations among its sites on 6 variables, Far West on 5, and Arizona on 4.

The major qualification to these findings is that all questions are dependent on whether the teachers received pre-service training. If we consider the types and sources of training received for only those teachers who report pre-service training, the differences between models and particularly the differences between sites within models are greatly reduced. Moreover, on those variables on which model differences remain significant, Kansas and Oregon do not always have the highest means; Florida and High Scope also tend to be high. Thus, we can conclude that the large differences between models and sites on aspects of pre-service training are due in part to the fact that some teachers report no pre-service training. This is not to say that it is more correct to only include the teachers who received pre-service training in the analyses, but simply that the conclusions drawn depend on the perspective taken.

Two other factors must be acknowledged in drawing conclusions about pre-service training. First, the low correlation between sponsors' and teachers' reports of the same variable cannot be ignored. This finding together with the large proportion of variance which is not explained by either models or sites[20] suggest that there may be different understandings as to what these questions about in-service training are intended to tap. Vague category labels may be part of the reason for different responses to the question of types and sources of training. This, then, is a second factor to be kept in mind in interpreting these findings. Because of these ambiguities, particularly in identifying who gave the training, it would be unwise to elaborate on the implications of the findings for individual variables. Thus, we can conclude there are differences between sites and models in pre-service training, but our knowledge of the nature of those differences is limited.

In-service Training:

In-service training may be more important to implementation in Planned Variation than pre-service training because it continues throughout the year rather than occurring for

[20]This finding has not been discussed previously, but can be found in Table 8 under the column headed "Class Within Site"; with two exceptions, over 60% of the variance on each variable lies between teachers in the same sites.

a limited time during the summer. Tables 13, 14, 15, and 16 parallel Tables 8 through 11 for pre-service training. Table 13 included analyses of PV classes only and the analyses are based on 9 models and 27 sites in the standard design. Table 14 presents the PV-NPV comparisons for the same variables, but with a smaller sample of 12 sites and 6 models. Table 15 shows analyses on comparable items from the Sponsor Implementation reports. Only variables with significant differences are included in any of these tables. Table 16 gives the means on these variables for the models included in the analyses; Table 2 in Appendix D contains the means and standard deviations for all sites and models.

Since essentially all teachers responded yes to the question, "As you were teaching during the year, was help and/or training available to you?" (TQ#8), no further analysis was done. There were differences, however, in the responses to the second part of the same question: "If yes, how often?" Table 13 shows significant differences among both sites and models. Examination of Table 16 indicates that model means range from a low of 1.0 for Oregon (training occurs on the average of once a week)[21] to a high of 4.9 for Kansas (every other month). The other models fall between training

[21] This question is scaled from 1=daily to 7=once this year, so that a high mean represents infrequent training.

TABLE 13

In-Service Training

Analyses of Variance for Planned Variation Classes

	Variable Name	Sources of Variance: Model	Site Within Model	Class Within Site	df Class
TQ#8	Frequency of in-service training	27.52%† ***	25.50% ***	46.98%	110
TQ#9	Kind of in-service training: Demonstration-lessons	10.29† *	20.83 **	68.88	135
	Lectures	13.06† **	20.38 **	66.57	135
	Group discussions	10.56† *	18.09 *	71.35	135
	Discussion of video-taped lessons	14.11† ***	20.45 **	65.44	135
	Observations	4.43† NS	22.87 **	72.70	135
	Role playing	11.39† **	30.10 ***	58.51	135
TQ#9	Who gave in-service training: (number of different types of training given by): Sponsor Representative	5.84† NS	18.73 *	75.44	129
	Consultant	12.26† **	19.73 **	68.01	129
	Local HS Office	5.30† NS	21.34 **	73.36	129
	Other	13.09† **	12.77 NS	74.14	129
TQ#9	No response (kinds of training not received)	8.42† NS	24.04 **	67.54	129

Note: † indicates that model effects are not significant when sites are used as random factors. See notes for Table 8 for further description of the table.

TABLE 14

In-Service Training
PV-NPV Comparisons

	Variable Name	PV-NPV	($\bar{X}$)	Model	Site	Model X PV-NPV	Site X PV-NPV	Class	df Class
TQ#8	Frequency of in-service training	1.97% NS		18.91% ***	31.77% ***	4.34% NS	9.56% *	33.45%	63
TQ#9	Kind of in-service training:								
	Demonstration lessons	1.94 NS		22.18 ***	14.11 **	7.61 *	3.98 NS	50.18	79
	Lectures	1.10 NS		27.59 ***	2.79 NS	6.31 NS	6.72 NS	55.50	79
	Group discussions	0.20 NS		5.69 NS	13.55 NS	9.02 NS	7.91 NS	63.64	79
	Discussion of video taped lessons	0.76 NS		15.02 **	2.76 NS	6.64 NS	11.31 *	63.51	79
	Observations	0.13 NS		6.32 NS	3.95 NS	14.32 **	9.67 NS	65.62	79
	Role playing	0.01 NS		15.31 **	13.45 **	11.94 **	5.64 NS	53.65	79
TQ#9	Who gave in-service training: (number of different types of training given by):								
	Sponsor Representative	5.85 **	PV=2.6 NPV=1.5	9.68 NS	9.61 NS	5.45 NS	8.57 NS	60.84	3
	Consultant	1.83 NS		8.06 NS	8.23 NS	3.89 NS	13.62 *	64.36	73
	Local HS Office	0.39 NS		19.72 ***	12.46 *	3.71 NS	7.51 NS	56.21	73
	Other	3.34 *	PV=.36 NPV=.90	9.52 NS	10.97 *	6.35 NS	10.31 NS	59.53	73
TQ#9	No response	0.15 NS		21.07 ***	3.33 NS	8.01 NS	7.36 NS	60.08	73

Note: See Table 9 for explanation of this table.

TABLE 15

In-Service Training

Analyses of Variance from Sponsors' Reports

	Variable Name	Sources of Variance: Model	Sites Within Model	df sites
SI#4b	In-service training			
	(1) Total no. of hours			
	For Teachers	89.8% ***	10.2%	12
	For Aides	90.2% ***	9.8%	12
	(3) How many days did the in-service training cover?			
	For Teachers	75.0% **	25.0%	12
	For Aides	70.8% **	29.2%	12
	(3a) Frequency of training:	78.7% **	21.3%	12

Note: For a description of the table, see the note for Table 10.

TABLE 16a

In-Service Training
Means for Models in Standard Design

	Variable Name	Far West	Arizona	Bank St.	Oregon	Kansas	Hi-Scope	Florida	EDC	Enablers
TQ#8	Frequency of in-service training:* (1=daily, 7=once this year)	2.7	3.0	3.6	1.9	4.9	3.2	3.3	4.2	3.5
TQ#9	Kind of in-service training received: (1=check, 0=no check)									
	Demonstration lessons*	.87	.82	.70	1.00	.93	.55	.58	.83	.64
	Lectures*	.92	.90	.38	.79	.80	.72	.50	.61	.83
	Group discussions*	.92	.96	.94	.95	.62	.92	.83	.89	1.00
	Discussion of video taped lessons*	.64	.68	.31	.64	.48	.59	.24	.28	.12
	Observations	.70	.63	.74	.73	.68	.62	.56	.50	.70
	Role playing*	.82	.26	.50	.61	.52	.63	.53	.22	.56
TQ#9	Who gave in-service training: (number of types of training given by)									
	Sponsor Representative	3.1	3.3	3.6	3.2	2.9	2.8	1.7	3.6	2.3
	Consultant*	1.7	2.7	3.3	3.4	2.7	4.7	2.8	1.5	3.1
	Local HS office	2.5	1.3	1.7	2.6	2.3	2.1	2.7	1.3	2.8
	Other*	.9	.1	.2	.3	1.7	.8	.0	.0	.6
TQ#	No response (kinds of training not received)	1.2	2.0	2.4	1.4	2.1	1.6	2.4	2.8	2.3

Note: * indicates significant differences between models; extreme means on those variables are underlined: single underlining indicates high means and double underlining indicates low means.

TQ: Teacher Questionnaire

TABLE 16b

In-Service Training
Model Means from Sponsor Reports

	Far West	Bank St.	Oregon	Kansas	Florida
SI#4b In-service training					
(1) Number of hours for teachers*	73	317	60	178	46
(2) Number of hours for aides*	73	263	60	178	46
(3) Number of days for teachers*	26	45	30	30	9
Number of days for aides*	26	38	30	30	9
Frequency of training (6= daily, 1=twice this year)	4.5	3.0	5.0	3.0	3.0
(4) Who gave the training?* (1= sponsor, 2=local)	2.0	1.0	1.0	1.0	1.0

Notes: Only means for models included in the analyses of variance are given.

*: indicates variables with significant model effects; on these variables extreme means are underlined, with high means having a single line, and low means a double line.

SI: Sponsor Implementation Report

monthly and twice monthly. There is also large variation among sites within the same model. The sites in the Kansas model show the largest variation with Portageville reporting that in-service training occurs between once a week and once every two weeks (2.5), Mounds reporting twice a year (5.3), and Oraibi reporting approximately once a year (6.8).

Surprisingly, there is no difference between PV and NPV on this variable: Planned Variation teachers, as a group, receive significantly no more or less frequent training than non-Planned Variation teachers (Table 14). There is a moderately significant PV-NPV by site interaction which appears to be due in part to one site (Johnston County) in which NPV teachers report more frequent training than PV teachers; in the other sites the PV mean tends to be greater than or equal to the NPV mean. The lack of overall PV-NPV differences, however, means that the differences among models in frequency of training are not necessarily due to the experimental treatment. They may instead reflect differences in the sites other than model intervention.

As with the pre-service training data, the sponsors' reports about the in-service training they gave (S1#6b(3)a)) do not correspond exactly with the teachers' accounts of how much they received (See Table 16). Although the data from the sponsors' reports also show significant differences among sites and models, the correlation between the two

sources on frequency of in-service training is only .52. Oregon is the only model for which there is agreement between sponsor and teachers as to how often training was given. Both sources report that training occurred once a week (2.0); this is the highest mean of all the models. The discrepancies in the other models suggest that there may be differences in definitions of in-service training.

The Sponsor Implementation Reports also give information on the total number of hours and days of in-service training given by the sponsors (SI4b(1) and (3)). Both variables show significant differences among models, with Bank Street reporting the highest number of hours of training given for both teachers ($\bar{X}$=317 hours) and aides (262) and the most days of training ($\bar{X}$ for teachers is 45; aides=38). Florida is the lowest on both variables with 46 hours and 9 days of training for both teachers and aides. The finding for the Florida model is understandable because the focus of their training is the Parent Educator rather than the teacher,[22] but is not consistent with the finding that the teachers in this model received a relatively high amount of pre-service training.

A second aspect of in-service training for which there are data is the kinds of training received (TQ#9). Table 14

---

[22]There are no PV-NPV comparisons for the Sponsor Implementation Reports.

shows that there are significant differences both among models and among sites within models on 5 of the 8 types of training list d in the questionnaire (demonstration lessons, lectures, group discussions, discussion of video taped lessons, and role playing) and differences among sites but not among models on a sixth type (observations). Examination of model means for this group of variables (Table 16) indicates that, unlike the findings for types of pre-service training, the same models do not have the highest means for all types of training: models which are high on one type of training may be low on another.[23] Thus, it appears that models vary in the types of training which teachers receive. Moreover, the means of the sites within models show that all models (except Oregon) have a great deal of variability among sites on at least one type of training. Kansas, however, is the only model with large site to site variation on 3 kinds of training: group discussions (Portageville, 100% of the teachers report this kind of training; Mounds, 25% of the teachers had it); discussion of video taped lessons (Portageville, 100%; Oraibi, 20%) and observations (Portageville, 100%; Mounds, 25%). This suggests that although there is variation in all models, that Kansas has the most variability between sites in the types of training which the teachers receive.

[23] It should be remembered that these figures do not indicate how much of a particular kind of training was received.

The lack of overall PV-NPV differences in types of training is not surprising because there is no reason to expect that all Planned Variation sites and models would be different from non-Planned Variation in the same way. We did, however, expect some differences between the experimental and the control treatments. It is surprising, then, that there are few PV-NPV by model interactions. Significant PV-NPV by model interactions are found for only demonstration lessons, observations, and role playing. The PV-NPV by site interaction is significant for only one type of training: discussion of video taped lessons. Hence, it appears that the lack of differences between the experimental and control groups on these variables are fairly uniform across models and sites.

A third aspect of training for which there are data is the source of training: who gave it? Four sources of in-service training are listed in the Teacher Questionnaire: sponsor representative, consultant, local Head Start office, and other (TQ89).[24] Analyses of variance (Table 13) show that there are no significant differences among models in the number of types of training given by either the sponsor representative or the local Head Start office. The means for

---

[24]These variables have the same problems of ambiguity as do the corresponding pre-service variables: they do not tell how much training each person gave. See footnote 15, p.72 for a more detailed discussion of those difficulties. There were no significant differences for "individual meetings": a high proportion of teachers in all sites report this kind of training.

all models, except Florida and the Enablers, range around 3 for the number of different kinds of training given by the sponsor representative. The means of the kind of training given by the HS office range around 2.5 for all models except SDC and Arizona (Table 16). There are significant differences, however, among sites within models on both of these variables. On the local HS office variable, three models have large variations among sites: Oregon (Tupelo, 0.2 kinds of training are given by the local HS office; E. St. Louis, 5.2), High Scope (Greeley, 0.3; Seattle, 4.3), and Far West (Buffalo, 1.4; Duluth, 4.0). For the sponsor representative, there are large variations between sites in four models: Far West (Duluth, 1.3 kinds of training are given by the sponsor; Buffalo, 4.0), Arizona (LaFayette, 1.4; Lakewood, 4.2), High Scope (Greeley, 1.3; Seattle, 4.7) and the Enablers (Bellows Falls, 0.2; Colorado Springs, 4.0). The finding of significant variation among sites within models but not among models on the local HS office variable suggests that there are local differences in the variety of training given which do not result from the treatment. The variation among sites but not among models in the types of training given by the sponsor representative suggests that the sponsors do not provide the same input to all sites within their models.

The number of kinds of training given by the consultant differs significantly among both models and sites within models. The teachers in the High Scope model report the most training from this source (4.7 kinds of training given by the consultant), while the teachers in EDC (1.5) and Far West (1.7) report the least. The largest amount of variation among sites is in the Arizona model (LaFayette, 1.5; Lincoln, 6.0), High Scope (Fort Walton Beach, 218; Greeley, 5.7) and EDC (Patterson, 0.0; Washington, 3.0) also have large amounts. The problem with interpreting these findings is that we do not know to whom "consultant" refers. Because of the range of people who might be included in this category, and because of the variation in responses, we suspect that the term has different meanings to different people. Without knowledge of those differences we cannot draw conclusions about this variable. This same difficulty holds for the "other" category where there are significant differences between models but not between sites. Kansas has the highest mean (1.7), but the implications of this finding are unclear because we have no notion of who is included in this category.

Thus, in looking at the four sources of training, we can conclude that there are substantial differences, either among models or among sites in the number of types of training given by each of these categories of people. On those categories which have significant model differences,

different models have high means; on those categories which have significant site differences, High Scope consistently shows large variation among sites[25].

Table 14 shows only three significant PV-NPV differences in the number of types of training given by these groups of people. As we would expect there is a main effect for the sponsor representative with PV teachers (2.6) having a higher mean than NPV teachers (1.5). What is surprising is that non-Planned Variation teachers reported any training from the sponsor since they are intended to constitute a control group which receives no training.[26] There is also a main effect for "other," but with the NPV mean (90% of the teachers checked "other") being larger than the PV mean (36%). Since we cannot interpret this category, it is reassuring to note that it is checked less often than the others. Finally,

---

[25]Also, the finding that the overall means, for all sources of training, except "other" are approximately the same--between 2 and 3 types--indicates that the sponsor may not be the sole source of training. This supports the contention that teachers may be reporting non-model training in these questionnaires. This conclusion is further supported by the sponsor implementation report (SI#4b(4)), Table 16, which shows that only Far West reports that the in-service training for the model was given by the local staff; in the other models it was given by the sponsor's staff.

[26]Despite the overall PV-NPV relationship, in two sites (E. St. Louis and Mounds), the NPV mean is actually higher than the PV mean on this variable. This suggests either that the teachers are confused as to who is included in the category of "sponsor representative" or that the contamination of experimental and control conditions in these sites is tremendous.

there is a site by PV-NPV interaction for the consultant: in 7 sites, PV teachers report more kinds of training from the consultant than do NPV, and in 5 sites, the relationship is reversed.

There is one additional variable which can be developed from the question of what kinds of training were received and who gave them. This variable, no response, indicates the number of types of training not checked as being given by anyone, and if the order of the means are reversed, it gives a ranking of sites and models as to where the most types of training were checked. A high mean in this variable, then, means that several types of training were not given. The analysis reveals no significant model differences for this variable (Table 13), indicating that no model gives a substantially wider variety of training than another. There also are no PV-NPV differences (Table 14). There are, however, significant differences among sites within models. Three models have particularly large variations among sites: Arizona (Lincoln, 0.6 types of training not checked; LaFayette, 3.2), Kansas (Portageville, 0.8; Mounds, 3.5), and Gordon (Houston, 0.7; Jonesboro, 3.7).

Summary of in-service training: Since no one or two models were consistently high on all in-service training variables, the model to model differences (Tables 13 and 15) suggest that teachers in different models receive different training

in terms of frequency, types, and sources. Oregon reports the most frequent in-service training and tends more often than other models to have demonstration lessons and group discussions. Bank Street reports the most hours and days of training and seems to avoid lectures. Kansas gives more demonstrations than other models, but fewer group discusssions. Far West and Arizona give more lectures and discussions of video tapes, and Far West also does more role playing. There are no significant model effects on the number of types of training given by the sponsor or by the local HS office, nor are there model differences in the variety of training received (number of types of training not checked). The finding of training being given by people other than the sponsor indicates that the teachers are receiving non-model training.

The significant site effects emphasize the importance of recognizing that teachers within the same model report very different training experiences. From this data, however, it is not possible to determine whether the differences result from differential sponsor input to the sites or from non-sponsor input.

The failure to find significant PV-NPV differences (main effects or interactions by models) on over half the in-service variables is important because without differences between experimental and control groups, we cannot responsibly

draw inferences about the effects of models. These findings may be explained in two ways. The lack of differences may be the result of contamination: non-Planned Variation teachers as well as Planned Variation teachers are being exposed to model training. The finding that NPV teachers report training by the sponsor supports this view. Or, the effects of model and site differences for PV only may result from differences in the sites rather than from differences in model treatment. In this case, we would expect either that PV-NPV differences would not occur, or that if they did occur, they would be unique to the site. If this explanation is correct, it may also be true that model-related treatments do not vary from site to site.

In using this information on in-service training, it must also be remembered that it has the same problems that were present for the pre-service variables. In addition to the non-model input mentioned above, we must acknowledge the limitations which result from vaguely labelled categories and from lack of information on how much of each kind of training was given. Because of the problem of ambiguous categories, it is possible that some of the variables, particularly the sources of training, have different meanings for different teachers. This possibility is supported by the finding of only moderate correlations between sponsor and teacher reports, and by the large proportion of

variation which lies among teachers in the same site and is not explained by site or model effects. This problem indicates that we must be cautious in drawing specific inferences from this data. The lack of information on how much of each type of training was given is limiting because it precludes conclusions about the relative importance of differences in the kinds of variables.

The most striking findings from this group of in-service variables, then, are the large number of significant differences between sites and models (although they are fewer than for the pre-service variables), and the relative absence of PV-NPV differences.

Continuing Support and Feedback:

The frequency and types of training given are not the only aspects of training which may vary within Planned Variation. One additional aspect may be loosely defined as continuing support and feedback. In identifying this area, we are proposing that formal training sessions may not be a sufficient basis for implementation of a model program. Teachers may also need help with specific problems and questions, support to bolster their confidence, and individual feedback on their performance. This aspect of training, then, will considered in this section.

In addition to data from the teachers' reports, there are four items on support and feedback in the sponsor and

the consultant reports.[27] The format for presenting the data is similar to that of the previous sections: Table 17 presents the analyses of variance for the Planned Variation classes, Table 18 shows PV-NPV comparisons, and Table 19 shows the means on these variables for the models included in the analyses. Table 20 presents the correlations between the support variables.

The first question presented in Table 17 is, "To whom do you go (most often) for help and information in implementing the program?" (TQ#12). Of the four response categories analysed,[28] three show significant differences among models and among sites within models.[29] On the "Sponsor

---

[27]One variable from the Teacher Questionnaire, satisfaction with training, which is included in this section is more appropriately described as a reaction to the training rather than as an indicator of what support was given. When we examine our theory of implementation, these two dimensions will be separated. For purposes of data presentation, however, it is more instructive to combine them.

[28]The original question has a fifth category, "no one available." Since only 2 responses of a total of 448 were entered under this heading, it was dropped from further analyses. The "other" category was analyzed but showed no significant differences.

[29]Again we are faced with the problem of vague categories. In most models, primary responsibility for model implementation is given to a local person who is trained by the sponsor and who works full time in the site. It is not clear, however, where such a person was considered in this question. It is possible that some teachers included her in the sponsor representative category while others may have included her as an assistant director or even as "other."

TABLE 17

Continuing Support and Feedback
Analyses of Variance for Planned Variation Classes

	Variable Name	Sources of Variance: Model	Site Within Model	Model Within Site	df Class
TQ#12	To whom do you go most often for help in implementing the program?				
	Sponsor Representative	11.84%† **	24.36% ***	63.79%	135
	Head Start Director	8.97† NS	9.58 NS	81.46	135
	Another teacher	10.70† **	22.15 *	67.15	135
TQ#13	Did trainers stay long enough to be really helpful?	12.23† **	17.21 *	70.56	126
TQ#16	How satisfied were you with training offered?	16.20 **	11.06 *	72.75	127
SI#3m	Sponsor feedback to teachers	42.1 *	57.9		(df Site) 12

Note: SI: Sponsor Implementation Report

† indicates that model effects are not significant when sites are used as random factors. See notes for Table 8 for further description of the table.

TABLE 18

Continuing Support and Feedback

PV-NPV Comparisons

	Variable Name:	Sources of Variance: PV-NPV	(X̄)	Model	Site	Model X PV-NPV	Site X PV-NPV	Class	df Class
TQ#12	To whom do you go most often for help in implementing the program:								
	Sponsor Representative	5.42 *	PV=.17 NPV=.03	8.24 NS	4.55 NS	13.44 **	4.55 NS	63.79	79
	Head Start Director	13.45 ***	PV=.15 NPV=.51	1.70 NS	10.12 *	7.48 NS	13.75 **	53.50	79
	Another Teacher	0.87 NS		7.58 NS	7.82 NS	23.91 ***	7.82 NS	52.01	79
TQ#13	Did trainers stay long enough to be really helpful?	0.57 NS		18.44 ***	19.69 ***	2.18 NS	10.88 *	48.23	76
TQ#16	How satisfied were you with the training offered?	1.85 NS		13.01 **	16.37 ***	7.79 *	12.86 **	48.11	78

Note: See Table 9 for explanation of this table.

TABLE 19

Continuing Support and Feedback
Means for Models in Standard Design

		Far West	Arizona	Bank St.	Oregon	Kansas	High Scope	Florida	EDC	Enablers
TQ#11	Have you personally requested help or training? (1=yes, 2=no)	1.31	1.06	1.08	1.08	1.08	1.00	1.00	1.06	1.23
TQ#12	To whom do you go most often for help in implementing the program? (1=check, 0=no check)									
	Sponsor Representative*	0.32	0.25	0.00	0.33	0.37	0.06	0.00	0.17	0.13
	Head Start Director	0.13	0.36	0.41	0.13	0.07	0.00	0.10	0.33	0.23
	Another teacher*	0.13	0.16	0.18	0.00	0.25	0.06	0.41	0.00	0.06
	Other	0.03	0.17	0.09	0.13	0.00	0.36	0.24	0.22	0.07
TQ#13	Did trainers stay long enough to be really helpful? (1=yes, 2=no)	1.17	1.10	1.00	1.00	1.20	1.39	1.42	1.17	1.07
TQ#16	How satisfied were you with the training offered? (1=very satisfied, 5=very dissatisfied)	2.66	2.10	2.59	1.78	1.67	2.48	2.84	2.17	1.60
TQ#46	Would you choose PV next year? (1=yes, 2=no)	1.10	1.03	1.11	1.31	1.00	1.00	1.05	1.17	1.00

Note: * indicates significant differences between models; extreme means on those variables are underlined: single underlining indicates high means and double underlining indicates low means.

TQ: Teacher Questionnaire

Representative" variable, Bank Street and Florida ave the lowest means with no teachers going to the sponsor most often for help. Far West, Oregon, and Kansas have the highest means (32%,33%,38% respectively), but even in these models, only one-third of the teachers report that the sponsor representative is the most frequent source of assistance. There is also large variation among sites within models on this variable. Although Arizona and Oregon have the largest variation among sites (Arizona: Lakewood, 75% of the teachers go to the sponsor rep, Lincoln and LaFayette, 0% go to the sponsor; Oregon: Tupelo, 75%, E. Las Vegas, 0%), five additional models (or a total of 7 of the 9 models in the analysis) have at least one site in which no teachers check the sponsor representative on this variable. These findings are surprising. The low model means are contrary to the expectation that the sponsor representative would serve as the primary authority for implementation. In part, however, the model means are misleading because of the large variation among sites. This latter finding is perhaps more interesting because it suggests either that sponsors are varying their input from site to site, or that their input is the same but the teachers are perceiving it differently.

The other source of help with significant effects for both sites and models is "Another Teacher." Considering the differences among models, Oregon and EDC have the lowest means with 0% of the teachers going to another teacher most often for help, while Florida has the highest with 41%. Kansas and Florida have the highest variability among sites (Kansas: in Portageville, 75% of the teachers go most often to another teacher for help, in Mounds and Oraibi, 0% do; Florida: in Jonesboro, 67% go to another teacher, in Houston 0% do).

There are no significant model or site effects for choosing the Head Start director but the variable was included because of PV-NPV differences. The overall mean for this variable indicates that 20% of the Planned Variation teachers go to the Head Start director most often for help. This is approximately the same as the proportion which go to the sponsor representative most (18%) and slightly more than the proportion who go to another teacher (14%) or to other (15%).

The PV-NPV comparisons on these variables show overall PV-NPV effects for the sponsor representative and the Head Start director, and interactions with sites or models for the sponsor representative, the Head Start director and another teacher (Table 18). For the sponsor representative, the overall effect indicates that PV teachers tend more often to go to the sponsor most frequently than NPV teachers, which is to be expected, although the PV mean is not large.

On the other hand, the PV-NPV by model interaction on this variable indicates that this relationship does not hold for all models. The PV-NPV analysis for another teacher shows no overall effect but a strong interaction with models. The Head Start director variable shows a strong overall PV-NPV effect (PV=15% of the teachers go to the Head Start Director most often, NPV=51%), which might be explained in terms of non-Planned Variation teachers having fewer alternative sources of help than Planned Variation teachers. At the same time, there also is a PV=NPV by site interaction which complicates interpretations.

It is apparent, then, that there is a great deal of variability in whom teachers go to most frequently for help in implementing the program. It also appears that overall, teachers are as apt to go to the Head Start director as to the sponsor representative. PV-NPV comparisons show that PV teachers go more often than NPV teachers to the sponsor representative and vice-versa for the Head Start director although there are interactions with sites or models on both variables.

Another question which is relevant to the kind of support which teachers are getting is "Did those who trained you stay long enough to be really helpful to you?" (TQ#13). The analysis of variance on PV classes shows significant differences among both models and sites within models

(Table 17). Examination of the model means (Table 19) indicates that overall, teachers tend to respond yes, the trainers did stay long enough. In the Florida and High Scope models, however, 42% and 40% of the teachers, respectively, report that their trainers did not stay long enough. This suggests that, according to this measure at least, support is inadequate in these models. The variation within models is highest in Kansas (Portageville and Mounds, 0% of the teachers say trainers did not stay long enough; Oraibi, 60% say no) and High Scope (Fort Walton Beach, 0% no; Seattle, 67% no). This indicates that trainers are judged differently in different sites. Whether this is due to differential input or to different receptions is not clear. Nor is it clear that the trainer referred to is always the sponsor or his representative. Moreover, there is almost no overall PV-NPV difference on this variable (Table 18). Again, this raises problems for interpreting model differences (i.e., do no PV-NPV differences stem from contamination, or are model differences the result of factors unrelated to the model?). The moderate interaction with the site factor suggests that the PV-NPV differences which do exist are the result of local site conditions rather than of Planned Variation treatments.

Sources other than the Teacher Questionnaire provide data on support and feedback. Both the sponsors and the OCD consultants completed site assessments containing two relevant items: availability of sponsor and sponsor feedback to the teacher (items 3-1 and m i. the sponsor reports and items 13 and 14 on the consultant forms). These are the most direct measures of the dimension we would like to tap, and they provide perspectives different from those of the teachers. Analyses of variance, however, show significant differences among models for only the sponsor's rating of the amount of feedback they give to the teachers. Table 19 shows that the high mean on this variable is for Bank Street (5.0, high frequency).[30]

An additional support variable, taken from the question, "During the past year have you personally requested help or training?" (TQ#11), shows no significant differences among models, sites, or PV-NPV. Essentially all teachers, both PV and NPV responded yes to this question; the overall

---

[30]Differences between sites cannot be tested because only one observation per site was made, and sites, consequently, are the basic unit of analysis. Using the site as the basic unit also means that the number of degrees of freedom in these analyses is small. In such a situation it is difficult to determine whether the lack of significance on the other variables is due to a real lack of differences between models or to the low power of the tests (power (1-B) is the probability of finding out that the null hypothesis is wrong). No PV-NPV comparisons are available for this data.

proportion for PV teachers is 90%. The variable is not included in the analysis of variance tables, but the model means are shown in Table 19.[31]

We have presented a number of different variables which were intended to measure continuing support and feedback as a dimension of sponsor input. Given this intention, we would expect the variables to be highly correlated. In fact, however, they are only moderately correlated. A discussion of these correlations can be found in Appendix C, Table 3.

Finally, there is a related but distinct variable which can be included in this section: "In general, how satisfied are you with the training offered you during the year?" (TQ#16; 1=very satisfied; 5=very dissatisfied). This question is probably best interpreted as a general reaction to training rather than as a specific indicator of support and feedback, but is interesting in the present context. Table 17 shows significant differences among models but not among sites within models on this variable. The teachers in Enablers ($\bar{X}$=1.6), Kansas (1.7), and Oregon (1.8) have the highest mean levels of satisfaction. The

---

[31] Although most teachers reported requesting help, we hypothesized that there might be differences among models or sites in the number who actually received help. To test this expectation, we attempted to analyze the number of times teachers received help as a proportion of the number of times they requested help (TQ#11 b÷a). The analyses, however, failed because a number of teachers responded inappropriately: they recorded that they received help more times than they requested it.

other models are also fairly high; Florida which has the lowest mean (2.8) is in the range of "somewhat satisfied." Comparison of PV and NPV means, however, shows that there are no differences in the overall means. The finding of no overall PV-NPV effect indicates that Planned Variation teachers as a group are not more satisfied with their training than non-Planned Variation teachers (Table 18). The significant interactions indicate that in some models and sites, non-Planned Variation teachers are actually more satisfied with their training than are Planned Variation teachers.

Summary of Continuing Support and Feedback: In this section an attempt was made to tap an aspect of training other than straightforward accounts of the types and extensiveness. One important finding which resulted is that there are differences among models and even larger differences among sites within models in the number of teachers who go to the sponsor representative most often for help in implementing a program. Because the sponsor representative would be expected to be the authority on the model, it is surprising to find that in some sites no teachers go to the sponsor rep most often and that, overall, teachers are as apt to go to the Head Start director most frequently as to the sponsor rep. Moreover, the PV-NPV by model interaction on this variable indicates that in some models PV teachers do not go to the

sponsor rep more often than NPV teachers. These findings may be partially explained by vague category labels, and by the fact that the Head Start director is always present at the site.

Another interesting finding is that the teachers are generally satisfied with their training. There are significant differences among models on this variable, but even in the model with the lowest mean, training is rated as somewhat satisfactory. There are, however, no overall PV-NPV differences in satisfaction.

The Florida model is extreme on more of the variables analyzed than the other models. We find that it is characterized by teachers not going to the sponsor representative for help most often, but instead, going to another teacher (although there is high variability within the model on the latter variable). Moreover, Florida is low on satisfaction with training and high in judging that trainers do not stay long enough to be really helpful. We might conclude, then, that the teachers in the Florida model do not receive strong support from the sponsor. This is consistent with the model's emphasis on the role of parent educators rather than teachers. Other models do not present as clear a picture, but are extreme on several measures: in Bank Street, the sponsor reports that she gives a great deal of feedback to the teachers but none of the teachers go to

the sponsor representative for help most frequently. The teachers in Kansas and Oregon report that they are quite satisfied with their training (although in Oregon, the satisfaction of PV teachers is not greater than that of NPV teachers) and more of them than in other models go to the sponsor for help most frequently (Oregon also has high variability among sites on this variable). Moreover, Oregon has a low mean on going to another teacher most frequently, and Kansas has large differences among sites in the number of teachers who go to other teachers for help and in whether the teachers felt that the trainers stayed long enough to be really helpful.

In interpreting these findings, two additional factors must be considered. First, low correlations between the support measures suggest that the data may be unreliable, and second, the lack of overall PV-NPV differences indicates that Planned Variation teachers do not receive more support than non-Planned Variation teachers in all cases.

Sponsor Input Variables as Determinants of Implementation:

In attempting to explain the variation in levels of implementation, we propose that the variables just presented, the length, types and sources of training as well as support factors, will be important. More specifically, we expect that more training, both in-service and pre-service, and more support will be related to more successful implement-

ation.[32] While we cannot make similar simple statements about the relation between kinds and sources of training and levels of implementation, we do not feel they are unimportant: we expect that successful implementation is related to an interaction of the "right" type of training at the "right" time; with determination of right depending, no doubt, on the people involved and their stage of training. We also expect that more training from the sponsor should lead to better implementation because the sponsor knows the model best. But merely giving more types of training, as is reported here, is not necessarily related to implementation. Finally, we assume that training given by the local Head Start office or by a consultant may influence implementation, but without further information about the nature of the training, it is not clear whether it would be a help or a hindrance. We predict, then, that if these variables in part determine the extent of implementation, then variations in them may help to explain variation in levels of implementation.

But while these variables may be partial determinants of implementation and while there are more data available for sponsor input than for the other areas to be discussed, we would argue that additional data on sponsor input is

---

[32]Terms such as successful or better implementation are used synonymously with a higher level of implementation; all are defined in terms of the objective of fully implementing the model.

necessary for an adequate study of implementation. One reason is the previously discussed problem of vagueness and unreliability in the present measures. Second, there are additional dimensions of sponsor input which may affect implementation but which have not been studied. Since we predict that these dimensions vary among sites, we expect that they may help explain variations in levels of implementation. The need for studying these additional aspects emerges from the present findings. First, the data suggest that input varies, but beyond the clue that sponsors do not give all the training, it does not indicate why. We do not know whether the input made by people other than the sponsor's staff, or non-sponsor input, accounts for all the variation (with sponsor input remaining the same to all sites), or whether part of the variation is the result of sponsors changing their training from site to site. Second, we do not have full knowledge about who is giving the training. We do not know how the local sponsor representative fits in; nor do we know how much of the total training in a site is given by non-sponsor people, or what kind of training they are giving.

These questions suggest two broad areas of study. One area is the sponsor's organization, or staffing pattern. This concept includes the responsibilities and expertise of the staff involved in transferring a model from the sponsor to the teachers in a Head Start site, and the relationships between them. Anecdotal evidence from consultant reports and

from conversations with some of the sponsors underlines the importance of this concept to implementation.

One important position in the path between the sponsor and the teachers is the member(s) of the sponsor's central staff who visits the sites. Beyond the general conditions that the sponsor's staff or representatives do visit, other factors vary, sometimes within models as well as between them. One such factor is the size of the sponsor's staff: the number of people responsible for implementing a model at the sites. Another factor which varies is arrangements with the sites: the frequency of sponsor visits, and the decision as to whether a site is visited always by the same person or by a rotating staff. A third, less tangible, but perhaps more important, factor is the role of the sponsor's representatives. Variations in the sponsor's role were discussed in Chapter 1; anecdotal evidence suggests that in some sites, the sponsor's representative may be concerned only with working with the teachers to implement the model in the classroom, while in others he may become involved in a much wider range of problems. We might predict that in situations where the sponsor's representative was more involved, the teachers be more responsive to the demands of the model and might work harder to implement it. If this, or some other relationship with implementation exists, then variation in this factor might help explain variation in levels of implementation.[33]

---

[33]The issue of variation in the sponsor's role is important

Another key position in the path between the sponsor and the teachers is that of the local person who is reponsible for helping the teachers with the model on a daily basis. A person in this capacity is present in most models, but the specific arrangements for the position vary. In some models, the local sponsor representative is given special training, while in others she receives the same training as the teachers. In some cases, the local representative is the primary liaison with the model, with the teaching staff having little direct contact with the sponsor's staff, in others the sponsor's staff provides the bulk of the training directly to the teachers and the local representative provides support between sponsor visits. The importance placed on this position also varies from site to site. In some sites, working with the model is the local representative's only responsibility. In others, the model is only one duty among many. We would expect variations such as these to have an impact on implementation. If this link between the sponsor and the site is weak, implementation will suffer. One OCD consultant asserts, for example, that the local representative

> "is weak and needs strong support from the modeller.... Teachers are not given consistent help in deepening their understanding of the model....The modeller does not have field support."

---

not only to explain variations in levels of implementation, but also to define the treatment: if two models vary on dimensions which might affect child outcome measures, then those variations should be documented.

Or, in an even more extreme situation, an OCD consultant reports his conversations with teachers after the local representative has returned from a week's training session at the sponsor's home base:

> "When I asked what sort of follow-up the teachers had, two could remember none while one teacher could vaguely remember 15 minutes being devoted to it at one meeting but couldn't remember what was covered."

The consultant concludes that "...training carry-over does not seem to be being accomplished."

It would seem that in models where the sponsor's staff attempts to train the teachers directly, the loss from a weak local representative would be on-going support and the answers to practical, day-to-day questions. We might predict that in this situation implementation would be difficult, but not impossible. On the other hand, in the models where the local representative receives the major portion of the sponsor's training, and has primary responsibility for training her staff, an ineffective person will be disastrous for implementation.[34]

The importance of the concept of the sponsor's staffing pattern can be summarized in a statement by one of the OCD consultants about the problems which a breakdown in this com-

---

[34]Another person who may not be directly involved in the training path between the sponsor and the teachers, but who is probably crucial to implementation, is the Head Start Director. The role of the Director will be examined in the section of this chapter on Context Variables.

ponent creates:

> "The major problem [this site] is facing -- and yet which is the key to successful implementation of this type of model -- is lack of communication from top to bottom."

And in another report, the same consultant commenting on the same site states that "The communication path seems to stop before getting to the teacher." Variations in this concept, then, may explain variation in levels of implementation.

A second area for which the data are inadequate is inputs from people other than the sponsor's staff. Since our data indicate that training from people other than the sponsor occurs, this area should be examined. While it is possible that these non-sponsor inputs augment the sponsor's training in the model, it is also possible that they are totally unrelated, and perhaps even confuse the sponsor's training. It is important, then, to determine who the non-sponsor people are, how much of this additional input is made to the teachers, and what the content is. In relation to the last question, we are further interested in whether the input is formal training or assistance with specific problems; whether training is in one special area, such as music, which might not be essential to the model, or whether it covers broad approaches to teaching; and finally, whether it is consistent or inconsistent with the model. If training from someone other than the sponsor, such as the local Head Start office or an RTO consultant, is inconsistent with that

from the sponsor, for example, we would expect implementation to suffer.

Conclusions on Sponsor Input:

From the data available on pre-service and in-service training, and on limited aspects of continuing support and feedback to the teachers, it appears that sponsor input varies among models and among sites within models. Moreover, there are fewer PV-NPV differences than differences among sites and models. Many of the differences which do exist are interactions that indicate that different models, and sometimes different sites within the same model, show different relationships between experimental and control classes.

These findings support the argument that implementation is a more complex process than had been imagined originally. We have discussed the need for a broader definition of input in explaining variation in levels of implementation in terms of both an expanded view of sponsor input and a recognition of non-sponsor inputs. While we emphasize the importance of these additional areas in an adequate study of implementation, we believe that they are not the only factors which should be considered. We have argued that the process of implementation is interactive, with sponsor input being only one aspect. We turn now to the second category of factors which may affect model implementation.

LOCAL STAFF REACTION AND INPUT

The variables to be presented in this section are intended to provide information on the personal background and professional experience of the local teaching staff in Planned Variation. From one point of view, we are interested in these variables descriptively, simply for what they tell us about the types of teachers and aides working with the models and whether the Planned Variation staffs are different from non-Planned Variation staffs.

From another point of view, we are interested in exploring these variables in terms of whether they explain variation in levels of implementation. We have asserted that implementation is an interactive process which is influenced by the local staff as well as by the sponsor. We expect that model input will not be effective unless the model is adopted and used by the staff. The determination of whether the model is adopted depends in part on staff input. The skills, preferences, and previous teaching methods of the local staff may affect levels of implementation. In part, adoption of the model also depends on the staff's reaction to it. Liking the model and the people on the sponsor's staff, and therefore being willing to make the effort of trying to work with the new program will undoubtedly affect implementation.

Our expectations about the relationship between staff factors and implementation also assume that the models represent new ways of teaching for the majority of teachers in Planned Variation. As a result, implementation will be a difficult process because of lack of mastery and knowledge of the new skills and techniques necessary for full implementation; and, in some cases, because of the need to break old habits. The Consultant Reports support this assumption. One consultant, for example, states that "...the pull of the traditional seems to wipe out the new model." And another concludes her discussion on the difficulties of working with a new program with an observation: "...I could not help but wonder, what is the toll on teachers of this model?" Thus, it seems reasonable to assume that implementing a new model is difficult. At the same time, our hypothesis about the importance of previous skills and preferences suggests that implementation will be less difficult for some teachers than for others. As with the sponsor input, we find that this category has many dimensions, which do not all influence implementation in the same way.

In order to examine the staff background variables from both viewpoints, we will first present data to provide a description of the staff in this study. Second, we will examine the same variables as determinants of levels of implementation and will suggest other areas of study. In

addition, we will consider the implications of our findings for the design of the Planned Variation study.

Data:

The data presented in Tables 20, 21 and 22 follow the organization of the previous sections. Table 20 summarizes the unweighted means analyses of variance for the standard balanced design of PV classes only; significance levels and allocations of variance are given. Table 21 shows the analyses for the comparison of Planned Variation and control classes, again using balanced designs but in this case involving only 6 models and 12 sites. It should be remembered that the model and site effects in this table are based on PV and NPV classes combined, and therefore are difficult to interpret. Table 22 shows the means for the models included in the analyses of the PV classes. For variables with significant model differences, the extreme means are underlined. The means and standard deviations for all sites and models on these variables can be found in Appendix D, Table 4.

We will not examine the specific variables contained in these tables in detail as we did in the previous sections, because they are not directly affected by the experimental treatment. The major impression to be drawn from the tables is that there are large differences among models and sites

TABLE 20

Staff Background

Analyses of Variance for Planned Variation Classes

	Variable Name	Sources of Variance: Model	Site Within Model	Class Within Site	df Class
Teacher:					
TQ#33	Live in the neighborhood where most of the children live	21.58% ***	17.20% **	61.22%	133
TQ#39	Circle the highest grade completed:	22.09 ***	17.69 **	60.22	130
TQ#40	Check any of the following which you have had:				
	Early childhood development course	14.74† ***	24.55 ***	60.71	135
	Nursery School teaching course	10.93† **	21.57 **	67.50	135
	Kindergarten first or second grade course	10.34 *	9.31 NS	80.35	135
TQ#41	Do you have a state or city teaching certificate?	19.15† ***	26.71 ***	54.14	130
TQ#42	How many years of teaching experience have you had in H.S.?	8.67† *	33.24 ***	58.09	130
TQ#38	Please check your ethnic group:	13.82† **	21.94 **	64.24	129
TQ#43	How did you happen to teach in this center rather than another?	11.63† **	19.87 *	68.50	128
TQ#45	How did you choose to participate in PV?	13.30† **	27.80 ***	58.90	107
Aide:					
AQ#21	Circle the highest grade completed:	14.12† **	19.45 *	66.43	124

Note: † indicates that model effects are not significant when sites are used as random factors. See notes for Table 8 for further description of the table.

TABLE 21

Staff Background
PV-NPV Comparisons

	Variable Name	PV-NPV	(X̄)	Model	Site	Model X PV-NPV	Site X PV-NPV	Class	df Class
Teacher:									
TQ#33	Live in the neighborhood where most of the children live	0.29% NS		23.82% ***	2.23% NS	4.04% NS	8.53% NS	61.10%	77
TQ#39	Circle the highest grade completed	1.50 NS		16.24 ***	12.95 **	6.17 NS	18.42 ***	44.72	75
TQ#40	Check any of the following which you have had:								
	Early childhood development course	5.66 **	PV=.57 NPV=.81	20.12 ***	2.65 NS	1.51 NS	7.77 NS	62.29	79
	Nursery School teaching course	0.36 NS		11.13 *	1.92 NS	9.30 NS	4.82 NS	72.48	79
	Kindergarten first or second grade	2.41 NS		6.25 NS	11.73 *	1.55 NS	11.54 *	66.52	79
TQ#41	Do you have a state or city teaching certificate?	4.56 *	PV=1.62 NPV=1.40	20.89 ***	7.41 NS	4.58 NS	10.58 *	51.98	76
TQ#42	How many years of teaching experience have you had in H.S.?	0.00 NS		21.78 ***	8.25 NS	4.82 NS	3.48 NS	61.66	74
TQ#38	Please check your ethnic group	0.01 NS		9.49 NS	17.35 **	4.83 NS	4.82 NS	63.50 NS	76
TQ#43	How did you happen to teach in this center rather than another?	0.61 NS		11.15 *	7.60 NS	1.99 NS	10.44 NS	68.21	74
Aide:									
AQ#21	Circle the highest grade completed	1.70 NS		7.91 NS	8.65 NS	9.24 NS	4.29 NS	68.21	77

Note: See Table 9 for explanation of this table

TABLE 22

Staff Background

Means for Models in Standard Design

	Variable	Far West	Arizona	Bank St.	Oregon	Kansas	High Scope	Florida	EDC	Enablers
Teacher:										
TQ# 33	Live in the neighborhood where most of the children live* (1 = yes, 2 = no)	1.79	1.42	1.79	1.52	1.32	1.42	2.00	2.00	1.61
TQ# 36	Age (# of years)	30.9	32.1	29.6	36.5	36.1	38.6	32.2	34.6	38.2
TQ# 39	Circle the highest grade completed:*	15.4	15.0	15.4	13.8	13.5	14.7	16.2	13.9	13.9
TQ# 40	Check any of the following which you have had: (1 = check, 2 = no check)									
	Early childhood development course *	0.76	0.70	0.76	0.87	0.85	0.65	0.24	0.67	0.89
	Nursery School teaching course*	0.76	0.36	0.73	0.35	0.67	0.37	0.24	0.56	0.44
	Kindergarten first or second grade course*	0.61	0.50	0.37	0.31	0.53	0.55	0.53	0.00	0.44

TABLE 22
(con't)

	Variable	Far West	Arizona	Bank St.	Oregon	Kansas	High Scope	Florida	EDC	Enablers
TQ# 41	Do you have a state or city teaching certificate? (1 = yes, 2 = no)*	1.33	1.63	1.69	1.92	2.00	1.46	1.38	1.67	1.79
TQ# 42	How many years of teaching experience have you had in HS? (# of years)*	3.1	3.2	2.6	3.7	4.0	3.5	2.4	4.1	3.2
TQ# 38	Please check your ethnic group: (1 = white, 0 = nonwhite)*	0.74	0.49	0.59	0.22	0.42	0.49	0.51	0.06	0.67
TQ# 43	How did you happen to teach in this center rather than another?* (1 = assigned, 2 = volunteered)	1.42	1.74	1.62	1.61	1.53	1.44	1.32	1.06	1.40
TQ# 45	Did you choose to participate in PV?* (1 = yes, 2 = no)	1.58	1.24	1.28	1.17	1.20	1.00	1.38	1.44	1.53
++	Number of years in PV (% second year)	0.50	0.38	0.36	0.70	0.31	0.54	0.57	0.50	0.60
Aide:										
AQ# 21	Circle the highest grade completed:*	11.7	12.0	11.0	12.3	10.6	10.6	10.4	11.8	11.7

Note: * indicates significant differences between models; extreme means are underlined.
++ number of years in PV obtained by comparing lists of teachers in 69-70 and 70-71.
TQ: Teacher Questionnaire AQ: Aide Questionnaire

on staff background characteristics, and relatively small PV-NPV differences.[35]

A brief consideration of two variables is useful in demonstrating how the data can be interpreted. For the teacher certification variable (TQ#41), there are strong effects for both sites and models. The model means range from Far West, where 77% of the teachers have their teaching certificates, to Kansas where none of the teachers do. Arizona has the highest variability within models with all teachers in Lincoln having their teaching certificates, and with none of the teachers in Lakewood having theirs. There is a moderate PV-NPV effect for this variable, which is reflected in the finding that, overall, 60% of the non-Planned Variation teachers are certified, while only 38% of the Planned Variation teachers are. There is also a moderate site within model interaction with PV-NPV, which indicates that the relationship between PV and NPV teachers

---

[35]This finding does not hold for the teacher aides. Analyses of their questionnaire responses reveal significant effects only for level of education; age, years of Head Start experience, ethnic background, and place of residence (whether aide lives in the same neighborhood as her students) show no significant differences among sites and models or between PV-NPV. In addition, some teacher variables also show no significant differences. Nursery school practice teaching and kindergarten or first grade practice teaching are two (TQ#40); overall, few teachers checked these items. In addition, analyses of types of experience other than in Head Start (TQ#42) could not be completed because very few teachers in the total sample checked these items (34 of 264). There are not enough responses to draw conclusions other than a general one that very few PV teachers have types of teaching experience other than Head Start.

is not the same in all sites. For the second variable, number of years experience with Head Start (TQ#42), there are moderately significant model effects. The means show that teachers in Bank Street have, on the average, taught for 2.6 years and teachers in the Florida models have taught for 2.5 years. Both of these means are lower than those for the other models which range from 3 to 4 years average experience. The site effects are stronger than model effects on this variable. Arizona and High Scope have the largest variation. In the Arizona model, the means range from an average of 1.6 years of experience in Lincoln to 4.6 years in LaFayette, and in High Scope the range is from 2.2 years of experience in Fort Walton Beach to 5.2 years in Seattle. There are no differences between Planned Variation and control teachers on this dimension.

These and the other findings for the staff background variables, then, indicate that there are wide variations among teachers in Planned Variation in their teaching experience and training, their level of education, their age, place of residence, and assignment to PV. Thus, it appears that the teachers with whom a sponsor works in one site are different from those in another site. Moreover, the situations which one sponsor faces are different from those which another sponsor faces. Finally, the relative

lack of PV-NPV effects suggest that on many dimensions, the differences among teachers are not unique to Planned Variation.

Staff Reaction and Input as Determinants of Levels of Implementation:

The finding of variations in staff background characteristics supports the expectation that these variables might be useful in explaining variation in levels of implementation. In considering them in this framework we shall follow the division introduced earlier, between staff input and reaction to the model.

By staff input we mean the characteristics and experiences which an individual brings to the situation in which he or she is to implement a model program. The data just presented fit into this category. The assumption that implementing a model is a difficult process is important here. We have asserted that full implementation is difficult because it requires knowledge and mastery of new skills and techniques in teaching. We also asserted that implementation will be more difficult for some teachers than for others. The consideration of staff input variables provides a basis for making predictions about which teachers fall into which group.

The predicted relationship with implementation is clear for the number of years in Planned Variation variable: the longer a teacher has worked with a model, the more

successful she should be at implementing it.[36] For other variables, the relationship with levels of implementation is less obvious. In general, we might predict that a young, inexperienced teacher would tend to implement a model to a greater extent than would an older, more experienced teacher. This prediction assumes that an older teacher has an established pattern of teaching. For her, implementing a new model involves not only learning new skills, but breaking old habits as well. Particularly if she is comfortable with her established methods, implementing a new model may be very difficult. In contrast, a new teacher might be very receptive to a model because it provides an organized framework within which to cope with her new duties.

For the impact of level of education, additional training, and certification on implementation, we can make opposing predictions. On the one hand, more education and more training may be detrimental to model implementation because a well-educated teacher may be aware of a range of alternative approaches to teaching young children, and may consider the model simply as another alternative. In such a case, she might ignore the model, or pick only parts of it to work with,

---

[36]Unfortunately, this prediction is not borne out when the sponsor ratings are used as measures of implementation. As we pointed out in Chapter 2, however, it is unclear whether the finding is due to a real lack of differences between first and second year teachers or whether it is an artifact of the instrument.

rather than trying to implement the entire model. On the other hand, a more highly educated teacher might be familiar with the concepts a sponsor is trying to introduce through his model, and therefore might understand the model better than a person who has less training. In this case, a higher level of education might be beneficial to implementation. We propose that the determination of which prediction is realized depends on the interaction of education and training with other variables such as reaction to the model and previous methods of teaching. We will return to these other variables shortly.

Race and place of residence of the teacher are teacher characteristics which are of interest in describing Planned Variation teachers, but which are difficult to relate to implementation as input variables. Place of residence is better considered as it relates to a total site, and will be considered as a context variable.

The second category of variables through which the local staff interacts with implementation can be called reaction to the model. We expect that if a teacher likes a model and the people who sponsor it, she will have a higher level of implementation than if she does not like it.

One aspect of a teacher's reaction to the model is her reaction to the sponsor staff and to the training they give. A variable presented in the sponsor input section

under continuing support and feedback is relevant here: satisfaction with training.[37] It is possible that high satisfaction with training indicates a positive reaction which may be associated with greater attempts to implement the model. The more important side of this dimension, however, is probably the negative one. We would predict that if training is very poor that people would have little motivation -- or skills -- to implement the models. Similarly, if the local staff views the sponsor's staff as unprepared or incompetent, we would expect implementation to suffer. We have no data on this latter variable.

There are several other variables for which there are no data, but which are crucial to understanding the relation of staff characteristic to levels of implementation. One of these is reaction to the content of a model, or to the model as an approach for teaching. This aspect of a teacher's reaction is probably more important than her reaction to the sponsor's staff and training. We expect that liking for a model is determined primarily by the teacher's philosophy of education. If the model is consistent with her views about the nature of children and their learning processes,

---

[37] It should be remembered that this variable asks about training in general, and can only be used as a rough estimate of satisfaction with model training.

then the teacher will be more apt to like the model and will make a greater attempt to implement it.[38]

Another very important variable for which there are no data is the method of instruction used by a teacher before she began working with a model. In one sense, this variable fits with the above discussion about the consistency of a model and a teacher's philosophy. We would predict that a teacher whose previous style is consistent with that required by the model will react more favorably to the model than will a teacher whose previous methods are very different from those of the model. In another sense, this variable can be seen as input and is directly related to the assumption that implementation is more difficult for some teachers than for others. We predict that implementation will be easier if the model is consistent with a teacher's previous way of operating a class and relating to children, because it will primarily depend on learning new skills rather than breaking old habits. For example, we would predict that a teacher

---

[38] There are two variables, whether the teacher volunteered or was assigned, first to the center and then to Planned Variation, which might be used as indirect measures of liking for a model. We might predict that if a teacher has chosen to work with the model, she may like it better than if she had been assigned to work with it. Choice indicates some preference for a model, even before beginning. Assignment without choice does not involve any feeling of preference, and in some cases, may even involve forcing a teacher to work with a program she does not like. This somewhat tortured reasoning, however, does not substitute for a direct question of whether a teacher likes a model.

who has been accustomed to working with children in structured academic lessons would find it easier to work with the Oregon model than would a teacher who gives the children a great deal of time for free play and individual exploration.

The extent to which the model is understood, a final variable for which there are no data, is an important dimension because a teacher cannot implement a model when she does not understand what is expected of her, or how she can carry out the requirements of the model.

In summary, it appears that staff variables may be important in determining the difficulty of and motivation for implementing a model. We have suggested that levels of implementation may be partially determined by factors such as the teacher's level of education, experience in Head Start, previous methods of teaching, as well as reactions to model training and the content of a model. We would expect, then, variation in staff factors might explain variation in these levels. Although the variables presented in this section tap several dimensions of staff input and reaction, there are no data for other crucial dimensions. For a complete understanding of the relation between staff characteristics and model implementation, we need information on the teacher's

liking for and understanding of the model, and her previous methods of teaching.[39]

Implications for Study Design:

The fact that we expect teacher background variables to influence implementation has implications for the design of the Planned Variation study. It is important to comment on these before continuing our discussion of factors which influence implementation. Ideally, in an experiment, one assumes that the situations in which the treatments are to be tested do not differ sytematically except in ways caused by the treatments. This assumption is important in drawing conclusions about outcomes of the treatments.[40] The finding of significant differences among sites and models on teacher background characteristics suggests, however, that this assumption is not met: teachers differ systematically in ways which are not caused by the treatment.

---

[39]This section focused almost exclusively on the relationship of teacher characteristics to implementation. Teacher aides were not included primarily because of the limited data available on them and because of the lack of model or site differences in the data which do exist. We expect, however, that in classes where the aide is included as a teaching partner that many of the same relationships will hold. The whole question of the teaching aide's role in implementation deserves further study.

[40]This is another aspect of an assumption in Chapter 1: that treatment is a primary determinant of what happens in the classroom.

The differences among sites on these dimensions are not hard to understand: all teachers within a site are directed by the same administrators, and administrative policies in Head Start vary from one center to the next. Sites may place different emphasis, for example, on having community members on the teaching staff, on requirements for certification, or on additional training. Different hiring policies or career ladders, then, may result in systematic differences among teachers in different sites.

The systematic differences among models on these variables are more difficult to deal with. Even with large differences among sites, we did not expect systematic model differences. A search for an explanation for these differences in such factors as regional differences among models reveals no pattern which relates to teacher characteristics. The finding that not only does each sponsor face a variety of situations in trying to implement his model, but that the situations differ from model to model, confounds the treatments in Planned Variation. It will be difficult to tell whether differences in child outcomes result from model differences or teacher differences. This is a serious problem with the study design.

OPERATIONAL CONTEXT OF THE STAFF

The third category of factors relevant to implementation is the context in which the staff operates: the sponsor and staff inputs do not operate in a vacuum, but rather interact with the situation around them. For example, it is probable that implementation will be difficult when the site is in turmoil over the loss of funding, even if sponsor support to a site is strong, and the teachers like the model. Three aspects of the operational context can be identified. In suggesting how context variables may be related to levels of implementation, each aspect will be considered separately. In the presentation of available data, however, the three dimensions will be combined because the data are inadequate for separate discussions.

The first dimension of the operational context, which we shall call site characteristics, includes variables which help to determine what a site is like. For the most part, it includes straightforward and easily obtained information such as the size of site or adequacy of facilities. In part, however, it also includes less tangible factors such as the atmosphere in the centers and the values on which they operate. The second area, site administration, includes those factors which determine how smoothly, or efficiently, a site operates. Our assumption is that if the site is not functioning well, the energies of the staff will be diverted

to administrative problems and away from model implementation. Thus, efficient site management may be a minimal condition for implementation. The third category, administrative support for the model, is based on the assumption that the Head Start director is a key person in the implementation process. We predict that the director's opinion about the model and about Planned Variation will influence implementation by affecting the efforts of the teaching staff to work with the model, and by determining the administrative arrangements made in relation to the experiment.

Data:

Three items from the Teacher Questionnaire are relevant to a discussion of the context of implementation. Table 23 shows the analyses of variance for these items for both PV classes only and for PV-NPV comparisons. The analyses for PV only are based on the standard balanced design. Those for the PV-NPV comparisons are based on the smaller sample of 6 models and 12 sites. Table 24 gives the means on the three variables for the models included in the analyses.

Table 23 shows that there are moderate differences among models on the teachers' satisfaction with their working conditions (TQ#24), on the extent of parent involvement in the classroom (TQ#29), and in the frequency of

TABLE 23

Context Analyses of Variance

Planned Variation Classes Only:

	Variable Name	Model	Site Within Model	Class Within Site	df Class
TQ#24	Satisfaction with working conditions	9.36 † *	19.12 *	71.52	134
TQ#29	Degree of parent involvement	12.18 † *	15.57 NS	72.31	126
TQ#30	Activities for parents	9.38 *	22.00 **	68.61	129

Note: See Table 8 for a description of these analyses.

PV-NPV Comparisons:

	Variable Name	PV-NPV	Model	Site	Model X PV-NPV	Site X PV-NPV	Class	df Class
TQ#24	Satisfaction with working conditions	0.2[illegible] NS	4.96 NS	10.77 NS	5.52 NS	9.31 NS	69.16	
TQ#29	Degree of parent involvement	0.20 NS	18.11 ***	5.79 NS	4.81 NS	13.99 **	57.10	
TQ#30	Activities for parents	0.00 NS	16.03 **	9.71 NS	3.99 NS	7.16 NS	63.11	

Note: See Table 9 for a description of these analyses.

TABLE 24

Context

Means for Models in Standard Design

		Far West	Arizona	Bank St.	Oregon	Kansas	High Scope	Florida	EDC	Enablers
TQ# 24	Satisfaction with working conditions* (1 = very satisfied, 5 = very dissatisfied)	2.0	2.2	2.5	2.5	2.4	2.3	2.3	2.1	2.3
TQ# 29	Parent involvement* (low number = low involvement)	2.9	2.4	2.8	3.4	3.0	2.5	2.5	3.0	3.2
TQ# 30	Parent activities* (1 = weekly, 4 = once a year)	2.6	2.6	2.9	2.6	2.1	2.6	3.0	2.6	2.7

Note: * indicates significant differences between models; extreme means on those variables are underlined: single underlining indicates high means and double underlining indicates low means.

TQ: Teacher Questionnaire

TABLE 25

Context

Means and Standard Deviations from Sponsor Site Assessment

	Far West	Bank St.	Oregon	Kansas	Florida
a. Turnover rate teachers	1.5 .50	1.8 1.30	3.0 1.0	2.3 1.89	3.8 1.64
b. Turnover rate aides	2.2 1.30	1.0 0.0	3.5 1.50	2.3 .94	1.8 1.30
c. Turnover rate children	3.5 .87	1.8 .83	2.0 0.0	1.7 .94	3.0 0.0
d. Intra-staff friction	2.5 1.12	4.0 .71	2.0 0.0	2.3 .94	2.8 1.30
e. Regular attendance teachers	4.2 .43	4.2 1.30	2.0 0.0	2.3 .94	0.0 0.0
f. Punctuality of teachers	4.2 .43	4.2 1.30	2.5 1.50	0.0 .00	0.0 .00
g. Regular attendance children	3.0 .71	4.0 1.22	3.5 .50	3.7 .94	3.0 .00
h. Support of local Head Start	4.2 .83	3.5 .87	3.0 1.00	3.7 .94	2.7 1.09
i. Support of the community	4.0 1.22	3.0 1.22	4.5 .50	3.7 .94	4.0 .71
j. Support of PAC	4.5 .50	3.0 1.58	4.5 .50	3.7 .94	3.8 .43
k. Adequacy of plant indoors	4.0 .71	3.2 .43	2.5 1.50	3.0 1.63	1.7 .47
k. Adequacy of plant outdoors	2.8 1.22	3.0 1.78	2.5 .71	2.3 1.50	1.7 .94
n. Rapport between administrator and staff	3.5 .50	3.5 .50	4.0 .00	3.0 1.63	2.2 .83
o. Rapport between staff and children	4.2 .43	4.0 .71	4.5 .50	3.0 .00	3.5 .87
p. Rapport between sponsor and local Head Start staff	4.2 .43	4.0 .71	4.0 1.00	4.3 .94	3.5 1.12

Note: Only means for models in the analyses of variance are included here. There were no significant model effects on any variances. The standard deviations give an indication of the variation among sites. The Enabler model is not included here because the Enabler consultants were not asked to complete the Sponsor Implementation Report.

TABLE 26

Context

Means and Standard Deviations from Consultant Site Assessment

CR#		Far West	Arizona	Bank St.	Oregon	Kansas	High Scope	Florida	EDC	Enablers
1. Turnover teachers	$\bar{x}$	1.2	2.0	1.5	3.0	2.0	1.0	2.8	2.0	1.5
	SD	.45	1.73	.58	1.73	1.00	0.00	1.50	0.0	1.00
2. Turnover aides	$\bar{x}$	1.8	1.7	1.0	3.0	2.3	1.5	2.8	2.0	1.50
	SD	.45	1.16	0.0	1.73	1.56	.58	.96	0.0	.58
3. Turnover children	$\bar{x}$	2.0	1.3	1.8	1.7	1.3	2.8	2.7	1.5	2.2
	SD	.71	.57	.96	1.16	.58	.50	.57	.71	.96
4. Intra-staff friction	$\bar{x}$	2.6	1.7	2.2	1.7	2.0	3.0	3.8	2.5	2.5
	SD	1.14	.58	.96	1.16	1.0	.82	.50	.71	1.0
5. Reg. attend. of teachers	$\bar{x}$	4.0	4.3	3.8	2.3	4.7	3.0	3.7	4.0	4.0
	SD	1.00	.58	1.26	.58	.58	1.41	.58	0.00	1.16
6. Punctuality of teachers	$\bar{x}$	4.0	4.7	3.5	4.3	4.0	3.5	4.0	4.0	4.2
	SD	1.16	.58	1.29	1.16	1.73	1.73	1.00	0.00	.96
7. Reg. attend. of children	$\bar{x}$	3.8	4.3	4.0	3.7	3.7	3.2	3.0	3.5	3.0
	SD	1.30	.58	.82	1.16	1.16	.96	0.00	.71	.82
8. Support local HS for model	$\bar{x}$	3.8	4.3	3.0	3.0	4.0	3.2	3.2	3.0	4.0
	SD	1.30	.58	1.63	1.73	1.00	1.50	1.26	0.00	1.41
9. Support PAC for model	$\bar{x}$	3.4	4.0	3.2	2.0	3.7	3.2	3.5	3.0	4.3
	SD	.89	1.00	.96	1.00	.58	1.50	1.29	0.00	.58
10. Support community for model	$\bar{x}$	3.6	4.3	3.5	3.0	4.0	2.8	2.8	3.0	4.3
	SD	1.14	.58	.58	1.16	1.00	1.26	.50	0.0	.58
11. Adeq. plant indoors	$\bar{x}$	4.0	3.7	3.0	2.7	3.0	2.2	2.5	3.0	3.5
	SD	.82	1.16	.82	1.52	2.00	.96	.58	1.41	.58

12. Adeq. plant outdoors	x	3.5	3.3	3.2	1.7	1.3	3.0	2.2	1.0	2.0
	SD	1.92	1.53	1.71	.58	.58	1.16	1.26	0.00	.82
15. Rapport between admin. & staff	x	3.2	4.0	4.0	4.0	3.7	3.0	2.8	4.0	2.8
	SD	1.50	1.00	.82	0.00	1.16	1.41	.96	0.00	1.26
Rapport between staff&child	X	4.4	4.7	3.8	4.0	3.3	3.5	3.8	3.0	4.2
	SD	.55	.58	1.26	1.41	1.53	1.73	.96	0.00	.96
17. Rapport between sponsor staff & local staff	x	3.6	4.0	3.5	3.3	4.3	3.0	2.2	3.5	4.7
	SD	.89	1.00	.58	1.16	1.16	1.41	.50	.71	.58

Notes: Only means for models in the analyses of variance are included here. The standard deviations give an indication of the degree of site to site variation within models.

activities offered for parents (TQ#40).[41] Examination of the means indicates that the teachers in Far West (2.0) and EDC (2.1) tend to be the most satisfied with their working conditions while the teachers in Florida (2.7) are somewhat less satisfied. Oregon has the highest level of parent involvement (3.4) and Kansas tends to have the most frequent parent activities, averaging around once a month (2.1), although all models are fairly similar on this variable. There are also significant differences among sites within models on satisfaction with working conditions and activities for parents. Bushell has the largest variation on both variables: the teachers in Mounds are satisfied with their working conditions (1.8) while the teachers in Portageville have mixed feelings (3.2). Activities for parents are offered weekly in Mounds (1.0), but only twice a year in Oraibi (3.1). Finally, there is only one PV-NPV effect for these variables: an interaction with sites on the amount of parent involvement. This indicates that the relationship between PV and NPV varies across sites.

---

[41] In analyzing each of these variables, the form was altered slightly from the original. For working conditions and activities for parents, an average was taken across the specific categories of conditions and activities. For the parent involvement item, the transformation was slightly different: the frequency of parent involvement (every day= 5, once a week=4...) was weighted by the number of parents in that category (e.g., 6 parents worked once a week = 6 X 5), summed across the categories and divided by the total number of parents involved. The resulting scores are a combination of frequency of participation and number of parents, where a high score indicates a greater involvement. The original format of the questions are found in Appendix B.

The primary sources of data for this area, however, are the site assessments from the Final Consultant Reports and the Sponsor Implementation Reports rather than the Teacher Questionnaire. The means and standard deviations for the variables from these site assessments are given in Tables 25 and 26. The analyses of variance for these reports differ from those for the Teacher Questionnaire because for each variable, there is only one observation per site. As a result, we can only test for differences among models and not for differences among sites. Because of missing data and single site models, only 5 models were included in the analyses. The analyses of the site assessments reveal no significant differences among models on the items relevant to the context of implementation. It is not clear, however, whether these findings result from a real lack of differences, or from the low power of the test. Although site differences cannot be tested statistically, the large standard deviations on some of the variables suggest that sizable differences do exist.[42] Thus, we might tentatively conclude that there are differences among sites within models on some of the items in the site assessments.

---

[42] In these reports, where the scores can range from 1 to 5, a standard deviation of 1.00, for example, results when the site scores are 1, 2, and 3. A standard deviation of 1.73 indicates a wider range of scores such as 5, 4, and 1, while a value of .58 indicates a narrower range such as 1, 1, and 2.

It should also be noted that the degree of correspondence between the sponsor and consultant ratings is not always high. Table 4 in Appendix C shows the correlations between the two sources for each item. The correlations range from a .10 for the turnover rate of children to .68 for adequacy of the physical plant inside. Since the items were essentially identical for the two groups of people making the assessments, the low correlations indicate that the ratings are unreliable: for one site, the ratings vary according to who is making them. It may be that the items are so general that they can be interpreted in several ways. In any case, it is clear that the findings from the site assessments should only be used as general indicators of the context in which implementation is taking place.

Within these limitations, we feel that the assessments provide interesting descriptions of the sites. Since it is difficult, however, to integrate 20 variables into a coherent picture, we performed a factor analysis on the consultant site assessment in an attempt to reduce the variables to a manageable number of dimensions. The consultants' rating rather than the sponsors' were used because there are more data for them. The results of the factor analysis are included as Appendix E at the end of the report.

A few general trends are apparent from the assessments. From the consultant reports, it appears that all models tend toward the lower end of the scale on friction, while they tend toward the high end on punctuality, regular attendance, support of Head Start personnel and PAC for the model, and rapport between the staff and the children. From the sponsor reports, all models tend toward the high end of the scale on regular attendance of the children, support of PAC and the community for the staff and the children, but all tend toward the low end on the adequacy of the physical plant outdoors.

We can conclude, then, from the data which exists for the context in which implementation takes place, that there are differences among sites on a number of dimensions. The signficance of those differences, however, cannot be tested for the site assessment data. The differences among models are significant only for three items from the Teacher Questionnaire and there are low correlations between the two sources of data for the site assessments.

Context Variables as Determinants of Levels of Implementation: Site characteristics: The first category of context variables which may be related to implementation revolves around the broad question of what the sites are like. We expect different dimensions of this category to interact with implementation in different ways. One characteristic of

the site, and the community of which it is a part, is directly related to model implementation: support for the model. There are three items on the site assessment which report support from the local Head Start staff, the PAC, and the community for the model.

Other site characteristics are less explicitly related to implementation, but are no less important. The satisfaction with working conditions and adequacy of physical plant variables are relevant here. We expect that the conditions under which teachers must work are important primarily in terms of establishing a minimum base for implementation. Above the minimum level, we do not expect this factor to have an appreciable influence on implementation. Below the minimum, however, poor working conditions will make implementation very difficult. Facilities would seem to be a particularly important element of this factor. If , for example, several classes are housed in one room, the lighting is poor, the noise level is high, or the supplies are inadequate, we would expect model implementation to be hindered.

Two additional variables for which there are data, parent involvement and activities for parents, may also be related to implementation. We expect that a center which attempts to involve parents in classroom and other Head Start activities will provide a different context for implementation than one in which parents are not deeply

involved. These variables may be part of a larger dimension which contrasts a community-oriented site with a school-oriented site. A community-oriented site is one which encourages full parent participation as opposed to a school-oriented site which maintains more formal, distant relationships with the parents. If this dimension is valid, it might also be reflected in other factors; such factors might be hiring priorities, as measured by the number of teachers who live in the same neighborhood with the children, or the number of teachers who are certified (Both of these variables were introduced as staff background factors). We would expect that a community-oriented site would emphasize hiring neighborhood people, while a school-oriented site would place a higher value on credentialled teachers. It seems unlikely that this dimension would have the same relation to model implementation in all situations. It may be that it interacts with models: some models may be better for community-oriented sites while others are better for sites with a school orientation. More study is needed on this aspect of the site characteristics; the present variables are not adequate for exploring it.

Another variable which in some ways is an extension of this school-community dimension, is the delegate agency. This is a question of who runs Head Start. It is a continuation of the previous discussion in the sense that

in many cases Head Start is run either by CAP, the Community Action Program, or by the public schools. If we are concerned with defining the context a site provides through its values or orientations, the school-community contrast applies fairly well,[43] and might provide a tool for increasing our understanding of the implementation process. In another sense, however, the variable is not an extension of the previous discussion. The delegate agency variable may be related to implementation in more concrete ways than values. It is possible that public schools provide contexts which facilitate implementation because they may have more materials, better facilities and more efficient operating procedures than other agencies which administer Head Start. This is another site characteristic which deserves exploration.

A third aspect of the delegate agency variable which may be related to implementation is the presence of day care. We would predict that model implementation will be more difficult in a Head Start center which is combined with a day care program than in one which is operated independently. We base this prediction on the assumption that teachers and children both tend to be tired from the longer day involved

---

[43]One problem with matching a delegate agency label with an orientation label is, for example, that there can be school-oriented centers which are not administered by the public schools. Implementation will be more difficult than in centers where there is less strain.

with day care. As a result implementation will be more difficult than in centers where there is less strain.

These variables do not exhaust the aspects of site characteristics which may be related to implementation. There are no doubt still more areas for which we have no information, such as the size of the site and location of the centers. In relation to the size variable, we would predict that if a site is very large and Planned Variation is only a small part of the total operation, then Planned Variation may receive less attention than in a site where it is a major concern. As a result of being less important, model implementation may suffer. Similarly, the location of classes within the site may influence implementation. We would predict that if Planned Variation classes are spread out among centers rather than being used together, the isolation of teachers trying to work with a new model may hinder implementation.

Thus, there are a number of site characteristics which may be related to implementation. Undoubtedly, we have not identified all relevant characteristics, but we have raised some important issues. While there is no evidence of variation in some of the variables considered here, there are differences among sites on others: hiring priorities, parent involvement, and delegate agencies. The existence of differences among sites on these variables, then, supports

the hypothesis that site characteristics may partially account for differences in model implementation.

Site administration: Site management, a second dimension of context which we would expect to be related to implementation, is primarily a question of how smoothly, or efficiently, the site is managed. If a center is functioning well, then the teaching staff within it can concentrate on the classrooms and on implementing the model. These variables are seen as minimal conditions for implementation: they alone do not guarantee implementation, but they do support it.

A number of variables from the site assessment can be considered as indicators of good management. One component is stability of the center, particularly of the staff. This includes elements such as turnover rates, punctuality, and regular attendance. These variables are found in the site assessments. We predict that implementation will be at a lower level if staff members are frequently late or absent, or if they work at the centers for only short periods of time, than if the staff are on the job regularly. Our assumption is that training new teachers, finding substitutes, and covering extra classes diverts energy away from model implementation.

Similarly, we predict that friction among the staff, a second component of site management, diverts resources from implementation. The friction variable from the site

assessment provides data which is relevant here. If there is disagreement about such concerns as role definitions, we would expect conflict and neglected duties. Such a situation may hinder implementation. On the other hand, if all staff members know their areas of responsibility and those of the people around them, this concept will not be an issue in implementation.

In addition to these aspects, there are other components of site administration for which we have no data but which may influence model implementation and therefore should be considered. One is administrative competence. Again, the negative end of the dimension is the important one in this context. We would expect that unless the administrative staff is at least minimally competent, -- that is, unless they can manage such tasks as getting the children to the school and fed, paying bills and salaries, and getting the buildings open and heated -- model implementation cannot proceed.

In contrast to the previous variables which refer to internal issues, funding security involves the concept of efficient relations with people and agencies outside the Head Start center. Clearly, funding problems hinder implementation when they prevent a site from hiring enough people to work with the model. Even when staff members do not actually lose their jobs, however, the threat of

budget cuts, late pay, and general funding instability may interfere with implementation: if staff members are beset with worrying about whether or not they will have a job next month, working to avoid this crisis will probably divert attention from the classroom. There are no data for this item either.

The same relation to implementation, then, holds for all the variables which have been introduced as site administration factors: problems in administering the site divert attention and energy from model implementation. While some of these variables may be more important to implementation than others, we would expect, in general, that they will be additive: the more problems a center has in its operations and relations with other groups, the less attention will be given to the model; as a result, implementation will suffer.

Administrative support for the model: We expect that all Head Start administrators will have opinions about the models with which their teaching staff must work, and about the importance of the Planned Variation experiment.[44] Further,

---

[44] A distinction between opinions about the model and opinions about the Planned Variation experiment is necessary because the former alone will not determine implementation. The Planned Variation design involves the assumption that the models will be replicated in the sites. Even if a director likes a model, she may interfere with implementation if she does not understand and respect the experimental design of which it is a part.

we expect these opinions to affect model implementation in at least two ways: by affecting the attitudes of the teachers toward the model, and through administrative decisions which facilitate or hinder implementation. Although there are no systematic data available through which to test these hypotheses, anecdotal evidence suggests the issues are at least worth raising.

Before discussing in more detail the ways in which administrative support may influence implementation, it will be helpful to comment briefly on the sources or determinants of such support, although without data we cannot test our hypotheses. This question, like many others, is complex and deserves more thorough study than we are able to give to it. We expect, however, three kinds of factors to determine an administrator's support for a model. The first is the adminstrator's opinion of the sponsor. Since curricular models are tied closely to the notion of sponsorship in this experiment, it seems probable that the relationship between the sponsor and the Head Start director will greatly influence the latter's opinion of the model and the experiment. In evaluating the relationship, we would expect that things such as the way the sponsor presents himself and his program to the community initially, his continuing attitude toward the site and toward the director in particular, and his competence in training and monitoring implementation will

be important. If the director has a high opinion of the sponsor, we would predict that she would support the model to a greater extent than if she did not like him. A second probable determinant of an administrator's opinion of the model is her philosophy of education. We would expect that if her values are congruent with those of the model, she will like the model. A third determinant of support can be labelled as priorities in operating the program. This factor is perhaps less obvious than the first two, and is better viewed in terms of the administrator's view of the importance of Planned Variation than in terms of liking the model. By priorities we mean the director's sense of what is important for the program; her perception of the roles and goals of Head Start. We expect, then, that the director's support for the experiment will be influenced by her priorities: if maintaining an experimental design has low priority, she may intentionally or unintentionally interfere with implementation by acting on higher priority items. Thus, the support an administrator gives the model may be determined by the interaction of her feelings about the sponsor and his representatives, the congruence of the model's approach to education with her own beliefs, and the relative standing of the PV model in her system of priorities for the Head Start program.

With some feeling about the determinants of why an administrator would or would not support the model, we return to a discussion of how such support then influences implementation. Unfortunately, we have no systematic data for these questions either, and must rely solely on anecdotes for illustrating the discussion. To repeat our previous hypothesis, we expect that support is related to implementation in at least two ways: by affecting the attitudes of the teachers toward the model, and by making administrative decisions which facilitate or hinder implementation. In the first case, the administrator's opinion is seen as an important determinant of the atmosphere created within the site in relation to the model. We would expect that in a situation where the director likes the model and thinks Planned Variation important, she will give moral support and encouragement to the staff working with the model. On the other hand, if she does not think highly of the model in particular, or the experiment in general, the atmosphere for the staff will be poor. For example, one sponsor reports[45] that "Morale problems were generated because paraprofessionals in Planned Variation were treated as temporary employees by administration." It seems probable that staff members in such a situation will be less motivated to work with the model than in one where the director supports the model.

---

[45] The Sponsor Implementation Reports contain several open-ended questions; this quote is taken from one of them.

The second way in which administrative support influences implementation is through decisions and policies. This aspect of support is more complicated than the first because it includes a larger proportion, and therefore a wider variety, of the administrator's actions and responsibilities. We cannot specify all the actions which might affect implementation, but we can give examples to illustrate the importance of this concept. In some cases, a director will not take the action necessary to support implementation because he or she does not like the model. For example, one consultant in discussing a number of problems a director has with a model and its sponsors states: "Since he's not sure of the model, he doesn't want to spend money on materials, books, etc. that are related to it. With the big financial cut, he's worried about resources."

In other cases, however, the important support factor for implementation is not liking the model, but respecting the experimental design of Planned Variation. A director, for example, may have a high opinion of the model as an apporach to teaching children, but unless she understands the experimental design and tries to maintain it, she may hinder implementation by the policies she pursues. The relation of such actions to implementation may not be immediately apparent until we remind ourselves of the goals of Planned Variation. One of the primary objectives of this

experiment is to test the effectiveness of implemented model curricula. It is assumed that the sites will try to replicate the models as closely as possible. The study was not organized to encourage sites to pick out some parts of the models and discard the rest, nor was it intended that they work with the models only intermittently. It is also assumed in testing the effectiveness of the specified models that other factors which may affect the outcomes are controlled. If there is a large amount of unsystematic input to the classes, the purpose of the study will be defeated. We will not be able to draw conclusions about the causes of differential effectiveness. A related, but slightly different issue is that of comparison classes. In Planned Variation, many of the comparison classes are drawn from the same sites as the experimental classes. If these comparisons are to be useful, it is important that the model not be implemented in them. Therefore, it is important that the two groups of classes not be given the same treatment. When we speak of administrative policies which do not respect design, we are referring to such actions as making demands on staff time which take concentration away from implementation, not providing necessary resources to the classes, making non-model input into the classroom programs, or contaminating comparison groups.

Examples of directors carrying out policies which interfere with implementation, unfortunately, are numerous. Among them are statements from both OCD consultants and sponsors. One sponsor states, for example, that

> The Resource Teacher . . . is interested in implementing the [model]. However her additional responsibilities assigned by the Director, tend to limit her time in the [model] classes and mitigates against her effectiveness. She is on an overload servicing 13 groups.

A similar example of demands on staff members which interfere with implementation comes from a consultant to another site who also reports that the Program Advisor

> is excellent, but her load is much greater than it should be, since it includes PV, the commitment to the 1969 [model] trainees [who are not the same as the 70-71 trainees] and the total program. As I understand it, the total program includes about 50 classes.

The important fact to realize is that the people referred to in both examples have primary responsibility at their sites for seeing that the models are implemented. The administrator's decision to make Planned Variation only part of the job responsibility may seriously hamper implementation.

Another type of administrative policy which affects implementation involves demands which take the staff's attention away from the model. One consultant, after reporting on a number of demands that were being made of

the teachers, concludes that all the non-model problems

> temporarily detract teachers from spending mental and physical time doing a better job in the classroom.

Or in another center, a consultant comments about what took place in a long conference she had with the Program Advisor about the model and the practicing staff. One of her statements is that

> We also expressed concern about the pressure of career development that seems to minimize the importance of the classroom. . .

Like the examples above, these support the belief that administrative policy can interfere with model implementation. Another aspect of the requirements of the experiment, and thus, of implementation, which may be hindered by administrative actions is input into the classroom by people other than the sponsor's staff. One consultant reports that

> The State RTO has been asked to come in to do some inservice training. This is a non-model person . . .She hasn't been asked to adapt anything she will say to the model . . .

A final problem is that it is possible that comparison classes may be contaminated because they are included in model training. One consultant states, for example, that the director cannot understand why all children cannot be included under Planned Variation.

Thus, it seems clear that although we do not have systematic data for all sites, we can conclude that it is very probable for administrative policy to affect model implementation. Actions which hinder implementation can be interpreted as lack of support. This is not to say that the directors consciously interfere with the model. We propose that the notion of priorities determining support, which was presented earlier, is the crucial one here. Unless Planned Variation has high priority, the director will base her decisions on other competing interests, and the resulting policies may interfere with implementation. In part, this may be explained as a lack of understanding of the experiment; a director may not recognize that his actions interfere because she does not understand the requirements of an experiment. The last example given above illustrates this point well: the director did not seem to have any notion of why it was important to include only certain groups of children in Planned Variation. This lack of understanding -- and its effects -- is apparent in other sites too. One sponsor, for example, states that the

> Administrator lacked clarification of her role, and the meaning of . . . sponsorship, i.e., last fall a second research design was introduced into [model] classrooms.
> No understanding of sponsorship -- difficult to keep clear which classes are being studied, who is being trained by . . . [the sponsor] staff.

Lack of understanding, however, is not the only explanation for giving other concerns priority over maintaining the research design. Another explanation can be viewed within the framework of the tension between a research program and a service program. We propose, first, that the demands of a service program are often in conflict with those of a research design. While an experiment calls for controlled inputs and differential treatments, for example, a service organization calls for bringing in all the resources possible and providing service to as many people as possible. Moreover, as the examples above suggest, some Head Start Centers have goals other than education for the children to be met. Second, we propose that Head Start is considered as a service program, not as an experiment by those who run it. This is particularly true since the sites chosen for Planned Variation were operating centers before and during the study, and will continue after the study. Thus, we conclude that the directors in Planned Variation sites are primarily service oriented, and as such, may understandably place the Planned Variation research design low among their priorities. Such a placement may be interpreted as lack of support for the model because it may result in administrative actions which interfere with implementation.

In this discussion of administrative support for the model we have asserted that administrators have opinions about the models and Planned Variation which may influence

implementation. The influence takes at least two forms: the atmosphere a director creates for the staff working with the model and administrative action. We propose that the second may be more complicated because it revolves not simply on liking the model but on the interests competing with Planned Variation in the administrator's priorities. If the research design has low priority, action may be taken on the basis of other interests and, as a result, may interfere with implementation. We cannot draw any conclusions about the validity of these hypotheses because we have no data to test them. It does appear, however, that these variables have a great deal of potential importance and therefore, merit further study.

Summary of Operational Context Issues:

In this section we gave only limited attention to relevant data. For the two primary sources of data, the site assessments completed by the sponsors and the OCD consultants, there appear to be differences between the sites on a great many variables. The differences, however, cannot be tested statistically. Moreover, there are no significant differences among models, and the correlations between the ratings from the two sources are low. The data, therefore, can only be used to suggest interesting differences among sites; strong inferences cannot be made from specific variables.

In contrast to the data presentation, we discussed at length the areas of the operational context which deserve further study. Administrative support for the model, particularly as it is reflected in a director's policies and priorities, seems especially important. We suggested that knowledge of these additional areas, together with the context variables for which there are already data, may help to explain variation in levels of implementation.

SUMMARY AND CONCLUSIONS

The organization of this chapter is based on the assumption that implementation is an interactive process depending not only on sponsor input, but also on staff input and reaction, and the context in which implementation takes place. Within each of the three major sections, we first presented relevant data. In examining the results of the analyses, it must be remembered that the findings are suggestive rather than conclusive. For all data, there are limitations of using sites as fixed factors, multiple significance tests on non-independent measures, and low reliability (as evidenced by low correlations between variables tapping the same dimension and by lack of agreement among sources in reports of the same variables). In addition, for the training variables there is also the problem of lack of specificity; it is not

clear whether the teachers are reporting on only model training or on all training. These limitations lead to difficulties in interpreting findings and may result in an exaggeration of actual differences.

From the available data, we find variations among sites within models in all areas and among models in all areas except those measured by the site assessments.

1. From the variations in sponsor input data, we can conclude, first, that sponsors apparently differ from one another in the inputs they make, and that at least some sponsors do not treat all of their sites in the same way. Second, it appears that people other than the sponsors and their representatives are giving training in the sites. Third, there are fairly strong PV-NPV differences in pre-service training, but not in in-service training or in support and feedback. This finding probably reflects a combination of treatment contamination and vague instruments.

2. From the variations in staff background characteristics, we conclude that each sponsor must work with a variety of teachers and that the teachers with whom one sponsor works are different from those with whom another sponsor works. The relative lack of PV-NPV effects suggests that, on most dimensions, the differences among Planned Variation teachers are the same as those among non-Planned Variation teachers.

3. The finding of significant differences among models on teacher characteristics has implications for the design of the study. One requirement of a good experimental design is that the situations in which the treatments are to be tested do not differ systematically. If this assumption is not met, as it is not in this study, then treatment effects are confounded with the other factors which differ systematically among models.

4. The findings for the operational context variables are limited. From the primary source of data, the sponsor and consultant site assessments, there are no model effects, and site effects cannot be tested. We can only conclude, then, that there appears to be some variation among sites within models on most of the context variables. A factor analysis which ties together some of the site assessment variables is presented in Appendix E.

All these findings support the assertion that implementation is a complex process. The variation among sites and models also supports the proposition that these variables are useful in explaining variation in levels of implementation. In the second part of each of the three sections, then, we tried to specify how the variables might be related to implementation and to suggest additional areas which should be considered in a more thorough study of implementation.

1. For sponsor input, we propose that longer and more frequent training from the sponsor may be related to higher levels of implementation, but that other dimensions are equally, or perhaps more, important. The types of training given and the extent of personal support and feedback from the sponsor may also influence implementation. Moreover, the present data suggest that we need to further explore training by people other than the sponsor: such additional training may either facilitate or interfere with model implementation, depending on who gives the training and whether it's consistent with the model. Finally, anecdotal evidence suggests that sponsors may become involved in activities in the sites which range beyond model prescriptions. Since it is possible that these activities may influence implementation, perhaps through the degree of commitment felt by the teachers, the area should be explored.

2. Under staff input and reaction, we propose that not all teachers have the same inclination or ability to implement a model. If staff characteristics determine the difficulty of and motivation for implementation, they might help to explain the variation in levels of implementation. We suggest that reaction to training, age, experience, level of education, and certification may be determinants of levels of

implementation. There are no data, however, for the other variables which may be crucial to the relation between staff characteristics and implementation: liking for the model as an approach to teaching, the extent to which model requirements are understood, and previous methods of instruction. If a teacher likes and understands the model, and if it is consistent with her approach to teaching, we predict that she will have a higher level of implementation than she would otherwise.

3. We suggest three areas of variables within the operational context of the staff category. Under the heading of site administration, we propose that unless a site is operating efficiently, the energies of the staff will be diverted to administrative problems, away from implementation. Site administration is reflected in such variables as staff stability, staff friction, administrative competence, and funding security.

4. Another important, but neglected area, is administrative support for the model. On the basis of anecdotal evidence, we propose that the administrator's opinion of the model and of the Planned Variation experiment will influence implementation through the atmosphere she creates and through her actions and policies. We also assume that the latter means of influence will be more complicated because it is dependent on

the director's priorities for her program as well as on liking the model. Further, we assert that we would expect to find that Planned Variation is a low priority concern to many Head Start directors because their primary conception of their program is as a service project. As a result, we would also expect that some of the director's actions would interfere with model implementation. Finally, we propose that site characteristics, such as size, adequacy of facilities, the delegate agency, and the values and orientation of the site may be important to implementation but the nature of the relationships are not always as clear as in other areas.

The final conclusion to this chapter is that we have raised many issues but resolved few of them. The finding of model and site differences on the majority of variables supports the argument that implementation is a complex process and that some of these variables may be useful in explaining the variation is levels of implementation. There are, however, many dimensions of the implementation process for which there are no data.

Chapter 4

PREDICTING LEVELS OF IMPLEMENTATION

In Chapter 2, we presented the sponsors' ratings of teachers as a measure of level of implementation. We concluded that there is a great deal of variation among classes and sites within models in the extent to which models are implemented. In Chapter 3, we explored a number of variables which might influence levels of implementation. From the data available for these variables, we found large variations both among and within models, and predicted that the variables might help explain the variations in levels of implementation.

In this chapter, then, we will use multiple regression analyses to relate the variables discussed in Chapter 3 to the sponsor ratings presented in Chapter 2.

Multiple regression analysis is a technique for estimating a functional relationship between one dependent variable and several independent variables. The resulting equation can be used to predict values of the dependent variable for given values of the independent variables. The equation for a linear regression can be expressed as:

$$\hat{Y} = b_o + b_1x_1 + b_2 x_2 + \cdots b_nx_n$$

where $\hat{Y}$ stands for the score on the dependent variable (sponsor ratings) which will be predicted from the

knowledge of a set of independent variables; x_1, x_2, ... x_n stand for the given scores on the independent variables; b_1, b_2, ... b_n are the coefficients of the independent variables which give the best linear prediction of the dependent variable; and b_o is the regression constant.[1] In our case, the sponsor ratings of teachers will be used as the dependent variable,[2] and the variables introduced in Chapter 3 will be used as the independent variables. From the regression analysis we can also obtain the square of the multiple correlation coefficient (R^2) which is the amount of variation in the dependent variable explained by the independent variables. R^2 is an important indication of success in predicting Y; one primary goal of the regression analysis is to maximize R^2.

In attempting to explain the total variation among levels of implementation, we can separate the variation among models, among sites within models and among teachers within sites. Considering first the differences among

---

[1]The size of a b coefficient depends on the scale in which the variable is measured. As a result, comparisons of the size of the coefficients cannot be made. To overcome this problem, we can use standardized coefficients, b*. For a more thorough discussion of regression analyses, see Draper and Smith, Applied Regression Analysis (New York: Wiley, 1966).

[2]In Chapter 2, the ratings for both February and May were examined. Since the May ratings were made any time from May to the following October, it was decided to use only the February ratings in these analyses because they are more representative of the overall level of implementation in the site during the year and are probably more valid. The analysis of variance of sponsors' ratings for February only can be found in Appendix C, Table 5.

models, we would not expect a knowledge of models to explain much of the variation in levels of implementation because in the analysis of variance of sponsor ratings, models explained only 3.1% of the total variance. As a check, however, we entered eleven of the twelve PV models as dummy variables to explain variation in the total system.[3] Table 27 shows the results of this analysis: models explain only 8.1% of the variation in sponsors' ratings of teachers. The F-test for this value is not significant ($p = .275$). In fact, only one model, REC, has a significant b coefficient (where "significance" can be described in terms of the extent to which a variable significantly adds to the equation predicting the dependent variable).[4] We can conclude, then, that model effects do not help to explain variation in levels of implementation. Therefore, we will ignore them in the rest of our analysis.

---

[3]The use of dummy variables is a technique for introducing variables which are not conventionally scaled into a regression equation. We define one dummy variable for each model: all teachers who work with a given model receive a 1 on the corresponding dummy variable, and all other teachers receive a 0. Only eleven models are entered as dummy variables because the twelfth is redundant in the sense that it adds nothing to the amount of explained variation. Moreover, any set of eleven models yields the same result.

[4]Inspection of the model means on sponsors' February ratings (Table 5, Chapter 2), shows that REC has the highest mean, but the model only has one site; there are other individual sites with equally high means.

TABLE 27

Regression Analysis
Model Differences in Explaining Total Variation in Levels of Implementation

Variable Name	b	SE_b	b*	T-Test	df	Significance
Far West	-0.1869	0.530	-0.038	-0.35	153	>.500
Arizona	0.0128	0.555	0.002	0.02	153	>.500
Bank St.	-0.4535	0.547	-0.088	-0.83	153	.409
Oregon	0.5971	0.630	0.091	0.95	153	.345
Kansas	0.0128	0.676	0.002	0.02	153	>.500
High Scope	-0.2244	0.559	-0.042	-0.40	153	>.500
Florida	0.0855	0.630	0.013	0.14	153	>.500
EDC	0.0128	0.690	0.002	0.02	153	>.500
Pittsburgh	-0.1342	0.916	-0.012	-0.15	153	>.500
REC	2.4224	0.869	0.240	2.79**	153	.006
NYU	0.0128	0.746	0.002	0.02	153	>.500
Regression Constant	5.102					

R^2 = 0.081

R = 0.285

SD_{res} = 1.724

F = 1.23 with 11 and 153 degrees of freedom (P = .275).

Notes: b = beta coefficient
b* = standardized coefficient
SE_b = standard error of the coefficient
R = multiple correlation
R^2 = multiple correlation squared
SD_{res} = standard deviation of residuals

Variation Among Sites:

It is apparent, then, that the variation in levels of implementation lies among sites and among teachers within sites. We are interested, first, in differences among sites, with sites considered both as independent and as dependent variables. At one level of analysis, we can use site differences as independent variables to explain the total variation in sponsors' ratings. At another level, we can use sites as dependent variables and try to explain the differences among them.

In order to deal only with variation among sites, as opposed to variation in the total system, we will consider variables on the "site level". Some variables, such as those from the site assessment, are already at the site level because there is only one observation per site. For other variables, including the sponsors' ratings of teachers, however, the original data has one observation per class (and consequently, there is variation within sites). For these variables, including the dependent measure, we will use site means in estimating site effects.

Because it is impossible to enter all possible variables in a regression equation, we selected interesting variables from among those which correlate most highly with the site mean of the sponsors' ratings of the teachers' performances. Table 28 shows the highest correlations,

TABLE 28

Site Level Correlations with Sponsor Ratings of Teachers

	Training:	r
SI#4b	Number days in-service training for aides	.58
SI#4b	Number days in-service training for teachers	.51
SI#4a	Number days pre-service training for aides	.43
SI#4a	Number days pre-service training for teachers	.44
TQ#11	Whether teacher requested help	.38
SI#4b	Number hours in-service training for aides	.37
TQ#9	Whether teacher had group discussion in in-service	.35
SI#4b	Number hours in-service training for teachers	.34
TQ#6	Whether teacher had group discussion in pre-service	.29
TQ#6	Whether sponsor gave some pre-service	.29
	Staff Input:	
AQ#26	Helpfulness of aide's training	-.54
TQ#42	Teacher's years in Head Start	-.48
TQ#38	Teacher's race	.45
AQ#10	Aide's satisfaction with equipment	-.44
TQ#40	Teacher had K-2 practice teaching	-.37
TQ#36	Teacher's age	-.34
	Context:	
SI#3.d	Intra-staff friction	-.80
SI#3.n	Rapport between administration and staff	.74
SI#3.o	Ratpport between staff and children	.62
SI#3.p	Rapport between sponsor staff and local staff	.46
SI#3.k	Adequacy of physical plant indoors	.40
CR#15	Rapport between adminstrator and staff	.40
CR#4	Intra-staff friction	-.39
SI#3.l	Availability of sponsor guidance	.39
SI#3.j	Support of PAC for the model	.39
++	Whether site housed and administered by public school	.37
CR#6	Puctuality of teachers	.36
SI#3.h	Support of local Head Start personnel for the model	.36
TQ#24	Teacher satisfaction with working conditions	-.33
CR#7	Regular attendance of children	.32
SI#3.i	Support of community for the model	.32
SI#3.k	Adequacy of physical plant outdoors	.31
++	Whether site housed by CAP	-.30

Note: These are Pearson product-moment correlations

which are divided into three groups corresponding to the three domains in Chapter 3 (with the exception that sponsor input will be defined more broadly as training). In order to explore the largest number of variables, we first ran a separate equation for each domain.[5] We then took the variables with the largest coefficients from each equation and combined them into one equation. In each case, the regressions were stepwise. Under this method the variables are entered one at a time, in order of maximum improvement to the regression equation; the variables which add most to the multiple correlation squared are entered first. This process can be stopped when the last variable adds less than 1% to the multiple correlation, or when the standard deviation of the residuals is at a minimum. The final equation for predicting site-mean levels of implementation shows a striking finding (see Table 29). We can explain 97.7% of the variation among sites in sponsors' ratings ($R^2 = 0.977$) with the knowledge of three variables: the amount of friction within the site, the rapport between the administration and the staff, and the adequacy of the physical plant inside--all as judged by the sponsor. (The ratings on the same dimensions by the OCD consultants

[5]When the number of predictor variables is large in relation to the number of observations in the study, the estimated true validity of the regression equation is very low. The three equations are shown in Appendix C, Tables 6, 7 and 8.

TABLE 29.

Accounting for Site-to-Site Variation in Levels of Implementation

	Variable Name	b	SE_b	b*	T-Test	df	Significance
SI#3.d	Intra-staff friction	-0.527	0.027	-0.658	-19.80	23	<.001
SI#3.n	Rapport between administration and staff	0.4232	0.031	0.466	13.50	23	<.001
SI#3.k	Adequacy of physical plant indoors	0.2025	0.025	0.265	8.05	23	<.001
Regression Constant		4.573					

R^2 = 0.977
R = 0.988
SD_{res} = 0.161

F = 326.96 with 3 and 23 degrees of freedom ($p < .001$)

Partial Correlations with Dependent Variable for Variables Not Entered:

AQ#2b	Helpfulness of aide's training	-1.278
SI#4b	Number days in-service training or teachers	0.134
SI#3o	Rapport between staff and children	1.021

Note: This is a stepwise regression: variables were allowed to enter in order of importance

b: regression coefficient
b*: standardized regression coefficient
R^2: multiple correlation coefficient squared
SI: Sponsor Implementation Report
AQ: Aide Questionnaire

do not enter). The friction variables has a negative coefficient, which indicates that a low level of friction is associated with a high level of implementation. The other two variables have positive coefficients, which indicate that high rapport and good facilities are associated with high implementation. Clearly, these variables together present a picture of a pleasant site. Moreover, because the friction and rapport variables alone explain 90.6% of the variance, it appears that harmony is the primary component of a "pleasant site".

Several explanations of this finding are plausible. It is possible that sponsors are simply equating pleasant sites with high levels of implementation: site differences may reflect only differences in pleasantness, and may be unrelated to actual differences in performance with the model. This interpretation is supported by the fact that consultants' judgments of the same three variables (friction, rapport, and adequacy of facilities) are not related to levels of implementation. It is possible, however, that pleasant sites actually do have higher levels of implementation than do unpleasant sites. This may mean, on the one hand, that high levels of implementation occur in sites which have good facilities and good relations among the staff before the introduction of the model; a pleasant site may be a prerequisite for implementation. On the other hand, it may be that the model creates the pleasant environment: the staff in a site

may be so enthused by the model, that they work well together in order to implement it. Although the available data do not allow us to resolve the differences in interpretation about this relationship, the importance of this finding should be recognized.[6]

Variation Within Sites:

Explaining almost all of the site-to-site variation and knowing there is almost no model-to-model variation leaves us only with classroom within site variation to explain. To deal with within-site variation, we will use a class level analysis. At this level, however, we are dealing with all the variation in the system, including

[6]Three other variables--the aide's judgment of the helpfulness of training, the sponsor's report of the number of days of in-service training given to the teachers, and the sponsor's judgment of the rapport between the Planned Variation staff and the children--were in the analysis but did not enter the equation because the large partial correlations mean that the correlation matrix is not positive-definite.

site-to-site variation as well as within-site. One strategy for removing site effects in order to deal only with variation within sites would be to enter all sites as dummy variables, as we did with models. This, however, would use up a large number of degrees of freedom (number of sites - 1) and the subsequent estimates would be less precise than if fewer degrees of freedom had been used. An alternative strategy, and the one which we followed, is to use the three variables which explain 98% of the variation between sites (staff friction, staff rapport, and adequacy of the physical plant) as proxies for sites. Table 30 shows that these variables account for 29.8% of the total variation in levels of implementation.[7] If we force them in an equation first, we remove the site-to-site variation and are essentially left with within site variation.

The highest class-level correlations between levels of implementation and a range of independent variables in the three domains are shown in Table 31. These correlations are obviously lower than the correlations at the site level. We followed generally the same procedure in developing within-site equations as we did with the site level. However, because of the small number of even

---

[7] It should be noted that in explaining 29.8% of total variance with site variables, we have exceeded the amount expected on the basis of the amount of variance due to sites in the analysis of variance for February sponsor ratings (22%). This is probably the result of different data bases: the regressions are based on data from all classes for which there were sponsor ratings, which is more data than were included in the balanced anova (in standard design).

TABLE 30

Site Differences Explaining
Variation in Levels of Implementation

	Variable Name	b	SE_b	b*	T-Test	df	Significance
SI#3.d	Intra-staff friction	-0.5301	0.104	-0.388	-5.11	161	<.001
SI#3.n	Rapport between administration and staff	0.4922	0.171	0.240	2.88	161	.005
SI#3.k	Adequacy of physical plant indoors	0.0862	0.126	0.052	0.68	161	.495
Regression Constant		4.637					

R = 0.298

R = 0.546

SD_{res} = 1.469

F = 22.75 with 3 and 161 degrees of freedom ($p < .001$)

Note: The variables were forced into the equation.

b: regression coefficient
b*: standardized regression coefficient
R^2: multiple correlation coefficient squared
SI: Sponsor Implementation Report

TABLE 31

Class Level Correlations With Sponsor Ratings of Teachers

	Sponsor Input:	r
TQ#9	Number types of in-service training given by local Head Start office	-.22
TQ#11	Whether teacher requested help	.18
TQ#6	Whether local Head Start office gave some pre-service training	-.17
TQ#9	Whether teacher had group discussions in in-service training	.17
TQ#5	Number days pre-service training received	.15
TQ#6	Whether sponsor gave some pre-service training	.15
	Staff Input:	
TQ#42	Teacher's years experience in Head Start	-.26
TQ#36	Teacher's age	-.22
TQ#38	Teacher's race	.21
	Context:	
TQ#39	Degree of parent involvement	-.15
TQ#24	Teacher's satisfaction with working conditions	-.13

Note: These are Pearson product-moment correlations. They include all scores, not just relations within sites.

moderate correlations, we only used two equations to begin rather than three: one equation with training variables, and the other with both staff input and context variables. Tables 9 and 10 in Appendix C show these equations. It is apparent that we cannot explain much of the within-site variation in sponsor ratings with either of these equations. Knowledge of training variables adds 8.1% to the explained variance above site differences, and knowledge of three staff input and context variables accounts for 1.4% of the variance after site-to-site differences have been removed.

It is also apparent that the adequacy of facilities variable is not adding significantly to the equation. Since the goal of a regression analysis is to explain the most variation with the fewest variables, this variable was dropped from the analysis. In the site level equations, removal of this variable reduces the percentage of explained variance from 97.7% to 91.3%, but in the class-level equations, it only reduces the percentage of total explained variance from 29.8% to 29.6%. Thus the two remaining variables, staff friction and staff rapport, can still be considered as proxies for between-site effects.

The final equation was formed by first entering the two site proxy variables, and then allowing the best predictors from the previous equations to enter in order of importance. Table 32, which contains the results of this regression analysis, shows that even with the strongest variables combined, we can explain only 39% of the total

TABLE 32

Predicting Levels of Implementation

	Variable Name	b	SE_b	b*	T-Test	df	Significance
SI#3.d	Intra-staff friction	-0.4437	0.098	-0.325	-4.51	157	<.001
SI#3.n	Rapport between administration and staff	0.3928	0.146	0.192	2.68	157	.009
TQ#9	Number types in-service training by local HS	-0.1034	0.052	-0.148	-2.00	157	.048
TQ#9	Whether teacher had group discussions	1.1247	0.402	0.183	2.79	157	.006
TQ#6	Whether sponsor gave some pre-service	0.5522	0.237	0.159	2.33	157	.021
TQ#6	Whether local HS gave some pre-service	-0.5529	0.261	-0.160	-2.12	157	.036
TQ#42	Teacher's years experience in HS	-0.1427	0.06	-0.139	-2.07	157	.041

Regression Constant 4.682

R^2 = 0.390
R 0.624
SD_{res} 1.387

F = 14.31 with 7 and 157 degrees of freedom ($p < .001$)

Partial Correlations with Dependent Variable for Variables Not Entered

TQ#11	Whether teacher requested help	0.082
TQ#38	Teacher's race	0.057
TQ#36	Teacher's age	-0.078

Note: The first two variables were forced into the equation, after which the other were allowed to enter in order of importance.

b: regression coefficient
b*: standardized regression coefficient
R^2: multiple correlation coefficient squared

SI: Sponsor Implementation Report
TQ: Teacher Questionnaire

variation in sponsors' ratings of levels of implementation. More importantly, we can only explain 9.4% of the variance with within-site variables, after site effects have been removed. The five variables which contribute to the 9.4% do have significant coefficients, however, and deserve to be considered individually. All five variables come from the Teacher Questionnaire; four relate to pre-service and in-service training and one is a teacher characteristic.

We will consider the teacher characteristic, number of years in Head Start, first, even though it is the last variable to enter, because it is the easiest to interpret. The coefficient for this variable is negative indicating that a teacher with more experience teaching in Head Start tends to implement the model to a lesser extent than a teacher with fewer years of experience. It may be that a well-established teacher has a comfortable pattern of instruction which she does not want to give up in order to fully implement a model.

The relationships of training variables to sponsors' ratings of implementation are more difficult to deal with because it is not clear whether the within-site variation on these measures results from real differences in experience or whether they reflect differences in memory or perception of the same experience. If we assume real differences, we can conclude from the two variables with positive coefficients that the teachers who received some

pre-service training from sponsors, and who participated in group discussions during in-service training have higher levels of implementation. These findings seem reasonable. It is logical--and in a sense, reassuring--that sponsor training has some impact on levels of implementation. Perhaps, the second variable is important because teachers need to talk things over in order to gain mastery of the model. From the two training variables with negative coefficients (still assuming real differences in experience), our conclusions are less clear. The regression analysis indicates that teachers who receive training, both in-service and pre-service, from the local Head Start office tend to have lower levels of implementation than teachers who did not receive training from this source. It may be that local training interferes with implementation in some way, either by directly contradicting the model, or by adding inputs which overload and confuse the teachers. If the second interpretation of within-site variation is correct, that the variation comes from different perceptions rather than different experiences, the conclusions are slightly different. We might speculate that the teacher who implements the model well is one to whom sponsor training is important. As a result, she emphasizes it more in her recall of what happened than do other teachers. Conversely, a teacher who does not implement the model may be the one who has focused on local training, and therefore,

remembers it more. The causal sequencing of these relationships is not clear. Regardless of interpretation, however, the finding that training from the local Head Start office is negatively related to implementation is important.

SUMMARY

Using regression analyses to explore the relation between a measure of the extent of model implementation and a variety of independent variables, we have made a number of findings.

1. We can explain only 39% of the variation in levels of implementation with the data available to us in this study.
2. Model-to-model differences explain an insignificant amount of the variation.
3. The within-site variation is the largest part of the total variance, but we cannot explain it well: after controlling for site-to-site differences, less than 10% of additional variance can be accounted for. The five variables which contribute to this 10% are, in order of importance, the number of types of in-service training given by the local Head Start office, whether there were group discussions during in-service training, whether the sponsor gave some pre-service training, whether the local Head Start office gave some pre-service training, and the teacher's years of experience in Head Start. All of these variables come from the Teacher Questionnaire and therefore, have within-site

variation.

4. The most impressive finding is that 97.7% of the site-to-site variation in levels of implementation can be explained with three variables, all of which are sponsor's judgments: intra-staff friction, rapport between the site administration and staff, and the adequacy of the physical plant indoors. Although several alternative interpretation of this finding are possible it appears that it indicates that sponsors are equating pleasant sites with high levels of implementation.

Chapter 5

CLASSROOM OBSERVATION

This chapter reports how model classrooms differ from one another in practice on several measures of adult and child behavior. The first section describes the classroom observation instrument, the fifty-one dependent variables derived from it, and the statistical analysis performed on these variables. The second section relates the results of the analysis on twenty-two variables--those variables which seemed to reflect genuine differences among models. In the third section, we attempt to summarize each model in terms of its scores on variables which should represent important aspects of model theory, and in terms of the variables on which each model is significantly different from other models. Thus, we report the data in two ways: first, in a straightforward manner by variable, and second, in an interpretive manner by model.

A complete list of the fifty-one variables is included in Appendix F , and tables of means and standard deviations for all eleven models on all the variables are in Appendix G. Although detailed descriptions of model theories are not included in this chapter, the reader can refer to "Design, Data Collection Activities, and the Curriculum Approaches" for more information about each sponsor.

CLASSROOM OBSERVATION INSTRUMENT

The Classroom Observation Instrument is a standardized instrument designed by SRI to assess child and adult behavior and interactions, types of activity, general atmosphere, and physical equipment in a classroom. Unfortunately, the variables used in this analysis are derived from only two of the six parts of the instrument.[1] These two parts, the Classroom Check List and the five-minute observations, describe types of activity and interactions. On the Classroom Check List, the observer codes all individuals and their activity for one minute intervals four times an hour. The five-minute observations consist of coding the interactions of a selected group of children and adults in a specific activity. The coding is in a sentence format with four components: who--to whom--does what--and how it is done. For example, a teacher (T) to one child (C) asks a direct question (1) firmly (F). Thus, the code sentence would be T-C-1-F. Or, a small group of children (S) among themselves (S) play (5) happily (H); S-S-5-H. An observer is expected to complete up to sixty frames (one frame includes one "code sentence") per five-minute observation or one frame every five seconds.

---

1. The full instrument and instructions for its use are described in SRI, Implementation of Planned Variation in Head Start (1969-1970), Appendix B.

Observers were trained by SRI and assigned to visit each Planned Variation site once in the fall of 1970 and once in the spring of 1971. Each observer spent two mornings in up to four Planned Variations classrooms and up to four comparison or non-PV classrooms. While the same person observed both PV and non-PV classrooms at a site, in about half of the sites, a different person observed in the fall and spring. PV classrooms were selected for observation on the basis of the teachers' years of experience with the model. In other words, SRI chose those classrooms which were more likely to be well implemented, based on the hypothesis that teachers with more experience in the model are better able to implement the specifications of the model. Comparison classrooms were selected on the basis of the likelihood that the children would later be in Follow-Through classrooms, in order to increase the probability of a large sample in a follow-up study.

Our analysis thus includes a fairly large amount of data on PV and comparison classrooms. Although the data do not cover all of the aspects of classroom behavior that we might want to examine, they do provide information about various kinds of activities and how children and adults relate to one another.

VARIABLES

The fifty-one dependent variables[2] on which our analysis is based are only a fraction of the entire group of variables that could be extracted from the classroom observation data. These fifty-one variables were designed to represent important aspects of Planned Variations models. Each variable should, theoretically, reflect differences among models if the models are implemented according to the sponsors' specifications. However, many of the variables are relevant to more than one model: for example, six sponsors emphasize "independent child activity" and no sponsor stresses "large group activity." (For further examples, the reader can refer to the tables of theoretically important variables for each model in the Model Summaries section). Thus, since several of the models have more similarities than differences in their theories, one cannot expect that each variable should differentiate every model. One can expect, however, that most of the variables should show differences among some models.

There are two further qualifications to the statement that the variables represent aspects of model theories. First, none of the variables is relevant to two models: University of Florida stresses training parents to teach in the home rather

---

[2] A list of all fifty-one variables is included in Appendix F.

than teaching children in the classroom; Enablers have no prescribed theory, but rather "enable" Head Start teachers to implement an educational program based on the needs and desires of the parents and community.

Second, because the variables are derived from the Classroom Check List and five-minute observations, both of which report readily observable behavior and require rapid coding, the variables are more relevant to those models which specify particular types of activity and interactions. In other words, we cannot construct variables from the Classroom Check List or five-minute observations which measure whether a teacher is responsive to children's needs, or whether a child is alert and interested in the classroom activity. In contrast, we can design variables which represent "academic activity," "children asking questions," "independent play." Thus, if a model specifies a certain amount of academic activity or emphasizes independent child activity, we have data on which to determine whether, in fact, these types of activity are occurring in their classrooms. However, if a model specifies no particular type of activity or curriculum, but rather that adults should aim to provide a setting in which children are encouraged to make their own discoveries, we do not have data on which to judge whether this setting is actually present. Therefore, the analysis of these data cannot be used to answer questions such as, "Which model is most successful in implementing its theory?" This analysis does

answer such questions as, "Which model has more academic activity in its classrooms?"

STRATEGY FOR FINDING MODEL DIFFERENCES

In order to find "model effects" or significant differences among models on our variables, we first examined means and standard deviations by model[3] and by site[4]. Then, we ran an exact least-squares analysis[5] which included all the data and provided F-tests for model effects, site-within-model effects, PV/non-PV effects, PV/non-PV by model interactions, and PV/non-PV by site interactions for all variables.

Originally, we planned to use PV/non-PV effects as a criterion for substantial model effects, under the assumption that if non-PV classroom scores were like PV classroom scores on some measure, then the PV scores did not indicate model influence. However, several factors persuaded us to change that strategy. First, the same person observed both PV and non-PV classrooms at a site, and if that observer

---

[3] 11 models with PV classrooms: fall and spring;
9 models with non-PV classrooms: fall and spring.

[4] 29 sites with PV classrooms: fall
26 sites with PV classrooms: spring
20 sites with non-PV classrooms: fall
18 sites with non-PV classrooms: spring

[5] Developed by Jeremy D. Finn of State University of New York at Buffalo.

coded differently enough from other observers, then the lack of difference between PV and non-PV classrooms could be a measure of observer effects. Second, we found that some PV classrooms were in the same building with non-PV classrooms and that PV teachers exchanged equipment and instructions with non-PV teachers. Thus, insignificant differences between PV and non-PV classrooms could indicate a high level of interaction among PV and non-PV teachers. Third, two of the eleven models had no comparison classrooms to be observed. Therefore, we decided to ignore non-PV scores and to run all further analyses with PV classroom data only.

Using PV classrooms only, then, we ran another least-squares analysis which provided F-tests for model effects and for site-within-model effects. This analysis also gave us t-values for each model and for sites-within-models on all variables. We recognize the problem of multiple comparisons in this analysis and intend to use these t-values merely as informal indicators of significance. Thus, with degrees of freedom close to 100, an effect in this chapter is significant at the .01 level if its t-value exceeds 2.6 and significant at the .001 level if its t-value exceeds 3.4.

We also ran correlations with only the PV data on all variables fall to fall, spring to spring, and fall to spring. We used the fall to fall and spring to spring correlations to test hypotheses about relationship among variables, i.e.,

if certain kinds of adult behavior correlated highly with certain kinds of child behavior. We used the fall to spring correlations to determine whether scores on a variable in the fall correlated with scores on that same variable in the spring. Because these fall to spring correlations were low on over a third of the variables, and because we found from the t-values that model effects often changed from fall to spring, we ran another exact least-squares analysis on difference scores (spring score minus fall score) to get t-values for model change effects. We looked, too, at F-tests for changes over all models from fall to spring on each of the variables. This analysis was helpful in indicating whether change over all models on a variable was significant, and in showing which models changed significantly more than others.

With these analyses, we examined each variable for significant model effects which could be interpreted as actual differences among models. Roughly speaking, an effect for a model was "convincing" if its t-value was significant at the .01 level and if site effects appeared less important than the model effect. In other words, if it was clear that only one site in a model with two or three sites scored higher than the grand mean on a variable and the other site(s) scored well below the grand mean, it could hardly be said that that "model" was significantly high. However, since we have five models with three sites, four models with two sites, and two

models with only one site, the "site effect importance" criterion could not be applied rigidly. Throughout this chapter, we report site effects where they appeared important.

We interpreted large fall-to-spring differences when an explanation seemed plausible and otherwise attributed the change to observer effects and to natural day-to-day change in a Head Start center.

Thus, we used all of our data together to explore for patterns of model effects. The F-tests for model effects were significant at the .001 level on almost all of our variables, but because site effects were also significant and because there were fall to spring differences, we could not assume the existence of actual differences in practice among models from the significance of the F-tests. If fall-to-spring changes on a variable were significant, and if site variation was substantial, then we assumed the variable to be of little value in finding differences in practice among models. Although it is possible to interpret site effects and possible to interpret fall to spring change, it is not possible to find convincing model effects when both site effects and fall-to-spring change occur together on the same variable. The following hypothetical tables illustrate (1) fall to spring change without site effects (which is interpretable); (2) site effects without fall to spring change

(which is also interpretable); and (3) site effects with fall to spring change (which, regrettably, is uninterpretable). We do not intend these hypothetical tables to represent actual variables in our analysis, but to serve as "caricatures" of the problems involved in finding convincing model effects.

HYPOTHETICAL TABLES

Situation 1: Fall to spring change without important site effects

	Fall	Spring
Model 1	Site A = 2	Site A = 5
	Site B = 3	Site B = 4
	Site C = 4	Site C = 6
	Model 1 Mean = 3	Model 1 Mean = 5
Model 2	Site E = 6	Site E = 3
	Site F = 3	Site F = 2
	Site G = 6	Site G = 4
	Model 2 Mean = 5	Model 2 Mean = 3
Model 3	Site H = 4	Site H = 6
	Site I = 5	Site I = 7
	Site J = 3	Site J = 5
	Model 3 Mean = 4	Model 3 Mean = 6

In Situation I, we see that model 2 is high in the fall and that model 1 is low. In the spring, however, model 3 is high and model 2 is low. Although the sites did not change at the same rate (for example, in model 2, site E decreased 3 points, site F 1 point, and site G 2 points), all of the

sites in one model either increased or decrea from fall to spring. Thus, if this variable were "children having tantrums," we could say that Models 1 and 3 showed an increase in tantrum behavior and that model 2 showed a decrease. This variable might correlate highly with other variables, such as "adult negative behavior," or it might not be correlated with any other variable. In any case, we can draw some conclusions about differences among models.

Situation II: Important site effects without fall to spring change

	Fall	Spring
Model 1	Site A = 3	Site A = 2
	Site B = 2	Site B = 3
	Site C = 1	Site C = 1
	Model 1 Mean = 2	Model 1 Mean = 2
Model 2	Site E = 8	Site E = 6
	Site F = 1	Site F = 2
	Site G = 3	Site G = 4
	Model 2 Mean = 4	Model 2 Mean = 4
Model 3	Site H = 3	Site H = 3
	Site I = 9	Site I = 8
	Site J = 3	Site J = 4
	Model 3 Mean = 5	Model 3 Mean = 5

Situation II shows model 3 consistently high and model 1 consistently low, with model 2 slightly above the grand mean. However, site E in model 2 and site I in model 3 are the only sites substantially above the grand mean. Thus, if

this variable were "children asking questions," we can say that children ask fewer questions in model 1 than in models 2 and 3, but that the higher scores in models 2 and 3 are mainly due to high scores in sites E and I. Although conclusions about model differances must be weak with such large site to site variation, at least the consistency of the scores indicates that there might be genuine differences on this variable among sites, if not among models.

Situation III: Important site effects with fall to spring change

	Fall		Spring	
Model 1	Site A	= 6	Site A	= 1
	Site B	= 2	Site B	= 10
	Site C	= 1	Site C	= 4
	Model 1 Mean	= 3	Model 1 Mean	= 5
Model 2	Site E	= 8	Site E	= 2
	Site F	= 1	Site F	= 6
	Site G	= 3	Site G	= 3
	Model 2 Mean	= 4	Model 2 Mean	= 3
Model 3	Site H	= 3	Site H	= 2
	Site I	= 9	Site I	= 3
	Site J	= 3	Site J	= 7
	Model 3 Mean	= 5	Model 3 Mean	= 4

Situation III illustrates the impossibility of finding convincing model effects when both site effects and fall to spring change occur together on a variable. No site or model is consistently higher than others. Although no variable in our actual analysis looked quite as helter-skelter as situation III,

this example certainly shows that we can draw no conclusions about differences among models when site effects and fall to spring change are significant on the same variable.

In the following section, we discuss only those variables which indicate convincing model effects. We do not discuss variables one through seven for two reasons. First, they represent the type of activity that observers selected for their five-minute observations, and thus only indirectly suggest the frequency of occurrence of these activities in the classroom. Second, only one of these seven reflects actual differences among models, and since that one is "academic activities" and we have five other variables to describe academic activity, we decided to disregard the first seven as a group.

Also, we do not discuss the few variables which showed insignificant F-tests for model effects. Neither do we discuss some variables which are combinations of other variables; e.g., we disregard "adult praise and acknowledgment" but rather deal with "adult praise" and "adult acknowledgment separately. Although we recognize that disregarding variables is a rather dangerous procedure, we feel fairly confident that we have included those variables which show genuine differences in practices among models.

DISCUSSION OF VARIABLES

According to our strategy, twenty-two of the total fifty-one variables indicate substantial differences among models. These twenty-two are divided into three groups for purposes of discussion: academic activity, independent child activity, and adult/child interactions. Along with the discussion of each variable is a table showing mean scores for each model. At the end of each of the three groups are summary tables showing which models are significantly high or low on those measures. In Appendix G are tables of means, standard deviations, and sample size for each model on all fifty-one variables.

I. Academic Activity:

Var 9: academic activity (frequency of occurrence)
8: adults with children in academic activities
13: aide's participation in academic activities
43: teachers and volunteers with children in academic activities
42: independent children in academic activities

Academic activity is defined as Activity C and Activity D (see variables 3 and 4 in Appendix F). "Academics" do not include games and puzzles, but do include numbers, alphabet, and "science" activities. It is interesting that none of the above variables changes significantly over all models from fall to spring. Although variables 9 and 42 show some significant within model changes from fall to spring, this group of variables seems to represent one of the more stable elements in classroom activity.

Further, site effects are less marked on academic activity variables than on any other group. Clearly, then, evidence

of significant model effects here reflect actual differences in practice among models.

Var 9: academic activity (frequency of occurrence)

The table on variable 9 shows Bank Street, and Kansas significantly high in the fall and Oregon, Kansas and Pittsburgh high in the spring. The fall Bank Street score is misleading because only the Wilmington site is above the grand mean, while the other two Bank Street sites are well below. Wilmington is still high in the spring, but not high enough for Bank Street's score to reach significance.

Pittsburgh's high spring score is puzzling: Pittsburgh has only one site and we cannot guess whether the fall observer saw more "typical" activity than the spring observer. We might assume, given that Pittsburgh was new to Planned Variation in the fall of 1970 and that the model emphasizes academic activity, that the model shows better implementation in the spring.

The Oregon and Kansas scores show strong evidence of a great deal of academic activity in their classrooms. Site effects are unimportant; fall to spring changes are within the limits of normal fluctuation; and both models stress academic learning in their classroom theory. Thus, it is fairly certain that children enrolled in Oregon or Kansas Head Start programs will be involved in more academic activity than children attending other Planned Variation programs.

Mean Scores by Model on Classroom Observation

Var. 9. Academic Activity (Frequency of Occurrence

	Fall	Spring
Far West	0.506	0.425
U. of Ariz.	0.484	0.717
Bank St.	1.029+	0.785
U. of Oregon	0.969	1.194++
U. of Kansas	1.220++	1.164++
Hi/Scope	0.207-	0.225--
U. of Florida	0.468	0.197--
EDC	0.323	0.458-
Pittsburgh	0.807	1.527++
REC	0.757	0.749
Enablers	0.305-	0.412-
Grand Mean	0.634	0.665

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Var 8: adults with children in academic activities

Variable 8 reaffirms our earlier statement about Oregon and Kansas classrooms and shows that most of the academic activity occurring there includes both adults and children. The scores on variable 8 for the Oregon and Kansas models are very similar to their scores on var 9, indicating that almost no academic activity occurs that is not directed, or at least observed, by adults. In contrast, Pittsburgh's high spring score on variable 9 is not repeated on variable 8, indicating that a high percentage of Pittsburgh's academic activity does not include both adults and children. We will investigate this hypothesis later with variable 42 (independent children in academic activity).

Var 13: aide's participation in academic activities

On variable 13, we see that the Kansas model remains high, indeed high enough to indicate that aides participate in more than half of the total academic activity in the classrooms (see variable 9). The Oregon model is only slightly higher than the grand mean in the fall and although it is significantly higher than the grand mean in the spring, only one of their sites (E. Las Vegas) is high. The other site (Tupelo) is quite low on this variable both in the fall and in the spring. Thus, because of large site effects in the

Mean Scores by Model on Classroom Observation

Var. 8. Adults with Children in Academic Activities

	Fall	Spring
Far West	0.209	0.228
U. of Ariz.	0.349	0.471
Bank St.	0.430	0.402
U. of Oregon	0.936+	1.126++
U. of Kansas	1.206++	1.124++
Hi/Scope	0.167--	0.174
U. of Florida	0.419	0.117--
EDC	0.242	0.301
Pittsburgh	0.542	0.649
REC	0.633	0.482
Enablers	0.206-	0.295
Grand Mean	0.476	0.483

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

Var. 13. Aide's Participation in Academic Activities

	Fall	Spring
Far West	0.059	0.087
U. of Ariz.	0.097	0.195
Bank St.	0.154	0.111
U. of Oregon	0.242	0.378+
U. of Kansas	0.791++	0.776++
Hi/Scope	0.032-	0.068
U. of Florida	0.170	0.046
EDC	0.087	0.093
Pittsburgh	0.148	0.207
REC	0.252	0.021
Enablers	0.061	0.077
Grand Mean	0.201	0.204

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Oregon model, we conclude that only the Kansas model is significantly high on variable 13.

Var 43: teachers and volunteers with children in academic activities

Variable 43 provides a distinct contrast to variable 13. This variable shows Oregon significantly high (at .001 level) on both observations and Kansas slightly higher than the grand mean. Thus, we have a further refinement of variable 9: more than half of the Oregon model's academic activity occurs with teachers and/or volunteers, while more than half of the Kansas model's occurs with aides.

Var 42: independent children in academic activities

Variable 42 completes the academic activity spectrum. Here the Oregon and Kansas models are low, as we might have predicted from their high scores on adults with children in academic activities. Bank Street is significantly high, but as with variable 9, only the Wilmington score is above the grand mean. The Pittsburgh model is high in the spring, indicating that, like the Wilmington site in the Bank Street model, a large proportion of their academic activity occurs without adult direct supervision.

Mean Scores by Model on Classroom Observation

Var. 43. Teachers and Volunteers with Children in Academic Activities

	Fall	Spring
Far West	0.150	0.141⁻
U. of Ariz.	0.252	0.276
Bank St.	0.276	0.291
U. of Oregon	0.694++	0.748++
U. of Kansas	0.415	0.349
Hi/Scope	0.135⁻	0.106--
U. of Florida	0.249	0.071--
EDC	0.155	0.208
Pittsburgh	0.394	0.442
REC	0.380	0.461
Enablers	0.145	0.218
Grand Mean	0.275	0.278

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

Var. 42. Independent Children in Academic Activities

	Fall	Spring
Far West	0.297	0.196
U. of Ariz.	0.135	0.245
Bank St.	0.599++	0.383+
U. of Oregon	0.033	0.068
U. of Kansas	0.014	0.040--
Hi/Scope	0.040	0.051-
U. of Florida	0.049	0.079
EDC	0.081	0.157
Pittsburgh	0.265	0.878++
REC	0.124	0.267
Enablers	0.099	0.117
Grand Mean	0.158	0.182

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

TABLE 33

First Summary Table

	Academic Activity		Adults wth. Childrn. in Acad. Acts.		Aide's participtn. in Acad. Acts.		Tchs. & vols wth. childrn in Ac. Acts.		Independ. childrn. in Ac. Acts.	
	Fall	Spr.	Fall	Spr.	Fall	Spr.	Fall	Spr.	Fall	Spr.
Far West								-		
U. of Ariz.		-								
Bank St.	+								++	+
U. of Oreg.		++	+	++		+	++	++		
U. of Kan.	++	++	++	++	++	++				--
Hi/Scope	-	--	--		-		-	--		-
U. of Fla.		--		--				--		
EDC		-								
Pittsburgh		++								++
REC										
Enablers	-	-	-							

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

II. Independent Child Activity:

Var 42: independent children in academic activities
10: independent child activity
44: independent children in non-academic activities
11: wide variety of activities
23: child self-learning with concrete objects

The evidence which differentiates models on this group of variables is less clear-cut than on academic activities variables. Two of the five variables (10 and 23) in the independent activity group change significantly from fall to spring over all models. However, all of these changes are positive, indicating, perhaps, that children who are more familiar with their classroom environment are more likely to behave independently. Within-model significant changes are comparatively few. Thus, fall-to-spring differences present only a minor difficulty.

Site effects, on the other words are large, especially in the Bank Street and Far West models. One would expect more site effects on this group of variables than on academic activity variables, since, as anyone who has ever worked with preschool children knows, it is often difficult to "structure" independent activity. Although adults in the classroom can encourage and reward independent activity, they cannot "schedule" it in the same way that they can "schedule" more structured academic activity. Further, the level of independent activity may be influenced by factors outside the control of teachers in the classroom. For example, independent acti-

vity might be higher just before holidays when children are keyed up, or when new toys or equipment are introduced into the classroom.

For these reasons, then, the requirements for significant model differences in this area are somewhat less stringent than in other areas.

Variable 42 was discussed in the previous section on academic activity, and shows mainly that the Wilmington site in the Bank Street model is outstandingly high, and that Pittsburgh is significantly high in the spring.

Var 10: independent child activity

The table on variable 10 indicates that the Bank Street model is significantly high in both fall and spring observations. However, some site effects confuse this finding. In the fall, Wilmington's mean is far higher than any other site, thus making Bank Street's mean significantly high. The other two sites (Boulder and Elmira) are both near the grand mean. In the spring, the Wilmington mean is lower than in the fall, but Elmira's mean is higher. Boulder is still near the grand mean. Thus, site effects, although less evident in the spring, are important in the Bank Street model.

The Pittsburgh model is significantly high only in the spring. As on two of the academic activity variables, this model's scores rise significantly from fall to spring. Since the Pittsburgh model was new to Planned Variation in the fall

Mean Scores by Model on Classroom Observation

Var. 10. Independent Child Activity

	Fall	Spring
Far West	1.656	1.461
U. of Ariz.	1.380	1.264
Bank St.	2.654++	2.450++
U. of Oregon	0.249-	0.677
U. of Kansas	0.561	0.550--
Hi/Scope	0.720	0.998
U. of Florida	0.649	0.922
EDC	0.843	1.138
Pittsburgh	1.244	2.454+
REC	1.484	1.532
Enablers	1.108	1.536
Grand Mean	1.141	1.296

Fall to Spring change over all models +:

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

of 1970, we attach more importance to its spring scores.

Var 44: independent children in non-academic activities

Variable 44 shows Bank Street again significantly high, but with the same type of site effects that occurred on variable 10.

Var 11: wide variety of activities

We include this variable in the independent activity group because it correlates so highly with variable 10 (.899 in the fall; .849 in the spring). Bank Street is high, but with the Wilmington site again far higher than any other site in the analysis. Boulder and Elmira are slightly above the grand mean on both fall and spring observations.

The fact that the Florida model is low on both observations may be the result of that model's emphasis on parent instruction rather than classroom activity.

Var 23: Child self-learning with concrete objects

Variable 23 is somewhat confusing: the variable includes three different observer codes: the code for respond, the code for comment, play, and the code for teach, inform. We are unsure whether a child playing with a truck or dolls would be included in this variable or whether "concrete objects" refer only to learning tools, such as Cuisenaire rods.

Mean Scores by Model on Classroom Observation

Var. 44. Independent Children in Non-Academic Activities

	Fall	Spring
Far West	1.358	1.265
U. of Arizona	1.245	1.018
Bank St.	2.054++	2.066++
U. of Oregon	0.217-	0.609
U. of Kansas	0.547	0.511-
Hi/Scope	0.680	0.947
U. of Florida	0.600-	0.843
EDC	0.761	0.981
Pittsburgh	0.979	1.576
REC	1.360	1.265
Enablers	1.008	1.419
Grand Mean	0.983	1.113

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

Var. 11. Wide Variety of Activities

	Fall	Spring
Far West	2.605	2.361
U. of Ariz.	1.987	1.915
Bank St.	3.836++	3.786++
U. of Oregon	1.413-	1.897
U. of Kansas	2.019	1.615
Hi/Scope	1.791	1.935
U. of Florida	1.553	1.559
EDC	1.647	1.758
Pittsburgh	1.835	2.658
REC	2.050	2.180
Enablers	1.860	2.173
Grand Mean	2.112	2.154

Note:- In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; --- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

Var. 23 Child Self-Learning with Concrete Objects

	Fall	Spring
Far West	1.841+	2.914++
U. of Ariz.	0.998	1.511
Bank St.	1.563+	1.156
U. of Oregon	0.0	0.552
U. of Kansas	0.050	1.005
Hi/Scope	0.111	0.052
U. of Florida	0.216	0.603
EDC	0.839	0.484
Pittsburgh	0.0	1.494
REC	0.835	1.564
Enablers	0.031	0.071
Grand Mean	0.578	0.910

Fall to Spring change over all models +.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

TABLE 34

Second Summary Table

	Ind. Child Activity		Independ. Childrn in Acad. Acts.		Ind. Chldrn. in non-Acad. Acts.		Wide Variety of Activities		Child self-learng wth concrete obj	
	Fall	Spr.	Fall	Spr.	Fall	Spr.	Fall	Spr.	Fall	Spr.
Far West									+	++
U. of Ariz.										
Bank St.	++	++	++	++	++	++	++	++	+	
U. of Oreg.	-				-		-			
U. of Kan.		--		--		-				
Hi/Scope				-						
U. of Fla.					-					
EDC										
Pittsburgh		+		++						
REC										
Enablers										

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

We tend to assume the latter, since the scores do not clearly resemble other independent activity scores.

With this confusion in mind, then, we see that the Far West model is significantly high in both the fall and spring. Site effects are important here, since the Tacoma site is below the grand mean. Duluth is far above the grand me n, as is Boulder in the Bank Street model. Thus, Duluth and Boulder are the only two sites in the analysis that are consistently high. Given Far West's strong emphasis on "autotelic" (self-rewarding) activities, we conclude that the Duluth site is well implemented in this area.

III. Child/Adult Interactions

A. Adult focus on groups of children

Var 12: adult interactions with one or two children
29: adult communication focus: one child
30: adult communication focus: small group

This group of variables provides information about how many children an adult in the classroom addresses at one time. Since these variables represent adult behavior rather than child behavior and since models are implemented through the adults in the classroom, we would expect these variables to show important model effects.

Var 12: adult interactions with one or two children

Variable 12 shows Pittsburgh and Bank Street consistently

Mean Scores by Model on Classroom Observation

Var. 12. Adult Interactions With One or Two Children

	Fall	Spring
Far West	0.320	0.362
U. of Ariz.	0.205	0.212
Bank St.	0.480^{+}	0.936^{++}
U. of Oregon	0.090^{--}	0.102
U. of Kansas	0.267	0.081^{-}
Hi/Scope	0.187	0.307
U. of Florida	0.140^{--}	0.208
EDC	0.151	0.278
Pittsburgh	0.683^{++}	0.968^{++}
REC	0.342	0.429
Enablers	0.243	0.203
Grand Mean	0.266	0.340

Fall to Spring change over all models $^{+}$.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); + indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

higher than other models. There are some site effects in the Bank Street model, however: Elmira is quite low in the fall, but that score rises in the spring -- indeed enough to give Bank Street a significant positive change from fall to spring.

Var 29: adult communication focus: one child

The table indicates that Bank Street is low, which is puzzling in view of its high score on variable 12. Either this discrepancy is anomalous or Bank Street adults often interact with two children, but rarely with one child.

The Kansas model is quite high, with small differences among its sites. This high score is probably related to the system of rewarding children's behavior with tokens. An adult gives a child a token and praises him verbally, and thus "communicates" with one child.

The Hi/Scope model is significantly high only in the fall. In this model, the Fort Walton Beach site dropped from a mean of 13.012 in the fall to 3.400 in the spring. thus giving that model a significant negative change from fall to spring. It is interesting that, over all models, adults communicated with one child significantly more in the fall than in the spring. Perhaps this finding is related to a significant positive change in child independent activity. Perhaps children need less "rewarding" or "reminding"

Mean Scores by Model on Classroom Observation

Var. 29. Adult Communication Focus: One Child

	Fall	Spring
Far West	10.408	9.590
U. of Ariz.	11.963	9.841
Bank St.	7.760--	8.255
U. of Oregon	12.755	10.519
U. of Kansas	15.777+	15.763++
Hi/Scope	16.924++	10.797
U. of Florida	8.941	7.321
EDC	9.695	10.112
Pittsburgh	16.687	12.778
REC	13.726	14.224
Enablers	13.359	10.695
Grand Mean	12.560	10.798

Fall to Spring change over all models --.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at.001 level.

in the spring.

Var 30: adult communication focus: small group

The Oregon model shows very high scores on this variable, scores which are probably related to their emphasis on small-group academic instruction. Because Oregon has low scores on variable 29 (adult communication focus: one child) and on variable 10 (independent child activity) we conclude that children in this model spend a great deal of time with adults in small groups.

Bank Street is low on this variable, possibly as a result of its stress on independent activity and on child interactions.

B. Adult attention (responses) to children

Var 22: adult acknowledgement to child
33: adult positive corrective feedback
46: adult praise of children

These variables represent three different ways of responding to a child's behavior. Acknowledgement is commenting in a more or less neutral way: "You're drawing a tree, Jimmy." Acknowledgement can also be non-verbal, such as smiling or nodding at a child. Positive corrective feedback is a method of maintaining discipline or correcting mistakes by suggesting an alternative, by giving a reason why behavior is inappropriate, by praising another child's more appropriate behavior, or by questioning the behavior. Examples: "Why don't you use these blocks instead of the ones that Rachel is playing with?" "If you tear the book, we won't be able to read it again." "Look at the way Henry is cleaning his paintbrushes."

Mean Scores by Model on Classroom Observation

Var. 30. Adult Communication Focus: Small Group

	Fall	Spring
Far West	2.339	3.202
U. of Ariz.	2.016	2.252
Bank St.	1.233	1.732
U. of Oregon	12.840++	17.131++
U. of Kansas	3.569	1.890
Hi/Scope	2.088	2.326
U. of Florida	2.706	2.554
EDC	3.890	3.421
Pittsburgh	3.377	2.835
REC	1.941	2.234
Enablers	2.275	3.424
Grand Mean	3.277	3.785

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). = indicates significantly low at .01 level; -- indicates significantly low at .001 level.

TABLE 35

Third Summary Table

A.	Adult Interactions with one or two children		Adult Communication focus: one child		Adult Communication focus: small group	
	Fall	Spring	Fall	Spring	Fall	Spring
Far West						
U. of Arizona						
Bank St.	+	++	--		-	
U. of Oregon	--				++	++
U. of Kansas		-	+	++		
Hi/Scope			++			
U. of Florida	--					
EDC						
Pittsburgh	++	++				
REC						
Enablers						

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

"Are you sure you want to stand on the table?" Praise is, of course, complimenting a child's behavior.

All three variables changed negatively from fall to spring, although the changes on variables 22 and 23 were not significant. We again suggest our hypothesis that as children and adults grow more comfortable with one another, there is less reason for adult behavior which rewards or alters children's behavior.

Var 22: adult acknowledgement to child

On variable 22, REC is consistently high, with Bank Street low. Hi/Scope and Enablers are significantly high only in the fall and Oregon high only in the spring. This variable is somewhat mysterious: although we understand acknowledgement to denote neutral attention, we wonder whether calling the roll would constitute an acknowledgement of each child. Since we have no information as to the content of this type of communication, we take the above scores at face value, and use this variable in conjunction with the next two, in order to find some characteristic differences among models on the type of attention adults give to children.

Var 33: adult positive corrective feedback

From the variable 33 table, we see that no score is significantly higher than the grand mean in the fall, but that Bank Street, Oregon, and Florida are significantly low. In the

Mean Scores by Model on Classroom Observation

Var. 22 Adult Acknowledgement to Child

	Fall	Spring
Far West	0.684	0.628
U. of Ariz.	1.097	0.984
Bank St.	0.241--	0.452--
U. of Oregon	0.592	2.076+
U. of Kansas	1.355	1.007
Hi/Scope	2.581++	0.993
U. of Florida	0.357--	0.648
EDC	0.694	0.876
Pittsburgh	1.822	1.985
REC	2.515+	3.207++
Enablers	1.983+	1.825
Grand Mean	1.263	1.179

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

Var. 33. Adult Positive Corrective Feedback

	Fall	Spring
Far West	1.099	0.282
U. of Ariz.	1.510	1.216
Bank St.	0.193--	0.415
U. of Oregon	0.420--	0.195
U. of Kansas	1.348	1.507++
Hi/Scope	1.276	0.286
U. of Florida	0.607-	0.855
EDC	0.610	0.513
Pittsburgh	1.577	0.510
REC	1.159	1.272
Enablers	0.881	1.336+
Grand Mean	0.947	0.759

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

spring, Kansas and Enablers are high, but no score is significantly low. There is a great deal of fall to spring change on this variable, with Far West, Hi/Scope, and Pittsburgh showing substantially less corrective feedback in the spring. We surmise that the changes in these models indicate more relaxed adult behavior in the spring.

The Kansas and REC models, on the other hand, show slight increases in corrective feedback in the spring, and thus, these two models are distinctive in the consistently high frequency of adult corrective feedback in their classrooms.

Var 46: adult praise of children

Here, Oregon, Kansas, and Pittsburgh are all consistently high. Oregon's mean is lower in the spring, and compared with a higher spring score on acknowledgement, we see that whereas Oregon teachers expressed more praise than acknowledgement in the fall, they expressed approximately equal amounts of praise and acknowledgement in the spring. Thus, the Oregon model also follows the pattern of more relaxed adult behavior in the spring. Kansas is high on both corrective feedback and praise, and just as we related their high score on adult communication focus: one child to their system of awarding tokens for good behavior, we also relate their scores on feedback and praise to this system.

Pittsburgh is high on feedback in the fall, and on praise in both fall and spring. Pittsburgh, therefore, is outstanding in amount of attention given to one or two children at a time, since they are high on two "attention" variables, significantly

Mean Scores by Model on Classroom Observation

Var. 46. Adult Praise of Children

	Fall	Spring
Far West	0.690^{--}	0.536^{-}
U. of Ariz.	0.529^{--}	0.552^{-}
Bank St.	0.298^{--}	0.310^{--}
U. of Oregon	3.614^{++}	2.137^{+}
U. of Kansas	3.141^{++}	2.615^{++}
Hi/Scope	1.819	1.014
U. of Florida	1.052^{-}	0.506^{-}
EDC	0.876^{-}	0.999
Pittsburgh	5.080^{++}	3.884^{++}
REC	0.634^{-}	0.595
Enablers	0.676	1.355
Grand Mean	1.540	1.249

Fall to Spring change over all models $^{--}$.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

TABLE 35

Third Summary Table

B.	Adult Acknowledgment to child		Adult Positive Corrective Feedback		Adult Praise of Children	
	Fall	Spring	Fall	Spring	Fall	Spring
Far West					--	-
U. of Arizona					--	-
Bank St.	--	--	--		--	--
U. of Oregon		+	--		++	+
U. of Kansas				++	++	++
Hi/Scope	++					
U. of Florida	--		-		-	-
EDC					-	
Pittsburgh					++	++
REC	+	++			-	
Enablers	+			+		

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ inidcates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

high on both adult interactions with one or two children, and above the grand mean on adult communication focus: one child.

C. Questioning

Var 15: adult direct questioning of child
16: child response to adult direct question
18: adult feedback to child response
20: adult asking thought questions
27: child asking questions

We will deal with the first three variables in this group together, as they are a "process" which is important in the Oregon, Kansas, IPI and REC models. These models have "prescribed learning" activity as part of their model specifications. In other words, these models attempt to teach children specific skills in a structured way. Thus, variables 15, 16, and 18 represent a sort of "drill" in which the teacher asks, for example, "What color is this apple, Susie?"; Susie answers, "red"; and the teacher responds, "That's right."

The tables on variables 15, 16, and 18 show, as we might have predicted, Oregon and Kansas high on all three variables. There are some site effects, however, within these models on the first two variables: Tupelo in the Oregon model and Portageville in the Kansas model are far higher than other sites in any model although the other sites in the Oregon and Kansas models are not at all low.

All three of these variables show a negative change from fall to spring and we assume that this finding supports our "rule" that children need less rewarding, reminding, or

Mean Scores by Model on Classroom Observation

Var. 15. Adult Direct Questioning of Child

	Fall	Spring
Far West	2.782^{-}	2.664^{--}
U. of Ariz.	4.808	4.454
Bank St.	3.248^{-}	3.435^{-}
U. of Oregon	13.359^{++}	10.894^{++}
U. of Kansas	8.415+	7.606^{+}
Hi/Scope	7.094	4.349
U. of Florida	5.208	4.266
EDC	5.969	6.104
Pittsburgh	8.091	4.569
REC	8.674	7.611
Enablers	3.523^{-}	6.647
Grand Mean	6.125	5.673

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

Var. 16. Child Response to Adult Direct Question

	Fall	Spring
Far West	2.092^{-}	2.208
U. of Ariz.	3.097	3.457^{-}
Bank St.	2.507^{-}	2.652^{-}
U. of Oregon	12.360^{++}	9.712^{++}
U. of Kansas	7.308^{+}	6.255^{+}
Hi/Scope	5.842	3.191
U. of Florida	4.295	3.582
EDC	2.777	4.019
Pittsburgh	6.488	3.385
REC	6.802	6.201
Enablers	2.097^{--}	4.829
Grand Mean	4.774	4.436

Fall to Spring change over all models $^{-}$.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

drilling as they become more familiar with their classroom environment.

Var 20: adult asking thought questions

Variable 20 shows a great deal of fall to spring change in several models. Far West, Oregon, Hi/Scope and Enablers all have lower scores in the spring. Only Pittsburgh and REC show more thought questioning in the spring. However, Arizona, Hi/Scope, EDC, and Pittsburgh remain above the overall mean on both observations, with Bank Stteet and Kansas significantly below the overall mean. Thus, although model differences on this variable are confused by fall to spring change, it is clear that there are consistently more thought questions being asked of children in the Arizona, Hi/Scope, EDC, and Pittsburgh models than in the Bank Street and Kansas models.

Var 27: child asking questions

We include this variable mainly to show that the Oregon model, which is outstandingly high on adult communication focus: small group, and has high scores on the question-response-feedback process, is significantly low on child asking questions. This finding assumes more importance in connection with variable 28 (child self-expression), which is discussed in the next section.

Mean Scores by Model on Classroom Observation

Var. 20. Adult Asking Thought Questions

	Fall	Spring
Far West	1.538	0.697
U. of Ariz.	2.096	1.764
Bank St.	0.646^{-}	0.452^{-}
U. of Oregon	3.268^{++}	0.133^{--}
U. of Kansas	0.405^{--}	0.397^{-}
Hi/Scope	2.689^{+}	1.671
U. of Florida	1.568	0.983
EDC	1.753	1.480
Pittsburgh	1.723	2.462^{++}
REC	0.786	1.670
Enablers	2.442	0.344^{-}
Grand Mean	1.726	0.985

Fall to Spring change over all models $^{--}$.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

Var. 27. Child Asking Questions

	Fall	Spring
Far West	4.183+	1.650
U. of Ariz.	2.727	1.706
Bank St.	3.378	1.851
U. of Oregon	0.775--	0.946-
U. of Kansas	2.302	1.789
Hi/Scope	3.224	1.546
U. of Florida	1.333	1.954
EDC	1.279-	1.668
Pittsburgh	1.628	2.252
REC	2.620	2.632
Enablers	3.553	2.418
Grand Mean	2.620	1.800

Fall to Spring change over all models --.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

TABLE 35

Third Summary Table

C.	Adult Dir. Questning. of Child		Child Resp. to Adult Dir. Quest.		Adult Feedback to Child Response		Adult Asking Thought Questions		Child Asking Questions	
	Fall	Spr.	Fall	Spr.	Fall	Spr.	Fall	Spr.	Fall	Spr.
Far West	-	--	-		--	--			+	
U. of Arizona				-	--	-				
Bank St.	-	-	-	-	--	--	-	-		
U. of Oregon	++	++	++	++	++	++	++	--	--	-
U. of Kansas	+	+	+	+	++	++	--	-		
Hi/Scope		-			+	-	+			
U. of Florida		-			-					
EDC					--	-			-	
Pittsburgh					++			++		
REC					+					
Enablers	-		--		-			-		

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicatss significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Also, it is interesting that the Enabler model is consistently above the grand mean on this variable with small differences among its sites. We would not expect the Enabler sites to be similar to each other, since there are no model specifications, but all three sites show high scores on child asking questions.

D. Child behavior (response)

Var 39: child positive affect toward adults
28: child self-expression

Var 39: child positive affect toward adults

Variable 39 is the only one of seven variables concerned with affect that seems to indicate any genuine differences among models, and even this variable shows significant site effects. The Hi/Scope model is consistently higher than other models, but their high t-value results from outstandingly high scores for the Fort Walton Beach site. Pittsburgh is higher than the grand mean in the fall and is higher than any other model in the spring.

It is puzzling that scores on child positive affect decreased over all models in the spring, especially since scores on variable 28 (child self-expression) show a significant increase.

Var 28: child self-expression

Variable 28 is a good illustration of the difficulty

Mean Scores by Model on Classroom Observation

Var. 39. Child Positive Affect Toward Adults

	Fall	Spring
Far West	0.982	0.288
U. of Ariz.	0.129	0.129
Bank St.	1.060	0.506
U. of Oregon	0.037	0.364
U. of Kansas	0.273	0.149
Hi/Scope	2.418++	1.185++
U. of Florida	0.352	0.086
EDC	0.181	0.694
Pittsburgh	0.909	1.272
REC	0.226	0.148
Enablers	0.234	0.360
Grand Mean	0.692	0.478

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). --indicates significantly low at .01 level; -- indicates significantly low at .001 level.

Mean Scores by Model on Classroom Observation

	Fall	Spring
Far West	16.885+	17.323
U. of Ariz.	15.012	17.857
Bank St.	14.268	17.040
U. of Oregon	4.501--	7.332-
U. of Kansas	11.304	10.895
Hi/Scope	8.103	12.497
U. of Florida	16.739+	21.239+
EDC	6.533-	13.271
Pittsburgh	11.361	14.802
REC	13.260	11.941
Enablers	14.356	15.508
Grand Mean	12.008	14.368

Fall to Spring change over all models ++.

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

involved in finding actual differences among models on some of our variables. Child self-expression scores increased from fall to spring significantly more than others. However, we see from the table that only the Oregon model is consistently low and that only the Florida model is consistently high.

The Oregon model was also low on the child asking questions variable and thus, children in the Oregon model apparently initiate interactions less frequently than children in other Planned Variation models.

There are large site effects on this variable within models other than the Oregon model, and so we can make no further assumptions about differences among models on child self-expression.

MODEL SUMMARIES

In this section we will summarize the findings on our twenty-two variables by model. Also, we will attempt to relate these findings to model implementation. In other words, we will suggest which differences among models can be related to model theory. For example, it is obvious that the high scores on academic activity for the Oregon and Kansas models are a result of sponsor emphasis on teaching children specific academic skills. On the other hand, it is difficult to understand why, for example, REC is high on adult acknowledgement and low on adult praise of children,

TABLE 35

Third Summary Table

D.	Child Self-Expression		Child Positive Affect Toward Adults	
	Fall	Spring	Fall	Spring
Far West	+			
U. of Arizona				
Bank Street				
U. of Oregon	--	-		
U. of Kansas				
Hi/Scope			++	++
U. of Florida	+	+		
EDC	-			
Pittsburgh				
REC				
Enablers				

Note: In the above table + indicates significantly high at .01 level (t value = 2.6 or over); ++ indicates significantly high at .001 level (t value = 3.4 or over). - indicates significantly low at .01 level; -- indicates significantly low at .001 level.

since we find no model specification which states that adults should acknowledge children rather than praise them. Thus, we will attempt to relate model theory to scores on variables when some relationship seems plausible.

There are several problems with drawing conclusions about model implementation on the basis of the classroom observation data. These data are derived from an instrument which requires rapid coding of readily observable behavior. While some models (Oregon, Kansas, Pittsburgh and REC) specify in fairly explicit terms what activities should occur in classrooms and how adults should interact with children, others emphasize "experiences" for children. Our data cannot measure whether teachers are providing "self-rewarding" experiences for children (a major aspect of the Far West model), nor whether adults are "expanding children's language" (a part of the Arizona model). Thus, models vary in the degree to which implementation can be checked by classroom observation data.

Further, scores on our variables reflect frequency of behavior. We have no way of knowing, for example, whether Bank Street's score on independent child activity (which is significantly higher than other models) is high enough to rate that model as highly implemented. Planned Variation sponsors do not express their theory in terms of contrast with other sponsors; therefore, even though there may be a great deal of independent activity (as compared with other

models) in Bank Street, we cannot draw a firm conclusion that Bank Street is well-implemented in that area.

With these limitations in mind, we chose a few variables for each model which might be theoretically important. A table of means and standard deviations on these theoretically important variables is included in each model's[6] summary. Site scores are shown, since site variation within models is often larger than variation among models. We would expect a model to be high on the variables included in its table, except where "low" (in parentheses) appears after the variable, in which case we would expect the model to be low.

A quick perusal of the tables of theoretically important variables serves to emphasize the limitations of classroom observation data in judging whether models are well implemented. The four models (Oregon, Kansas, IPI and REC) which specify certain types of overt behavior and supply academic curricula to their teachers appear to be better implemented than the other models. Although these models may indeed be well implemented, the fact that our variables are more relevant to their model specifications than to the other models, makes us wonder whether the other models would appear to be well-implemented if we used a different type of observation instrument.

---

[6] Enablers and Florida have no tables of theoretically important variables because they have no model specifications for classroom behavior.

Thus, our data do determine which models are higher than others on certain kinds of readily observable behavior. Using this data to make assumptions about one model's degree of implementation would be improper, since model theories differ from one another too much to make any one observation instrument relevant to all models.

FAR WEST

The Far West model is significantly higher than the grand mean in both the fall and spring on only one of the twenty-two variables included in the previous discussion. That one is variable 23 (child self-learning with concrete objects). Although child self-learning is theoretically important in the Far West model, it is only the Duluth site which is significantly high. Significantly low scores appear on praise and on the adult direct question-child response-adult feedback complex. Thus, the Far West model is not particularly distinctive on most of our measures of classroom behavior.

The table of theoretically important variables shows that the two sites in the Far West model are substantially different from each other. For example, Tacoma has high scores on adult communication focus: one child, while

TABLE 36

Far West

		Duluth		Tacoma		Far West		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 4	N = 4	N = 8	N = 8	N = 93	N = 97
Var 2. Activity B: group time (low)	mean	0.175	0.296	0.184	0.196	0.180	0.246	0.249	0.241
	s.d.	0.126	0.110	0.061	0.038	0.099	0.096	0.118	0.146
Var 10. Independent child activity	mean	2.706	1.774	0.605	1.149	1.656	1.461	1.141	1.296
	s.d.	0.565	0.397	0.143	0.489	1.128	0.544	1.060	0.979
Var 22. Adult acknowledgement to child	mean	0.802	0.620	0.566	0.635	0.684	0.628	1.263	1.179
	s.d.	0.121	0.353	0.151	0.335	0.181	0.344	1.099	1.012
Var 27. Child asking questions	mean	1.745	1.569	6.621	1.731	4.183	1.650	2.620	1.800
	s.d.	0.582	0.169	1.397	0.690	2.663	0.509	1.615	0.966
Var 29. Adult communication focus: one child	mean	5.885	7.435	14.930	11.746	10.408	9.590	12.560	10.798
	s.d.	1.232	0.870	1.860	0.925	4.790	2.335	4.966	4.768
Var 30. Adult communication focus: small group	mean	1.527	3.119	3.151	3.285	2.339	3.202	3.277	3.785
	s.d.	1.084	1.731	0.989	0.600	1.318	1.298	3.592	4.968
Var 31. Adult communication focus: large group (low)	mean	3.492	5.220	2.927	3.362	3.209	4.291	4.571	3.977
	s.d.	2.485	3.957	1.298	1.519	2.003	3.092	4.099	3.028
Var 50. Child informing self other than symbolically or with concrete objects	mean	0.0	0.0	0.430	0.086	0.215	0.043	0.406	0.119
	s.d.	0.0	0.0	0.213	0.080	0.263	0.071	1.117	0.356

Duluth has high scores on independent child activity.

Fall to spring changes are puzzling; Far West's spring scores decrease significantly more than other models on twelve of the total fifty-one variables; spring scores do not increase significantly more than other models on any variable.

UNIVERSITY OF ARIZONA

The Arizona model is significantly lower than the grand mean on adult praise of children, and on adult feedback to child response. We see from the table that Arizona's scores on adult asking thought questions are consistently higher than the overall mean. Also, we can see that the two sites do not much differ from each other, except perhaps on adult positive corrective feedback where Lafayette is high. This model seems to typify the "average" P.V. model, at least insofar as our classroom observation variables measure classroom activity.

BANK STREET

The Bank Street model differs from other models mainly on the amount of independent child activity. From the following table, we see that although the Wilmington site is much higher than Boulder or Elmira, all three sites have scores near or above the grand mean. Related to their high

TABLE 37

University of Arizona

		Lafayette		Lincoln		U. of Arizona		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 4	N = 4	N = 8	N = 8	N = 93	N = 97
Var 11. Wide variety of activities	mean	1.891	1.765	2.083	2.064	1.987	1.915	2.112	2.154
	s.d.	0.169	0.287	0.255	0.204	0.237	0.290	1.143	0.964
Var 20. Adult asking thought questions	mean	2.062	0.999	2.131	2.529	2.096	1.764	1.726	0.985
	s.d.	0.910	0.429	0.706	1.493	0.815	1.338	1.425	1.001
Var 27. Child asking questions	mean	2.347	1.827	3.107	1.585	2.727	1.706	2.620	1.800
	s.d.	0.292	0.235	0.699	0.331	0.657	0.311	1.615	0.966
Var 29. Adult communication focus: one child	mean	10.639	9.98	13.286	9.701	11.963	9.841	12.560	10.798
	s.d.	2.286	2.168	1.654	2.182	2.394	2.179	4.966	4.768
Var 30. Adult communication focus: small group	mean	1.154	2.252	2.857	2.252	2.016	2.252	3.277	3.785
	s.d.	0.741	1.432	1.969	0.172	1.709	1.020	3.592	4.968
Var 32. Adult praise/acknowledgement	mean	1.980	1.911	1.274	1.160	1.627	1.536	2.803	2.428
	s.d.	0.822	0.541	0.203	0.883	0.695	0.823	1.898	1.746
Var 33. Adult positive corrective feedback	mean	2.211	2.019	0.810	0.414	1.510	1.216	0.947	0.759
	s.d.	0.401	0.532	0.233	0.243	0.774	0.903	0.780	0.848

scores on independent activity are low scores on adult acknowledgement, praise, and feedback. Strangely, the Wilmington site has low scores on child self-expression, compared with Boulder and Elmira. Thus, child self-expression seems to be negatively correlated with independent activity in the Bank Street model.

We decided that Elmira's fantastic score in the spring on child positive affect to other children must be anomalous, perhaps due to observer effects or unusual circumstances in the classroom.

UNIVERSITY OF OREGON

The Oregon model has high scores on all variables that are related to academic activity. Although there is some site variation, both sites are above the grand mean on all academic activity variables, with the one exception that Tupelo is low on aide's participation in academic activities. Oregon is high on adult communication focus: small group, on adult praise, and on the adult direct question-child response-adult feedback sequence.

Low scores appear on child asking questions and child self-expression. Since the Oregon model specifies academic instruction in small groups, it is clear that our variables are measuring important aspects of this model.

TABLE 38

Bank Street

		Boulder		Wilmington		Elmira		Bank Street		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 4	N = 4	N = 3	N = 3	N = 11	N = 11	N = 93	N = 97
Var 4. Activity D: Inquiry activities	mean	0.067	0.130	0.029	0.034	0.067	0.0	0.053	0.060	0.076	0.052
	s.d.	0.042	0.099	0.032	0.059	0.047	0.0	0.044	0.085	0.081	0.068
Var 10. Independent child activity	mean	1.069	1.228	5.155	3.733	1.432	2.368	2.654	2.450	1.141	1.296
	s.d.	0.340	0.119	0.302	0.546	0.380	0.877	1.926	1.211	1.060	0.979
Var 11. Wide variety of activities	mean	2.102	2.851	6.902	5.456	2.059	2.806	3.836	3.786	2.112	2.154
	s.d.	0.424	0.426	0.189	0.700	0.390	0.454	2.344	1.377	1.143	0.964
Var 28. Child self-expression	mean	14.041	15.848	6.894	13.896	24.403	22.822	14.268	17.040	12.008	14.368
	s.d.	0.772	2.788	2.623	9.932	3.048	9.551	7.284	8.764	6.357	7.000
Var 32. Adult praise/acknowledgement	mean	0.768	1.494	0.0	0.011	0.951	0.787	0.539	0.762	2.803	2.428
	s.d.	0.327	0.263	0.0	0.020	0.142	0.729	0.464	0.755	1.898	1.746
Var 38. Adult positive affect to child	mean	3.395	0.536	0.0	0.0	0.0	2.270	1.235	0.814	0.647	0.513
	s.d.	1.054	0.600	0.0	0.0	0.0	1.356	1.752	1.216	2.037	0.995
Var 49. Child positive affect to other children	mean	3.251	0.506	0.018	0.0	0.576	26.893	1.346	7.519	1.209	3.005
	s.d.	2.478	0.274	0.031	0.0	0.102	5.735	2.088	12.240	1.896	6.965

TABLE 39

University of Oregon

		Tupelo		E. Las Vegas		U. of Oregon		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 3	N = 4	N = 7	N = 8	N = 93	N = 97
Var 8. Adults with children in academic activities	mean	0.784	1.002	1.138	1.250	0.936	1.126	0.476	0.483
	s.d.	0.073	0.039	0.215	0.165	0.231	0.173	0.446	0.432
Var 9. Frequency of academic activities	mean	0.784	1.054	1.214	1.334	0.969	1.194	0.634	0.665
	s.d.	0.073	0.088	0.244	0.163	0.272	0.192	0.540	0.500
Var 13. Aide's participation in academic activities	mean	0.075	0.016	0.464	0.741	0.242	0.378	0.201	0.204
	s.d.	0.091	0.027	0.091	0.137	0.213	0.376	0.284	0.306
Var 15. Adult direct questioning of child	mean	17.380	14.325	7.997	7.463	13.359	10.894	6.125	5.673
	s.d.	2.427	1.230	1.022	0.801	5.038	3.585	3.984	3.248
Var 16. Child response to adult direct question	mean	16.617	12.802	6.683	6.623	12.360	9.712	4.774	4.436
	s.d.	2.357	1.344	0.631	0.769	5.246	3.278	3.753	2.921
Var 17. Adult praise and corrective feedback	mean	4.457	8.276	5.557	5.390	4.980	6.833	4.393	3.858
	s.d.	0.631	1.327	1.083	1.083	0.990	1.884	2.522	2.521
Var 18. Adult feedback to child response	mean	3.182	5.650	2.486	3.045	2.883	4.348	1.305	1.270
	s.d.	0.249	1.491	0.408	0.542	0.472	1.719	1.268	1.482
Var 30. Adult communication focus: small group	mean	16.115	24.920	8.473	9.341	12.840	17.131	3.277	3.785
	s.d.	0.894	3.088	1.963	1.104	4.051	8.128	3.592	4.968

University of Oregon (cont.)

		Tupelo		E. Las Vegas		U. of Oregon		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 3	N = 4	N = 7	N = 8	N = 93	N = 97
Var 32. Adult praise/ acknowledgement of children	mean	4.489	5.440	3.829	2.987	4.206	4.213	2.803	2.428
	s.d.	0.646	0.660	0.797	0.796	0.786	1.428	1.898	1.746
Var 45. Adult informing children other than symbolically or with concrete objects	mean	2.665	4.040	2.260	4.444	2.491	4.242	2.962	2.357
	s.d.	1.037	1.225	0.652	1.298	0.915	1.278	2.780	1.953

UNIVERSITY OF KANSAS

The Kansas model, like the Oregon model, is high on variables related to academic activity. In fact the Kansas model differs from Oregon in only three aspects: Kansas has more aides and fewer teachers participating in academic activities; the adult communication focus is one child rather than small group; and the Kansas adults use much more positive corrective feedback.

Kansas is also low on independent child activity. Obviously, our variables measure the "behavior modification" process in some detail.

It is puzzling that the Portageville site is so much higher than Oraibi or Mounds on the adult direct question-child response-adult feedback sequence as Portageville is not much higher on academic activity. Perhaps Portageville teachers employ the direct question sequence in areas other than academic activity. Except for this variation, the three Kansas sites are quite similar to one another.

HI/SCOPE

The Hi/Scope model is significantly high in both fall and spring on only one of our twenty-two variables: child positive affect toward adults. On adults asking thought questions, Hi/Scope is significantly high in the fall

Table 40

		Oraibi		Portageville		Mounds		U. of Kansas		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 4	N = 4	N = 4	N = 4	N = 12	N = 12	N = 93	N = 97
Var 8. Adults with children in academic activities	mean	1.390	1.308	1.375	1.412	0.853	0.652	1.206	1.124	0.476	0.483
	s.d.	0.527	0.357	0.190	0.220	0.123	0.233	0.415	0.436	0.446	0.432
Var 9. Frequency of academic activities	mean	1.412	1.328	1.375	1.431	0.873	0.734	1.220	1.164	0.634	0.665
	s.d.	0.537	0.341	0.190	0.189	0.144	0.204	0.419	0.399	0.540	0.500
Var 13. Aide's participation in academic activities	mean	0.916	0.892	0.925	1.031	0.532	0.403	0.791	0.776	0.201	0.204
	s.d.	0.350	0.402	0.109	0.200	0.088	0.173	0.284	0.387	0.284	0.306
Var 15. Adult direct questioning of child	mean	5.729	5.367	13.425	11.987	6.090	5.465	8.415	7.606	6.125	5.678
	s.d.	0.099	1.099	1.578	0.965	0.622	0.248	3.679	3.214	3.984	3.248
Var 16. Child response to adult direct question	mean	4.232	2.912	12.531	11.400	5.162	4.454	7.308	6.255	4.774	4.436
	s.d.	0.352	0.494	1.445	1.134	0.485	0.462	3.821	3.770	3.753	2.921
Var 17. Adult praise and corrective feedback	mean	8.269	6.889	6.700	6.612	4.226	4.103	6.398	5.868	4.393	3.858
	s.d.	1.934	1.149	0.694	0.646	0.692	0.356	2.082	1.481	2.522	2.521
Var 18. Adult feedback to child response	mean	2.303	1.345	4.300	4.462	1.736	1.375	2.780	2.394	1.305	1.270
	s.d.	0.297	0.396	0.628	0.862	0.322	0.384	1.185	1.577	1.268	1.482
Var 29. Adult communication focus: one child	mean	14.294	17.540	19.881	18.350	13.154	11.398	15.777	15.763	12.560	10.798
	s.d.	1.163	1.951	2.838	1.275	1.268	1.195	3.509	3.453	4.966	4.768
Var 30. Adult communication focus: small group	mean	6.439	2.231	2.156	1.819	2.113	1.620	3.569	1.890	3.277	3.785
	s.d.	2.917	0.500	0.846	0.603	0.543	0.650	2.700	0.641	3.592	4.968

University of Kansas (cont.)

		Oraibi		Portageville		Mounds		U. of Kansas		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 4	N = 4	N = 4	N = 4	N = 12	N = 12	N = 93	N = 97*
Var 32. Adult praise/ acknowledgement of children	mean	5.728	4.270	5.069	4.700	2.693	1.896	4.496	3.622	2.803	2.428
	s.d.	0.844	0.853	0.520	0.706	0.592	0.627	1.464	1.435	1.898	1.746
Var 45. Adult informing children other than symbolically or with concrete objects	mean	5.023	3.810	1.087	0.462	2.026	1.078	2.712	1.783	2.962	2.357
	s.d.	2.239	0.414	0.340	0.383	0.792	0.471	2.176	1.515	2.780	1.953

and above the grand mean in the spring. Hi/Scope is significantly low on several variables relating to academic activities.

From the Hi/Scope table, we see a very mysterious pattern. The Fort Walton Beach site is quite different from Greeley and Seattle in the fall on several variables but Fort Walton Beach changes radically from fall to spring. Since we have both site variation and fall to spring change in this model, it is difficult to discover any consistent pattern among its scores. Perhaps the Fort Walton Beach spring observation occurred under unusual circumstances.

UNIVERSITY OF FLORIDA

Because the Florida model does not specify any particular classroom activity, but rather stresses training parents to teach their children by means of home visits, we have no table of theoretically important variables. As we might have predicted, site variation is large. The Florida model is, however, significantly high on child self-expression. Perhaps the lack of sponsor pressure on adults in the classroom results in children feeling more free to express themselves.

EDC

The EDC model, like the Arizona model, seems to be around the grand mean on almost all of our variables. Site

TABLE 41

Hi-Scope

		Ft. Walton Bch		Greeley		Seattle		Hi-Scope		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring	fall	spring
		N = 4	N = 4	N = 4	N = 4	N = 4	N = 4	N = 12	N = 12	N = 93	N = 97
Var 10. Independent child activity	mean	0.402	0.654	0.988	1.890	0.771	0.450	0.720	0.998	1.141	1.296
	s.d.	0.275	0.201	0.200	0.794	0.098	0.166	0.316	0.799	1.060	0.979
Var 14. Adult informing children symbolically	mean	0.615	0.0	0.0	0.0	0.0	0.063	0.205	0.021	0.104	0.377
	s.d.	0.333	0.0	0.0	0.0	0.0	0.108	0.348	0.069	0.359	0.747
Var 15. Adult direct questioning of child	mean	6.043	1.627	7.412	5.181	7.826	6.237	7.094	4.349	6.125	5.673
	s.d.	2.478	0.813	0.698	1.787	1.950	2.269	2.015	2.625	3.984	3.248
Var 20. Adult asking thought questions	mean	5.581	2.149	1.281	2.270	1.204	0.593	2.689	1.671	1.726	0.985
	s.d.	0.908	0.383	0.502	0.317	0.568	0.342	2.156	0.839	1.425	1.001
Var 21. Adult informing child with concrete objects	mean	4.439	0.0	0.0	0.0	0.0	0.0	1.480	0.0	0.477	0.189
	s.d.	1.093	0.0	0.0	0.0	0.0	0.0	2.186	0.0	1.088	0.690
Var 23. Child self-learning with concrete objects	mean	0.250	0.0	0.0	0.0	0.083	0.157	0.111	0.052	0.578	0.910
	s.d.	0.081	0.0	0.0	0.0	0.144	0.125	0.141	0.103	1.353	1.502
Var 26. Child self-learning with symbols	mean	0.037	0.0	0.0	0.0	0.0	0.0	0.012	0.0	0.051	0.085
	s.d.	0.065	0.0	0.0	0.0	0.0	0.0	0.041	0.0	0.198	0.264
Var 28. Child self-expression	mean	2.064	12.528	11.455	16.418	10.790	8.546	8.103	12.497	12.008	14.368
	s.d.	0.856	1.669	1.697	3.568	1.972	2.086	4.562	4.117	6.357	7.000

variation is large, as we might have predicted since EDC emphasizes adaptation of the model to the particular needs of the site. EDC is one of the Planned Variation models which stress "attitudes" and "providing for children assuming responsibility" rather than specific types of activities. Children should be absorbed in their classroom activity, and thus adults should provide "absorbing" materials and activities. Our variables cannot measure what is important in the EDC model; our variables can only measure immediately recognizable kinds of interactions and activity, not whether those activities and interactions indicate that children are involved or whether they are being creative.

UNIVERSITY OF PITTSBURGH

The Pittsburgh model's spring scores indicate interesting classroom behavior; the model is high on independent children in academic activities, adult interactions with one or two children, and adult acknowledgement. Pittsburgh is also high on adults asking thought questions. Thus, Pittsburgh teachers do not supervise academic activity, but rather acknowledge children's work and ask thought questions of one or two children individually.

Pittsburgh's spring scores seem to be consistent with model theory: the model emphasizes independent learning and teacher reinforcement of child learning. Of course,

Table 42

		Washington		Paterson		Johnson Co.		EDC		Overall	
		fall	spring	fall	spring	fall	spring	fall	spring	fall	spring
		N = 1	N = 4	N = 4	N = 4	N = 3	N = 4	N = 8	N = 12	N = 93	N = 97
Var 10. Independent child activity	mean	1.455	1.482	0.501	0.590	1.094	1.343	0.843	1.138	1.141	1.296
	s.d.	0.001	1.046	0.063	0.307	0.314	0.406	0.410	0.778	1.060	0.979
Var 11. Wide variety of activities	mean	2.409	1.494	1.002	1.354	2.253	2.426	1.647	1.758	2.112	2.154
	s.d.	0.001	0.564	0.031	0.251	0.312	0.464	0.675	0.652	1.143	0.964
Var 22. Adult acknowledgement to child	mean	1.591	1.616	0.688	0.515	0.404	0.497	0.694	0.876	1.263	1.179
	s.d.	0.001	0.434	0.107	0.120	0.066	0.701	0.374	0.711	1.099	1.012
Var 28. Child self-expression	mean	14.318	18.245	5.118	5.852	5.824	15.716	6.533	13.271	12.008	14.368
	s.d.	0.002	2.534	1.054	1.073	3.216	4.002	3.633	6.037	6.357	7.000
Var 3. Adult "negative" corrective feedback (low)	mean	0.227	0.691	0.768	0.564	0.0	0.0	0.412	0.419	0.254	0.253
	s.d.	0.000	0.223	0.289	0.198	0.0	0.0	0.416	0.346	0.378	0.326
Var 38. Adult positive affect toward children	mean	0.0	0.023	0.110	0.942	0.198	0.015	0.129	0.327	0.647	0.513
	s.d.	0.0	0.039	0.091	0.507	0.181	0.025	0.143	0.526	2.037	0.995
Var 49. Child positive affect to other children	mean	0.182	0.892	0.872	2.155	0.141	0.104	0.512	1.050	1.209	3.005
	s.d.	0.000	0.320	0.287	0.901	0.109	0.146	0.419	1.012	1.896	6.965

TABLE 43

University of Pittsburgh		Lock Haven Pittsburgh		Overall	
		fall	spring	fall	spring
		N = 4	N = 4	N = 93	N = 97
Var 10. Independent Child Activity	mean	1.244	2.454	1.141	1.296
	s.d.	0.525	1.103	1.060	0.979
Var 11. Wide variety of activities	mean	1.835	2.658	2.112	2.154
	s.d.	0.408	0.630	1.143	0.964
Var 12. Adult interactions with one or two children	mean	0.683	0.968	0.266	0.340
	s.d.	0.311	0.398	0.252	0.421
Var 18. Adult feedback to child response	mean	2.901	1.349	1.305	1.270
	s.d.	0.484	0.516	1.268	1.482
Var 33. Adult positive corrective feedback	mean	1.577	0.510	0.947	0.759
	s.d.	0.471	0.180	0.780	0.848

there is only one site in this model, and so we cannot compare scores across sites. However, there is strong evidence to suggest that in at least this one site, Pittsburgh made great strides toward good model implementation from fall 1970 to spring 1971.

REC

REC is significantly high in both fall and spring on only one variable: adult acknowledgement of children. From REC's theory, we would expect higher scores on independent activity and on adult communication focus: one child. In fact, from the REC table we see that REC scores are above the grand mean, but not high enough to be significant.

Although REC, like Pittsburgh, was new to Planned Variation in 1970-71, it did not change from fall to spring. In fact, REC changed less than any other model. Thus, REC scores are consistently around the grand mean on almost all our variables.

ENABLERS

The Enabler model has no prescribed classroom theory, and thus, we could not select theoretically important variables for this model. Because each site is to be implemented

TABLE 44

Responsive Environments Corporation		Kansas City		Overall	
		fall	spring	fall	spring
		N = 4	N = 4	N = 93	N = 97
Var 2. Activity B: Group time (low)	mean	0.248	0.325	0.249	0.241
	s.d.	0.068	0.087	0.118	0.146
Var 10. Independent Child Activity	mean	1.484	1.532	1.141	1.296
	s.d.	0.420	0.513	1.060	0.979
Var 18. Adult feedback to child response	mean	2.613	2.472	1.305	1.270
	s.d.	0.386	0.818	1.268	1.482
Var 29. Adult communication focus: one child	mean	13.726	14.224	12.560	10.798
	s.d.	1.700	3.123	4.966	4.768
Var 30. Adult communication focus: small group	mean	1.941	2.234	3.277	3.785
	s.d.	0.890	0.839	3.592	4.968
Var 31. Adult communication focus: large group (low)	mean	5.025	5.222	4.571	3.977
	s.d.	4.234	1.154	4.099	3.028
Var 32. Adult praise/ acknowledgement of children	mean	3.149	3.802	2.803	2.428
	s.d.	0.834	0.883	1.898	1.746
Var 33. Adult positive corrective feedback	mean	1.159	1.272	0.947	0.759
	s.d.	0.473	0.582	0.780	0.848

according to the community's needs and desires, we would anticipate more variation among the Enabler sites than in any other model. In fact, there is a great deal of site to site variation in this model, but not more than in the Hi/Scope, Far West, and Arizona models.

Enabler sites, as a group, are significantly high in the fall on adult acknowledgement and on children asking questions. All three sites are low on academic activity variables. The fact that Enabler sites are not supposed to be alike perhaps indicates that the implementation of a model in two or more different sites is a complex and difficult process.

CONCLUSION

It is apparent from the discussion of the classroom observation data that interpretation of model differences is inhibited by several factors. First, the research design is not balanced: some models have three sites, others two, and still others only one. Further, the number of classrooms observed differs from site to site (from one to four P.V. classrooms). Such an unbalanced design limits the statistical analysis of model effects, because site effects as well as classroom effects make model differences difficult to establish.

Further, our variables are limited to measures of overt, even gross, behavior in the classroom. Those models which emphasize behavior modification appear distinctive on our variables, but models which do not emphasize particular methods of instruction do not appear distinctive. An instrument which measured the context and purpose of adult behavior might be more effective in revealing differences among models.

In order to use classroom observation data as evidence of progress toward model implementation, we not only need a more sensitive observation instrument, but also need more than two observations over the year. With only two observations, it is impossible to distinguish whether fall to spring differences are a result of actual change over the year or merely a result of random day to day differences in classroom activities.

Despite the above limitations, two statements are possible:

1. We find strong evidence of differences among models on academic activity, independent child activity, and type of adult attention to children. For example, the Oregon and Kansas classrooms appear very different from Bank Street and Pittsburgh classrooms.
2. However, Far West, Tucson, Hi/Scope, Florida, EDC, REC and Enabler classrooms seem to be rather similar. Perhaps these classrooms actually are similar, but the classroom observation data does not provide enough information about classrooms to establish that there are no important differences among these models.

Chapter 6

CONCLUSIONS

Summary of Findings:

This report explored four issues: the degree to which models are implemented; the variation in factors which may affect implementation; the relation between these factors and levels of implementation; and the variations among models in classroom activities.

We addressed the question of how well the models are implemented through an analysis of the sponsors' ratings of teachers. We find that:

1. All models on the average are moderately well implemented.
2. There are no significant differences among models.
3. There is large variation among classes within sites in levels of implementation as ~~well~~ as significant differences among sites within models.
4. The levels of implementation for second year sites or for second year teachers are not significantly higher than for first year, and not all models show improvement over the year.

From these findings, we conclude that the implementation of models is not as simple as had been originally supposed; we

cannot say that models can be delivered to all classes, nor can we say that the experimental treatments are fully implemented. Moreover, these findings raise two additional questions: why is there variation in level of implementation, and if models are not being fully implemented, which parts are being implemented?

In trying to explain the variation, we explored a variety of factors which might affect implementation. We propose that implementation is an interactive process depending not only on sponsor input, but also on staff input and reaction, and on the operational context of the staff. Because there are problems with the data available for each of these areas, we must be cautious in our interpretations. Some findings, however, are clear:

1. There are significant differences among models and among sites within models in the types and amounts of training received by the teachers.
2. There are significant site and model differences in the training given by people other than the sponsor.
3. There are significant differences among sites and models in the characteristics of the teachers.
4. There are no significant differences among models in the contexts in which implementation takes place, but there do appear to be differences among sites within mode .

The finding of model and site differences on a large number of variables supports the argument that implementation is complex, and that some of these variables may be useful in explaining variation in level of implementation. At the same time, the discussion of factors which affect implementation reveals that more study is needed because there are many important dimensions of implementation for which the present data are inadequate, or for which there are no data at all.

The discussion of factors which influence implementation also suggests two areas in which traditional experimental requirements are not met. The findings of systematic differences among sites, and more importantly among models, in staff characteristics and in the training given by people other than the sponsor indicate that inputs other than the models may be confounding factors. These factors should be considered, then, when drawing conclusions about model effectiveness.

After identifying the factors which might affect implementation, we entered some of the variables in a series of regression equations in an attempt to explain the sponsors' ratings of levels of implementation. We find that:

1. As we expected from the earlier analyses of variance, models do not explain a significant amount of the variation in levels of implementation.
2. Site-to-site variation accounts for approximately 30% of the total variance.

3. We can explain 98% of the variance among sites in levels of implementation with three variables: sponsor's judgments of intra-staff friction, rapport between the staff and the administration, and adequacy of the physical plant indoors.
4. Within-site variables add only 10% to the explained variance after the site variation is controlled.

This failure to explain within-site variation underlines the contention that we do not have adequate knowledge about the process of implementation. The most impressive finding is that we can explain essentially all of the variance among sites in levels of implementation. Two interpretations of this finding are plausible. It is possible that sponsors, in rating the teachers, are simply equating pleasant sites with high levels of implementation: site differences may reflect only differences in pleasantness, and may be unrelated to actual differences in performance within the model. This interpretation is supported by the fact that judgments on the same three variables by consultants are not related to levels of implementation. It is also possible that pleasant sites actually do have higher levels of implementation than do unpleasant sites. This may be true either because a pleasant site is a prerequisite for implementation or because the staff in a site are so enthusiastic about a model that they work well together in order to implement it. Although it is difficult to make a definitive statement about the,

causal relation between the site factors and levels of implementation, this finding is both important and interesting. The relationship of site atmosphere and model implementation should be explored further.

The Classroom Observation data are used to discover how models differ in terms of classroom activity and whether those differences are interpretable in terms of model theory. There are three main findings:

1. Oregon and Kansas appear to be particularly well implemented, showing high scores on the variables which reflect their model prescriptions.
2. Bank Street and Pittsburgh seem to be well implemented in the areas of independent child activity, and adult attention to children.
3. Far West, Arizona, High Scope, Florida, EDC, REC, and Enabler classrooms do not appear to be significantly different from one another on most of the variables. However, each of these models is significantly high or low on at least one classroom observation variable.

Thus, the Classroom Observation chapter reports that some models can be readily differentiated on several measures, while other models seem to be rather similar to one another.

Implications for Future Studies

It appears that traditional experimental requirements -- that the treatments are well-specified and fully implemented -- are not met in Planned Variation. The contention that treatments were not well-specified was discussed in Chapter 1: we do not have a clear picture of what a class using each model would look like because not all Planned Variation models can be fully described in behavioral terms. The finding that the models are not fully implemented is summarized above. Such factors complicate an experiment because they make the interpretation of effects difficult.

It is important, then, to ask why the treatment requirements were not met in Planned Variation, in order to determine whether the deviations will re-occur in future studies. If we could conclude that they resulted from inadequate planning, then we could expect to have studies with well-defined and fully implemented treatments by simply taking more care to plan well. Our evidence, both quantitative and anecdotal, however, suggests that deviations do not simply result from poor planning but stem from the nature of the treatments and of the implementation process.

The nature of some of the Planned Variation models indicates that they will never be well-specified in a conventional sense because their philosophy runs counter

to close prescription. Instead they set out general principles and encourage teachers to carry them out in a manner which best fits their style and the needs of their students. The EDC and Bank Street models are well-known examples of this type of program. This is not to say that models of this kind cannot be carefully defined -- we strongly recommend that this be done where it has not been already. The level of description, however, may not be overt classroom behavior which remains constant in all situations.

The nature of the models together with the nature of the implementation process also indicates that treatments will not be fully implemented in the sense of finding all classes to be replications of the ideal models. In part this assertion comes from the discussion above: if a model can take a variety of forms, then identical replication should not be expected. In part, it is a function of the interactive nature of the implementation process. We expect some variation in all models -- even in those models which do not encourage variation -- because implementation depends not only on the model and the sponsor's input, but on site factors, such as the attitudes and skills of the staff and the organization of the Head Start center, which are beyond the control of the researchers. Thus it appears that the deviations in Planned Variation stem from the nature of the situation and therefore, will tend to occur in any study

involving complex treatments which depend on local people for implementation. We would argue that given the inevitability of the "messiness" of such studies, it should be taken into account in planning other studies. Moreover, it should be recognized that the messiness of a field study, aside from the experimental problems it presents, is not really a negative phenomenon. In a sense, it is preferable to laboratory research because it has greater ecological validity: treatments are tested under conditions which approximate those they will face in the real world. We should accept the vagaries of a study like HSPV, then, not only because they are inevitable, but because we can learn a great deal from them. Two broad recommendations can be made for future studies.

1. Alternative standards for a valid experiment must be developed. This involves both defining full implementation realistically and taking differences in implementation into account in analyzing treatment effectiveness. The important issue in redefining full implementation is allowing some variation among classes: we must accomodate the fact that a model may be adapted, by the sponsor and the teaching staff, to the unique situation in a site, or even to a class within a site, and to changes in a situation over time. Even in those models which do not emphasize adaptation, identical replication should not be expected. We need, however, to set limits on the variation: we must determine

how much variation can exist while still considering a class to be a fully implemented program. We must be able to decide at which point the variation is so great that full implementation becomes partial implementation, and at which point partial implementation becomes nothing -- no evidence of the no specific operational definition of full implementation. This is not to say that sponsors cannot identify a well-implemented class, only that they have not systematized, or at least communicated, their criteria.

Provision for taking differences in implementation into account in the outcome analyses is important because we would expect, even with a redefinition of full implementation, that some classes will be less than fully implemented -- a teacher may only use portions of a model, or only use it part of the time. Taking differences into account can be accomplished either by setting a criterion level or by using the extent of implementation as a covariate. In the case of the former, we would decide what level of implementation represented an acceptable level of treatment, and only analyze effects in classes which exceed that level. We would essentially be making a simple distinction between treatment and no treatment. The critical issue in making this simple distinction is deciding where the cut-off point should fall. A disadvantage to this approach is that it potentially sacrifices a great deal of data. The second approach, using implementation as a covariate or as an

independent variable in a regression equation to control for the variation in effects resulting from differences in implementation, does make use of all available data. Several other issues, however, must be considered. First, it is important that the scales used to measure the extent of implementation be anchored such that standards for judgments, particularly the criteria for full implementation and for no implementation, are made explicit. This was not done with the sponsor ratings of teachers described earlier, and, as a result, the conclusions drawn from them can only be tentative. Second, the problem of comparing partial implementation should be resolved. Within a single model, one class which is, say, 60% implemented may not look like another which is 60% implemented because different parts of the model may be present in each. The problem is to determine when these classes can be considered to be acceptable replications of a single treatment. Criteria should be established to make this determination, but to date, it has not been done. Comparison across models is more difficult and probably should not be attempted because models have different components and are at different levels of operationalization.

Setting all these standards is a necessary basis for determining whether the treatment as implemented is an acceptable example of the treatment we set out to test. Although the standards may be different than in a conventional experiment, they must still be met. If the standards cannot be established -- for example, if limits cannot be

set on the variation allowed, at least to the extent of being able to specify what differentiates a model class from non-model classes, and to specify what the model classes have in common -- then, we do not have a valid experiment.

2. The process of implementation is, in itself, an important topic for study. Within a traditional experimental framework, the focus of the research is on the effect of the treatments. In this context, the study of implementation is seen primarily as a simple check on the presence of the treatments. When we cannot assume that all classes will reach full implementation, it is equally important to examine the question of whether and in what form a program will be implemented, and to identify the factors which determine the process.

In this report, a great deal of speculation and some analyses with data originally collected for other purposes have been done on individual factors which might affect implementation. What is needed, however, is a study of the implementation process: a study designed to ask what effects sponsors' involvement, training efforts, and staff organization had; to ask what characteristics of the local staff, of the Head Start organization, of the community in which the center is located affect implementation; and to develop a theory of how these factors are interrelated -- i.e., to develop a theory of process rather than of individual factors. This might be done either by studying

natural variation in delivery systems or by experimenting with pre-specified strategies. This type of study goes beyond the issue of experimental validity to the investigation of conceptions of planned change. The existing literature on innovations, on intervention theory, on organizations, and on decision making should be useful in further conceptualizing these issues.

BIBLIOGRAPHY

Bracht, G., & Glass, G. "The External Validity of Experiments." American Educational Research Journal, 1968, 5, 437-474.

Campbell, D., & Stanley, J. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally and Company, 1963.

Cochran, W., "Some Methods of Strengthening the Common X^2 Tests". Biometrics, 1954, 10.

Draper & Smith. Applied Regression Analysis. New York: John Wiley and Sons, 1966.

Lunney, G. "Using Analysis of Variance with a Dichotomous Dependent Variable: An Empirical Study." Journal of Educational Measurement, 1970, 7.

Monaghan, A. "Patterns of Implementation in Head Start Planned Variation", The Huron Institute, Cambridge, Massachusetts, forthcoming.

Office of Child Development, Head Start Planned Variation Study, November, 1971, Washington, U.S. Department of Health, Education and Welfare.

Searle, S.R. Linear Models. New York: John Wiley and Sons, 1971.

Smith, M. "Some Short Term Effects of Project Head Start: A Preliminary Report of the Second Year of Planned Variation, 1970-71", The Huron Institute, Cambridge, Massachusetts, 1973.

Stanford Research Institute. "Implementation of Head Start Planned Variation Testing and Data Collection Effort. Final Report". Menlo Park, California

Stanford Research Institute, "Implementation of Planned Variation in Head Start (1969-1970)," Menlo Park, California, 1971.