DOCUMENT RESUME

ED 104 519 PS 007 709

AUTHOR Featherstone, Helen J.
TITLE Child Characteristics by Model Interactions.
PUB DATE 30 Aug. 74
NOTE 40p.; Paper presented at the Annual Meeting of the American Psychological Association (82nd, New Orleans, La., Aug. 30, 1974); Small portions of text may not reproduce clearly

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS Age Differences; *Cognitive Development; Comparative Analysis; *Compensatory Education Programs; *Demonstration Programs; Hypothesis Testing; Interaction Process Analysis; Intervention; *Preschool Children; Preschool Learning; *Program Effectiveness; Program Evaluation; Research Design; Response Style (Tests); Sex Differences
IDENTIFIERS *Project Head Start Planned Variation

ABSTRACT

Data from the 1969-70 and 1970-71 Head Start Planned Variation (HSPV) Study were used to examine program-child interactions. An effort was made to determine whether different preschool programs have different cognitive effects on different types of children. Seven hypotheses for the analysis of the data were generated from the results of the HSPV 1969-70 study; these hypotheses, which were tested on the 1970-71 data, involved the variables of ethnicity, socioeconomic status, sex, age, initial achievement, prior preschool experience, the response style variables of passivity and competence, and the dimension of program "directiveness". General findings are presented for each of the hypotheses. The concluding discussion focuses on three areas: (1) the relative importance of interactions of model with different variables; (2) the usefulness of various model groupings in predicting and interpreting interactions; and (3) the overall importance of the interactions of child characteristics and model in explaining cognitive outcomes. Due to limitations in the HSPV experimental design, the conclusions are presented as guidelines to profitable areas of inquiry rather than study findings. (SDH)

SEP. 24 1974

Child Characteristics by Model Interactions

by Helen J. Featherstone

Paper presented as part of the Symposium:
"Head Start Planned Variation: Learning
from a Social Experiment" at the
American Psychological Association Convention
in New Orleans, August 30, 1974.

I. Introduction

This paper describes an effort to use the Head Start Planned Variation (HSPV)[1] data to look at program-child interactions. The analyses referred to are described at far greater length in the report on the same subject (Featherstone, 1973).[2] The present paper is a summary of the findings and conclusions of the larger report as well as an effort to put them in perspective.

This part of the Planned Variation analysis addresses the issue of whether different preschool programs have different cognitive effects on different types of children. Specifically, it focuses on three interrelated questions. First, what characteristics, or types of characteristics, of children interact most powerfully with characteristics of preschool programs? Second, what are the patterns of such interactions; how pertinent to an understanding of these patterns are the considerations which have traditionally been used in grouping different types of programs? Third, how important are these interactions in explaining the cognitive outcomes of different programs?

The Head Start Planned Variation data is not ideal for answering these questions for a number of reasons. First, as has been repeatedly documented in the accompanying papers and other Huron reports (e.g., Smith, 1973; Wesiberg, 1973), HSPV does not meet the criteria for a true experimental design. Without random assignment of children to treatments, observed interactions are sometimes difficult to evaluate and interpret. Second, the design is unbalanced in ways which preclude some interesting inquiries.

For example, some sponsors were assigned to communities where nearly all target children were white; others served a primarily black population; there were significant numbers of Indian children in only one model. Hence, even though several ethnicity-by-model analyses were performed, the design imposed real constraints on what we could learn about interactions of model with ethnicity or culture.

Third, the range of child characteristics whose impact we could study was very limited. Although we had considerable demographic and test data on children, we had no information at all on the children's behavior in the classroom or at home; we knew very little about the child as a learner.

Finally, although the cognitive outcome measures used in the interaction study (the Stanford-Binet and the Preschool Inventory) are the best available to the PV study, they are very limited both in content and in context. These tests tell us a certain amount about information the child has acquired, but little about the nature of his conceptual functioning. They cannot measure personality traits such as initiative, flexibility and perseverance, which many preschool educators consider an important part of cognitive competence and which they intend their programs to foster. Further problems are raised by the oft-noted difficulties of getting valid test scores with young children from low-income families (Zigler and Batterfield, 1968; Reissman, 1962; Labov, 1969).

These problems are important and real; there is not way in which they can be adequately dealt with or wished away. Therefore, even though we cannot design our analyses in such a way as to correct for all inadequacies in the data, we can be cautious in our aims for different analyses and our interpretation of them. Thus, we consider our conclusions as guidelines to profitable areas of inquiry rather than findings which can be confidently translated into classroom practice. Because the conditions necessary for a true experimental design are not met by the PV study, and because the patterns in the data are regarded primarily as a basis for formulating hypotheses, no tables are presented and few references will be made here to levels of statistical significance. The reader who wishes to know more about the size and significance of observed effects is referred to the longer report by Featherstone, 1973.

II. Design of the Study

Because the preschool literature sheds a rather uncertain light on several of the interactions under study, hypothesis generation was seen as a major task of the report. And although certain preschool studies were valuable resources, the approach to this problem has had to be in part empirical. For this reason I used the data generated in the first year (1969-70) of the Head Start Planned Variations Study, in conjunction with results reported in selected preschool studies, as a basis for proposing specific hypotheses concerning the interaction of preschool programs and child characteristics. These hypotheses were then tested on another body of data--that generated in the second year (1970-71) of the HSPV Study. Because quite different measures were used in the last year of the study, it was unfortunately impossible replicate the analysis on the third cohort.

A. Models

The models of preschooling are the eight included in all three years of the PV study:

1. Academic Preschool Program: Engelmann-Becker (E-B)
2. Behavior Analysis Model: Don Bushell, University of Kansas
3. Cognitive Model: David Weikart
4. Parent Educator Model: Ira Gordon, University of Florida
5. Tucson Early Education Model: University of Arizona
6. Responsive Model: Far West Lab
7. Open Education Mode; Education Development Center (EDC)
8. Bank Street Approach: Bank Street College of Education

Differences and similarities among these models have been explored elsewhere; I do not feel I could add anything to this discussion in a paper of this length. However, it seems important to emphasize again, as other researchers have done in other papers, that we know far less than we would like to about actual differences in classroom practice among PV models.

At the beginning of the study I faced the question of whether (and how) models should be grouped for interaction analyses. A number of grouping schemes have been developed and described by other writers (see, for example, Weikart, 1972; Bissell, 1970; Mayer, 1971; SRI, 1971). Most of these could be applied without major difficulties to the PV models. On reflection, however, I felt that groupings based on theoretical considerations and/or sponsors' model descriptions (reliable observation data were not available for the 1969-70 cohort) would be inappropriate to the exploratory purpose of the 1969-70 analysis. Looking at outcomes of each of the eight models separately reduced the likelihood of obtaining statistically significant results, but I hoped it might increase our understanding of the usefulness of particular types of grouping schemes in describing the educational process.

In discussing the interaction patterns observed in the 1969-70 analyses I found that one consideration (implicit in nearly all categorization schemes of other writers)) seemed to recur:

the degree to which the role of adults in the classroom would be described as "directive." Clearly, all teachers are in some sense directive: they make rules about social behavior, aggression, and the use of materials; they set up some kind of time-table, they arrange the room and moderate their own behavior so as to encourage some activities while discouraging others. Nonetheless, models (as well as individual teachers) vary widely in the degree to which adults are expected to specify and directly supervise the minute-to-minute activities of children. On the 1969-70 analysis this consideration appeared to illuminate interactions of particular child characteristics with models. For this reason, I decided to group certain models as "more-directive" and "less-directive" for a few of the 1970-71 analyses in order to test hypotheses arising out of the first year's analysis.

B. Characteristics of Children

The child characteristics examined are as follows:

1. Initial ability or achievement level[3]
2. Previous preschool experience[4]
3. Sex
4. Age
5. Socio-economic status (SES)[5]
6. Ethnicity[6]
7. Response style: particular aspects of the way in which children respond to the cognitive demands presented in the Stanford-Binet IQ test as measured by a procedure similar to the Hertzig-Birch scoring.[7]

These characteristics were chosen because they represent a range of variables, both demographic and psychological, and because there was reason, either empirical or theoretical, to suppose that they might interact with characteristics of preschool programs.

C. Outcome Measures

The outcome measures used are two cognitive tests, the Preschool Inventory and the Stanford-Binet.[8] There are three reasons for this choice. First, these two measures have higher reliability than any others used in the PV study in 1969-70.[9] Second, they are more often used in preschool studies than the other tests in the PV battery, and hence are better known and most easily interpreted. Third, although the goals of the Planned Variations models vary greatly, all models include as an aim facilitating some kind of cognitive growth.

III. Generating Hypotheses from the 1969-70 Data

Strategy of the 1969-70 analysis

The analysis of the 1969-70 data was desgined as a data-dredging operation: its purpose was the generation rather than the testing of hypotheses.[10] In general the strategy used was as follows: for each variable, two-way[11] analyses of co-variance[12] were used to evaluate the importance both of the variable and of the interaction between the variable and the models in explaining the variance in post-test scores on the Stanford-Binet and PSI and gains[13] on these two tests. Because the purpose of the 1969-70 analysis was to generate hypotheses the pattern of interactions was explored even where the overall interaction did not reach acceptable levels of statistical significance. In general, the results for the four different dependent variables were examined together, and where two or more analyses showed substantial and compatible interactions between the variable and a particular model, a hypothesis was proposed.[14] This criterion was based on the assumption that interactions which show up in only one analysis are less likely to reflect strong effects, and in consequence less likely to be replicated; it is, however, somewhat arbitrary.

Results of 1969-70 analysis

Two types of hypotheses grew out of the 1969-70 analysis. First, there were specific hypotheses describing the actual structure of first-order interactions between particular child characteristics and individual models. These hypotheses were based very directly on interactions observed in the 1969-70 data and on results reported in other preschool studies. In a paper of this length it would be impossible to discuss and adequately justify all the specific hypotheses (and again, the reader who wishes to see a complete list of hypotheses, along with tables and particulars on the analysis, is referred to Featherstone, 1973, Part I). I do want to say a few words about interactions of model with the response style variables, since these variables are somewhat experimental and by no means self-explanatory. As noted in Section II-B, these variables refer to particular aspects of the way in which children cope with the cognitive demands presented in the Stanford-Binet pre-test, as measured by a technique similar to[15] the Hertzig-Birch scoring. In the 1969-70 analysis four variables were constructed from this data: "extensions," "substitutions," "competence," and "passivity." (See footnote 7 for an explanation of the variables). Children were rated "high" or "low" on each of these variables, depending on the number of responses they made in each category. In an analysis of covariance I examined main effects and interactions with these response style variables. On the basis of this analysis I decided that while two variables, "substitutions" and "extensions," were of no interest whatever--neither main nor interaction effects of these variables ever approached significance--interactions with the two other variables, "passivity" and "competence," might merit further investigation. Although main effects of these variables were not strong (the main effect of "passivity" was insignificant on all analyses), interaction effects were in some cases quite sizable). On the basis of the data these hypotheses were proposed:

1. Within-model interactions will favor children making few competence responses in Engelmann-Becker programs. In Bank Street programs interactions will favor those making many such competence responses. More generally, in more directive programs interaction effects will favor children making few competence responses, while in less-directive models they will favor children making many such responses.
2. Within-model interactions will favor children making many passive responses in Bank Street, Bushell and Engelmann-Becker programs; in Weikart, Tucson and Far West programs interaction effects will favor those making few passive responses.

In addition to specific hypotheses describing interactions of each child variable with model, seven general hypotheses were proposed as answers to the broader questions which had motivated the study. These hypotheses were a kind of summary of the more specific findings of the 1969-70 data analysis. They are listed below:

1. Predictable, significant, interactions of model with ethnicity and SES will not be found.
2. For most models the effect of sex on test score will be small. For one or two models, however, the effect of this variable will be predictable and significant.
3. An interpretable interaction of age with model is unlikely.
4. Interactions of model with initial achievement, prior preschooling, and the two response style variables "passivity" and "competence," may follow patterns which are both predictable and interpretable.

5. The "directiveness" continuum may apply to interactions of initial ability, prior preschooling, and the response style variable "competence" with model; if an exception is made for Bank Street, this dimension may apply to interactions with another response style variable, "passivity."

6. The dimension of directiveness does not apply to interactions of model with sex. Interactions observed on the IQ measure may relate to quite another aspect of the learning environment: the degree to which adults use concrete objects in their teaching.

7. None of the eight models will produce optimum gains for all types of children across both cognitive measures.

IV. Testing Hypotheses on the 1970-71 Data

Strategy of the analysis

The hypotheses derived from the 1969-70 analysis were tested on the data generated in the second year of the HSPV study, 1970-71. Somewhat different methods were used on the second year of data; the differences are described below.

Questions involving interactions of individual models with particular child variables were investigated through analysis of covariance and regression.[16] Analyses of covariance were used to evaluate the magnitude and significance of model interactions with six categorial variables: sex, preschool experience, ethnicity, SES category, and the two response style variables "passivity" and "competence." Regression analyses were used to investigate interactions with several continuous variables: initial IQ, initial PSI score, age, and three SES components, mother's education, income, and family size.

Because the question of whether and how to group models had seemed important since the beginning of the interaction investigation, and because the 1969-70 analysis indicated that the possible importance of the dimension of "directness" referred to above, I performed on the 1970-71 data a set of analyses referred to in these pages as "grouped-model" regressions. These analyses were intended to contrast models where adults take a highly directive role with models in which children make many significant choices about how--and with whom--to spend their time. Bushell and Engelmann-Becker were placed in a "more-directive" group, while

Far West, EDC and Bank Street were put in a "less-directive" one. In order to make the two groups as different as possible on the score of teaching strategy, the other three models, Weikart, Tucson and Gordon, were omitted from the "grouped-model" analyses.

The groupings, and the choice of models to be included, are based on sponsor model descriptions. Bushell and Engelmann-Becker are placed in the more-directive category because in both models children's major learning is alleged to take place in adult-directed groups where decisions about what is taught are made by adults. Although adults in several other models, e.g., Tucson, teach -small-group lessons, the adult is generally expected to take account of the children's comments and demonstrated interests as he proceeds. Thus, the actual form of the lesson is supposed to be determined by children and teachers working together (as in Weikart's "open" programs where both adults and children are expected to initiate and to respond [Weikart, 72]).

Far West, EDC and Bank Street are grouped together as less-directive because they emphasize the importance of the choices made by the individual child in his learning. All of these sponsors see the process of choosing as essential to learning. None of them expects children to spend the major part of each day in an organized group, working with an adult.

Weikart and Tucson were omitted because they fall somewhere between these two groups. Although these models emphasize the importance of choice-making, both appear to expect children to spend a significant part of each day in small groups with an adult.

Although in practice either of these models might look quite a lot alike the "less-directive" models, the structure for learning described by sponsors seems to direct the outcome of children's choices a bit more. The Gordon model is omitted because it contains no specific directions for classroom practice.

The Classroom Observation data supports the notion that practice in Engelmann-Becker and Bushell classrooms is quite different from that in other models (see Lukas and Wohlleb, 1972). Engelmann-Becker and Bushell are significantly above all other models ($p<.001$) in total academic activity and in frequency of adults with children in academic activities. They are below other models in independent child activity, with Bushell being significantly low ($p<.001$). Bank Street, Far West, Tucson, and EDC rank highest on this variable. In general the evidence for placing Engelmann-Becker and Bushell in the more-directive group is stronger than that binding Bank Street, ECD and Far West together. However, these three programs do show evidence of considerable independent child acitivity, both academic and non-academic.

Results of the 1970-71 analysis

Many of the predictions made on the basis of the 1969-70 data were supported by the 1970-71 analyses; some require further qualification.

1. Predictable, significant interactions of model with ethnicity and SES will not be found.

This prediction was supported by analysis of the second year's data. In the 1970-71 PV data, interactions of model with ethnicity were significant for the PSI. However, these interactions were not in the direction predicted on the basis of the 1969-70 PV data. For the combined SES measure--which weights equally family size, mother's educational level, and family income--interactions with model were trivial for both the PSI and the Stanford-Binet. Interactions of model with individual SES components were in some cases significant, but they did not follow any consistent pattern.

2. For most models the effect of <u>sex</u> on test score will be small. For one or two models, however, the effect will be predictable and significant.

Although sex differences on the Stanford-Binet were in the directions predicted on the basis of the 1969-70 data, they did not reach statistical significance. For children without prior preschooling sex effects appeared to be small. There is, however, some indication that for children past the first year of preschool, sex might affect achievement more strongly:[17] the PSI data showed second year boys in more-directive models scoring substantially higher than boys in less-directive models; differences between the achievement of girls in more- and less-directive models were negligible.

3. An interpretable interaction of <u>age</u> with model is unlikely.

The 1970-71 data--particularly the IQ data--tended to support this prediction. However, there were some indications that on this second round of data the two behaviorist models, Bushell and Engelmann-Becker, fell together in favoring the PSI achievement of younger children (this was not true in the 1969-70 analysis). In the less-directive models age appeared to have inconsistent effects on test scores.

4. Interactions of model with initial achievement, prior preschooling, and the two response style variables "passivity" and "competence" may follow logical and predictable patterns.

This prediction was supported, in a limited sense, by the 1970-71 PV data. Interactions of the two response style variables with model and model-group were insignificant on the PSI. On the Stanford-Binet they were, however, as hypothesized: children who made few "passive" responses and/or many "competence" responses on the IQ pretest gained more in less-directive programs, while those who were high in "passive" and/or low on "competence" responses tended to do better in more directive models. A conspicuous exception to this formulation was Bank Street, which appeared to favor the achievement of children making many passive responses at the time of pretest.

Interactions of model and model-group with prior preschooling also followed similar patterns across the two years of PV data: on both tests children with prior preschooling appeared to gain somewhat more in the less-directive models, while first-year children did better in more-directive programs.

Interactions of initial achievement with model-group followed predicted patterns to some extent. However, this statement requires considerable qualification: the initial IQ variable, used in both 1969-70 and 1970-71 analyses, looked quite different on the two sets of data. But when initial achievement was measured by the PSI prescore on the 1970-71 analysis, interactions--at least for the PSI--were as predicted. Children of low initial score did better in more-directive models than in less-directive ones, while for high-scoring children the opposite was true. However, because of the different patterns found for the two independent variables initial IQ and PSI pre-score, we must be very cautious in assigning importance to these results.

Thus, we can say that interactions of models with these variables were, to some limited extent, predictable. But do the observed patterns make sense? I believe that they do, although I think it is important to offer interpretations very tentatively in the absense of observational data which related particular children's classroom behavior to their test responses. Looking first at the response style variables "passivity" and "competence," I would guess that in classrooms where teachers as well as children had some choice about how and with whom to spend time--this describes less-directive models-- a child's style of dealing with a too-difficult problem might well influence the amount and type of adult attention he received. Specifically, it seems likely that a preschooler who perceived and vocalized some limitation on his competence to perform the task (e.g., "I don't know how to do that;" or "I can make a tall tower, but I'm too little to count the blocks") might get a great deal of attention directed specifically toward his cognitive needs, while a classmate who responded "passively" to the same situation (i.e., did nothing) might get relatively little.[18] These facts of classroom life (if facts they be) could well be reflected in test results. The data for Bank Street--both in 1969 and in 1970-71--suggest that a strong model emphasis on affective development may be particularly beneficial (in the limited sense of promoting measurable gains on cognitive tests) to "passive" or withdrawn children, or may serve to focus teacher attention on children who behave in this way. Likewise, it is possible that behaviorist models tend to encourage passive responses over "competence" responses and that children to whom this style comes naturally fit into behaviorist classrooms somewhat more readily.

In regard to interactions with prior preschooling and initial achievement test I would venture a related suggestion: maybe that in order to learn the sorts of things that cognitive tests measure in the context of a "less-directive" classroom, one needs to have mastered certain skills for making choices, perceiving problems, and getting and maintaining adult attention. If so, children who have not yet acquired these skills--first year preschoolers, and/or those with lower initial levels of achievement--may make more measurable gains in more-directive classrooms, while those whose learning skills are more advanced may actually move more swiftly in a less-directive environment.

This is an elaborate tissue of speculation. Research data on the behavior of children both in classrooms and in testing situations and on adult responses to different styles is needed to tell us whether these steps toward a tentative explanation of our served patterns are even in the correct direction.

Interactions of model and model-group with prior preschooling also followed similar patterns across the two years of PV data: on both tests children with prior preschooling appeared to gain somewhat more in the less-directive models, while first-year children did better in more-directive programs.

5. The "directiveness" continuum may apply to interactions of initial ability, prior preschooling and "competence" with model; if an exception is made for Bank Street, this dimension may apply to interactions with "passivity."

The 1970-71 PV analysis tended to support this prediction. The interaction of PSI pre-score with model-group was significant ($p<.001$), and in the predicted direction, for PSI grouped-model regressions. Interactions of prior preschooling and "competence" with model-group followed predicted patterns and reached acceptable levels of statistical significance ($p<.005$) on IQ analyses, although not on the PSI. Interactions of "passivity" with model were as predicted on the IQ. Nevertheless, confirmation of predicted patterns was not as strong as it might have been because in general the interactions reached significance on only one of the two tests.

6. The dimension of directiveness does not apply to interactions of model with sex. Interactions observed on the IQ measure may relate to quite another aspect of the learning environment: the degree to which adults use concrete objects in their teaching.

The 1970-71 data did not suggest striking confirmation of this prediction. Although sex effects within models were in the predicted direction for the Stanford-Binet, they did not even approach statistical significance. Furthermore, there was some indication that hte dimension of directiveness might apply to interactions of model and sex in the PSI; the PSI grouped-model regressions showed a substantial and significant second-order interaction of sex, model-group, and prior preschool experience. According to this analysis, the contributions of sex and model-group to achievement were essentially additive for first-year children: girls did a little better than boys; more-directive models seemed to boost scores a bit more than less-directive ones. For children with prior preschooling, however, the situation seemed to be quite different: although girls scored about the same regardless of which type of program they were assigned to, boys did substantially (nearly three-quarters of a standard deviation on the PSI post-test) better in the more-directive models.

7. None of the eight PV models will produce optimum gains for all types of children across both cognitive measures.

This hypothesis was supported by the 1970-71 data. Although the main effect of the Weikart model on the Stanford-Binet was so strong as to dwarf interaction effects, this was not true for the PSI (see Smith, 73, for a detailed analysis of model effects on the PV cognitive battery). In terms of the IQ test, then, all types of children made largest gains in the Weikart model. However, on the PSI the situation was different: for this test interaction effects tended to be as large as model effects, so that the model which produced optimum gains for one type of child was not the one which worked best for another.

V. Discussion

These general hypotheses provide a good starting point for a discussion of broad questions about the relative significance of different child variables, the usefulness of conventional model groupings in predicting interactions, and the overall importance of interactions in explaining cognitive outcomes.

The first question raised in the introduction to this paper concerns the relative importance of interactions of model with different variables or types of variables. This is not always easy to evaluate, since the size and significance of an interaction does not necessarily tell us how important it is. Thus, ethnicity-by-model interactions are significant, or nearly so, for all analyses. But because the direction of the observed effects is consistent neither across years nor across tests this interaction does not help us to predict which children will benefit most from particular models. In order to decide which variables are most important, it seems reasonable to consider the size and significance of effects, whether they are consistent across two years of data, and the degree to which the observed pattern is interpretable. Using these criteria, the two response style variables, "competence" and "passivity" seem to be among the most important. Although interactions of these two variables with model and model group are insignificant on 1970-71 PSI analyses, the interactions are significant and in the predicted directions for the Stanford-Binet analyses. On the IQ measure, the pattern of observed effects is the same for both years of data, and makes sense in the light of what we know about preschool curricula.

To a more limited extent, we can say the same thing about prior preschool experience. Although not all the predicted interactions reach statistical significance, the pattern of effects is very similar across the two years of data: children with prior preschooling gain most in less-directive models, while first year children do better in more-directive programs.

For three other child characteristics, initial achievement, sex, and age, the 1970-71 analysis indicates the existence of interpretable interactions of some magnitude. However, since these interactions are unreplicated, I cannot speak about them with equal confidence.

Interactions of initial achievement, sex, and age with model appear to affect children's performance on the PSI post-test but not on the Stanford-Binet. The 1970-71 data suggest that behaviorist models favor the PSI achievement of young children, of those initially scoring low in the test, and of boys with prior preschool experience. Within-model differences are smaller and somewhat less consistent for less-directive models. The analyses relating to age and PSI prescore lead this writer to infer that behaviorist approaches are especially efficient in facilitating the learning of children who have not yet reached a certain basal level of achievement or cognitive functioning; they suggest that other, more open-ended approaches may work better for children who because of age, prior preschooling, or precocity start the year somewhat better prepared.

These results in some ways parallel those reported by Bissell (1970), in her analysis of the contribution of SES to final test score in different types of preschool programs, and to Bar-Yam's summary of ATI studies. Bissell's analysis shows that although children of higher SES outscore low SES children in "less-structured" or directive preschools, this is less often true in more-directive programs like Bereiter-Engelmann. Similarly, Bar-Yam reports that in a number of studies of older children's learning, students of low ability appear to gain more in "directive" programs than in "permissive" ones; for high-ability children the choice of curriculum appears to influence performance less. These studies differ from Planned Variation in a number of ways. Nonetheless, the similarity of the patterns is suggestive.

Interactions of model and model-group with two remaining variables, SES and ethnicity, appear to be of less interest. Although interactions of model with ethnicity are significant in both 1969-70 and 1970-71 analyses, the patterns are inconsistent, even contradictory, across the two tests and the two years of data. Interactions of SES and model are not significant in either years' analyses when SES is defined by a combined measure.

The PV data indicate that the finding of Bissell and others that more-directive models favor the achievement of low SES children may not hold when the models are implemented in the context of Head Start, and when a number of other variables are controlled. Interactions of model-group (Engelmann-Becker and Bushell vs. ECD, Bank Street, and Far West) and SES components do not even approach significance either on IQ or PSI analyses. Furthermore, the main effect of model-group is insignificant (and favorable to less-directive models) on IQ analyses, and rather fragile on PSI analyses. (Although the effect favoring more-directive models is significant on some analyses, it is insignificant on others; the magnitude of the effect depends on the choice of other independent variables.) Knowing that the whole PV sample is of low SES (compared to national norms), we had expected to find strong main effects favoring more both cognitive measures, but this is not the case.[19]

There are, I think, two general points we can make about the relative importance of the child variables considered in this paper. First, the variables whose interaction with model follows the most consistent pattern across the two years of data are those which relate most directly to the child's behavior and experience as a learner: the response style variables, and prior preschool experience. Second, none of the variables which interact interestingly with model or model-group refer to immutable characteristics of children. All of them describe the child at a particular point in his educational experience. Age, PSI achievement, prior preschooling, and response style: all these things change from year to year. Sex by itself shows no consistent interactions with model group; only when it is considered in combination with prior preschooling is the interaction strong. The impact of particular models on little boys may depend on whether or not they have been in school before. IQ, which changes less over the years than achievement level, interacts far less powerfully with model and model-group.

These results suggest that the strategy which works well for a child today will not necessarily be optimum next month or next year. This is nothing new: plenty of good teachers use this knowledge every day in their classroom, allowing first-graders more freedom, for example, as their reading skills improve and they are more able to work independently. Nonetheless, the point needs emphasizing: the PV analyses described here lend wieght to the idea that we can increase preschoolers' achievement by adapting curriculum in particular ways for particular children. None of the data supports the notion that the choice of curriculum for a particular child or group of children should be final. Those characteristics of children which do not change--ethnicity, SES and sex--are precisely the ones which do not show consistent or interpretable interactions with model.

The second question which this investigation sought to answer concerned the usefulness of various model groupings in predicting and interpreting interactions. The 1969-70 data indicated that the dimension of "directiveness" might be pertinent to interactions of model with initial IQ, prior preschooling, and the two response style variables, but not to interactions of model with sex, ethnicity and age. The 1969-70 analyses, taken together with some of the 1970-71 observation data, raised the possibility that interactions of sex and model might relate to another dimension: the degree to which concrete objects are used for teaching and learning.

The 1970-71 analysis supports the idea that the dimension of directiveness applies to interactions of initial ability, prior preschooling and certain response style variables with model. But the second year's data suggestthat the dimension may also relate, in limited ways, to interactions of age and sex with model. These interactions are observed only on PSI analyses, and not on the IQ. Sex effects on the IQ give very limited support to the theory that an emphasis on learning through concrete objects will favor the IQ gains of boys more than those of girls. However, the observed effects are not significant.

The dimension of directiveness thus looks more important to interactions observed in the 1970-71 analyses than to those found in 1969-70. But while saying this I want to emphasize again that in these analyses, the main effect of "directiveness" is small. Although children in directive models score a little bit higher on the PSI they do slightly (insignificantly) less well on the Stanford-Binet.[20]

The third question we have asked about interactions of child characteristics and model concerns their overall importance. Can we say that one model or type of model is "best"--in the limited sense of maximizing cognitive gains--for all children? Or are interaction effects in fact more substantial than model effects? The 1969-70 analyses indicated that the answers to these questions were different for different tests; although on the Stanford-Binet interaction effects were more substantial than model effects, this was less often true for the PSI. For all models except Weikart, a similar pattern is observed in the 1970-71 analysis. The effect of the Weikart model[21] on IQ post-test scores is, however, so substantial as to dwarf the importance of interactions. (For more on this, see Smith, 1973).

On those of the 1970-71 analyses in which five models are grouped as more- or less-directive, the interaction of model-group with child characteristics explains substantially more of the variance in post-test scores--both PSI and IQ-- than does the main effect of model-group. The main effect of model-group explains 1.9% of the unique variance in PSI post-test scores and .5% of the variance in IQ scores.[22] Interactions of background variables with model-group explain more than 12% of unique variance in both measures. Grouped model analyses indicate that neither approach (more-directive vs. less-directive) is optimum for all children.

This is less clearly true on the analyses where effects of the eight models are considered separately. Effects of models explain somewhat more variance, and interactions of model with child variables explain less. Nonetheless, for the PSI we can still say that no one model produces optimum results for all children (for example, analyses relating to prior preschooling indicate that Tucson maximizes PSI gains for children with preschool experience while Weikart favors first year children).

The model analyses of IQ post-test scores reveal quite a different situation. Although interactions of model with several child variables are significant, the main effect of the Weikart model on IQ post-test scores is so substantial as to reduce the importance of interactions. For 1970-71, we could say that, in terms of IQ, all types of children (at least as defined by the variables discussed in this paper) gain more in the Weikart model.[23] These findings have, I think, certain implications for educational policy and research. First, and most important, they support the idea that educational diversity can benefit children. Although the PV data indicate that one model, Weikart, may be astonishingly successful in promoting IQ gains with all kinds of children, they do not suggest that one type of model (more-directive vs. less-directive) maximizes cognitive gains for all kinds of children. The inconsistent patterns found in 1970-71 analyses on the PSI and IQ post-tests suggest that the choice of educational program should depend on the outcome sought as well as on the children, since main effects as well as interactions are somewhat different on the two tests. The Weikart model, for example, although outstanding in its effect on IQ post-test scores, is somewhat less effective than Engelmann-Becker in raising PSI scores.

If the PV analysis indicates that a choice of curriculum which takes into account the differences among children may raise test scores significantly, it also suggests that diversity should be created on the micro rather than the macro level--within schools, preschool centers, or classrooms, rather than just within cities or school systems. I say this for two reasons. First, all the evidence from the PV analyses points to the fact that global demographic variables like ethnicity and SES do not interact in a predictable way with model, at least in a Head Start setting. Second, these data indicate that children's educational needs change--that while one model may efficiently raise the scores of four-year-olds without prior preschooling, another approach may benefit these children a year later. If, as I have argued, the variables which matter most are those which relate to classroom behavior and learning style, then curricula should be flexible. Otherwise half the benefits of diversity will be lost.

What, if any, is the educational significance of the effects described in these pages? Given what we have begun to suspect both about the limited impact of school differences generally (Jencks, et al, 72), and about the mortality of preschool IQ gains (Stearns, 71), it seems quite possible that a difference of half a standard deviation on the PSI or Stanford-Binet will not in future years translate itself into higher earnings, greater social mobility, or even improved understanding of fifth grade artihmetic. Nonetheless, differences of this magnitude do suggest that in the short-run, over the course of a school year, children are learning some kinds of things considerably faster than they were before the Head Start experience. The analyses described here suggest that for particular types of children some education environments facilitate this learning more than others. And while these differences may make little impact in the long-run of people's live, they may reflect some important differences in the match between children's present needs and their preschool experience.

Given the very real limitations of the cognitive tests used here, and the great importance of other goals of preschooling, both cognitive and non-cognitive, we cannot be very sure that optimizing gains on the PSI or the IQ is of primary importance. But I do think it worthwhile to investigate <u>why</u> some children gain more than others in particular environments. The observational studies which would illuminate this point might well shed light on a prior question: what kinds of growth do these test gains reflect, and how important are they anyway?

This paper demonstrates, I think, the need for further research on the interaction of child and model variables. The specific findings discussed in these pages are nowhere near clear-cut enough to be translated into classroom practice. Replication of any patterns reported here would be interesting and important. Nevertheless, one point seems to stand out: research directed at the question of what kinds of programs will benefit particular children in particular ways should look at characteristics of children which relate as directly as possible to their behavior in cognitive situations. This is no new idea: a number of good studies have done exactly this, often with interesting results (for a summary and review of several such studies, see Lesser, 1971). However, too little work of this sort has been done on the preschool level: we need more sensitive indices of response style and ideas about what other variables relate to children's classroom needs; we also need observation studies which illuminate the reasons for observed interactions.

In designing studies which might help us understand which children are likely to make what kinds of gains in particular environments, we should bear two points in mind, both lessons of the Planned Variations Study. First, almost all classrooms provide a mixture of more and less-directive situations. While this may make interpretation of data more difficult, it provides a real opportunity for those interested in children's learning to observe one child in a range of learning situations and learn what "response style" means in practice. Perhaps we need instruments which help us observe the differences--and similarities--in ways in which children response to more and less formal situations within the same classroom and ways that children affect learning environments.

The second lesson which emerges from the analysis of Planned Variation data which is described here is that in order to learn about interactions we need small experimental studies designed to test specific hypotheses. In data dredging operations of the sort described in this report one too often lacks, in the end, the very information one needs in order to understand the most provocative findings.

Footnotes

[1]Planned Variations is a three-year study in which twelve different sponsors have been given funds and facilities to implement their model of preschooling in Head Start centers in several sites across the country. Data have been collected on children in these sponsored classrooms and in comparison classrooms (Head Start classes in which no sponsor is attempting to implement his model of preschooling). These data include both demographic information and pre- and post-scores on a variety of instruments.

For a description of the design and selected analysis of year 1 of Head Start Planned Variations see Implementation of Planned Variation in Head Start. Menlo Park, Calif.: Stanford Research Institute, 1971.

For a description of measures used in all the years of HSPV, see Walker, Bane and Bryk, The Quality of the Head Start Planned Variation Data. Cambridge, Mass.: Huron Institute, 1973.

[2]This report, Cognitive Effects of Preschool Programs on Different Types of Children, is available from the Huron Institute, 119 Mt. Auburn St., Cambridge, Mass. 02138.

[3]Measured by prescore on the Stanford-Binet (1969-70 analysis and 1970-71 analysis) and the PSI (1970-71 analysis only).

[4]Relates only to the question of whether the child has been to school before and not to the character of his earlier preschool experience.

[5]The HSPV study has data on family income, family size, and mother's education level. Standard scores for the three variables were weighted equally and added together, giving a measure of SES for each child.

[6]Only Black and White children were included in the analysis.

[7]These variables require considerable explanation, and interested readers are urged to consult the longer discussion in Featherstone, 1973 (see, in particular, Part I Section VII and Part II Section VII). When the Stanford-Binet was administered in 1969-70 and 1970-71 to the HSPV sample, children's answers were scores by a procedure similar to that originally worked out and described by Hertzig, Birch et al (1968) in a study of class and ethnic differences in response style. Each response was placed in one of eight categories: (1) Delimitation: correct, complete, and limited to requirements of problem; (2) Extension: correct, going beyond requirements of problem; (3) Incomplete: incorrect, but relevant to problem; (4) Negation: direct refusal to work; (5) Substitution: irrelevant verbalization or activity; (6) Competence: statement of limitation on ability to perform task or answer question; (7) Aid: request for help; (8) Passive: no response. Most responses were coded (1) or (3).

It was hoped that the data generated by this coding of responses might give interesting information about a child's style of responding to cognitive problems. For this reason the frequency with which children made particular responses on the pretest was used in interaction analyses as a basis for constructing four response style variables: "extension," "substitution," "competence" and "passivity." Each child was scored as "high" or "low" on each of the variables, depending whether he made more or fewer responses in each category than the sample median.

[8]The Preschool Inventory, developed in 1968 by Bettye Caldwell, is basically an achievement test, designed to measure knowledge in areas that are relatively independent of a child's particular background and experience. It tests the child's level of general information with items like "Where do you go when you are sick?" and "Color the traingle yellow." This test is very sensitive to maturation, with the greatest gains being made by the younger preschoolers in the PV sample. Because scores for this test have not been normalized for age (as are the Binet scores), raw gains are harder to interpret. Thus, the child who gains six points on the PSI in eight months may actually have lost ground in relation to his age group (this would depend on his age); this is not true for the Binet.

The Stanford-Binet, although it presupposes certain knowledge on the part of the child, is intended as a test of learning ability rather than achievement. Many although not all of the items require the child to solve problems, both verbal and non-verbal: he is asked to duplicate a block bridge built by the experimenter, to identify missing or incongruent objects in a picture, to complete sentences by supplying opposites (e.g., "The day is light, the night is ____.")

The Preschool Inventory was administered to all children in the study, while the Stanford-Binet was given only to a randomly selected 50%. Certain models are therefore excluded from particular IQ analyses, due to small cell size.

[9]The test-retest reliability for the 1937 scale of the Stanford-Binet for children 2 1/2 to 5 1/2 ranged from 183 to .91, depending on the IQ of the child (the reliability is highest for children in the lowest IQ ranges). The reliability of the 1960 scale (used by PV) is probably higher, since only the most reliable items are included in the revised form (exact figures are not, however, available for the 1960 Revision).

The internal (KR-20) reliability of the PSI is about .90. For technical information on these two instruments, see Walker, Bane, and Bryk, The Quality of the Headstart Planned Variation Data, Huron Institute, 1973.

[10]Although other preschool evaluation studies were also used as resources in generating hypotheses, the bulk of the hypotheses actually proposed in Part I of the report (Featherstone, 1973) were based on the PV data, so it is the analysis of this data which I will describe here.

[11]The possibility of second-order interactions suggests the advantages of using three-way analyses, rather than two-way. However, small cell sizes and an unbalanced design made most three-way analyses impractical.

[12]The Data-Text packaged program for unweighted-means analysis of covariance was used. An unweighted means analysis was selected because the sample size for a particular program is unrelated to any real properties of the model. Since the number of children in each model is a matter of chance, there seemed no reason to give greater weight to models which happen to have more children.

Covariates on these analyses include: age, sex, race, preschool experience, income, and family size. Pre-test score was included for post-test analyses but not gains analyses.

Additional information on the sample and analysis is given in Featherstone, 1973.

[13]The use of raw gain scores is currently in disfavor because of the problems deriving from unreliability of instruments--in particular regression-to-the-mean effect. The use of the post-test scores, adjusted by pre-test, is generally considered to be more satisfactory because it bypasses the regression problem. A strategy which includes the use of gain scores adjusted for covariates is used for reasons related to the design of the PV study. Because Planned Variations is not a true experimental design, with random assignment of sites to sponsors, certain variables are confounded with model (see Smith, 73). In consequence, in the 1969-70 analysis it seemed necessary to acopt a conservative strategy: we used two types of analysis for each test and limited hypotheses to occasions where two strategies--or else effects on two instruments--showed a measure of agreement.

Because results of gain score analyses appear similar to results of post-test analyses, the gain scores are not used at all in the 1970-71 analysis. Techniques employed on the 1970-71 data are described below.

[14]The decision to look at the data in this way reflects an early assumption that the PSI and the Stanford-Binet measured similar skills and learnings and that, in consequence, main effects and interactions could be expected to look very much the same. This assumption was not born out by the analyses and probably directed our attention from the interesting question of how these tests in fact differed from one another.

[15]The PV procedure for coding responses differed from the Hertzig-Birch procedure in two ways: first, in the PV study only the child's last response to a particular item was coded--not, as in the Hertzig-Birch study, the entire stream of his behavior. Second, the Hertzig-Birch study included a category which was the equivalent of spontaneous extensions (2) for wrong as well as right answers.

[16]The 1970-71 analysis includes only covaried post-test scores. Raw gain scores are not analyzed separately because these interactions appear on the 1969-70 analysis to be quite similar to post-test interactions. Since the PSI and IQ tests are significantly correlated a multivariate analysis of covariance might have been done. However, the two tests are analyzed separately for two reasons: first, the PSI sample facilitates comparisons which would otherwise be impossible. Second, although the tests are correlated (.541 at pretest) model effects on the two measures are quite different (see Smith, M.E., <u>Some Short-term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation, 1970-71</u>, Huron Institute, 1973).

[17]This fits in with the commonplace observation that elementary age boys are far more likely than girls to be academic under-achievers, with, for example, the ratio of boys to girls being over five to one in remedial reading programs.

[18]There is at least a bit of research data to support this explanation. Monaghan, in a study of the behavior of teachers and children in one "open" preschool (1971), found that children's own initiations toward teachers early in the year set a pattern of teacher-child interaction which was maintained by teachers as well as children throughout the school year. Children who talked to, questioned and actively sought out teachers in the fall were in turn talked to, questioned and sought out by teachers throughout the year.

[19]On the NYU booklet 4a, which tests knowledge of letters and numbers, the more-directive models do show stronger gains. This is not surprising, as the other models do not place major emphasis on this type of learning. (See Smith, 73).

[20]These observed differences between the two tests are consistent with early results reported by Robert Soar in his analysis of process and outcome in selected Follow-Through models (71). In the first two years' data, Soar found

> a tendency for abstract measures of pupil growth to relate positively to classroom behavior dimensions that reflect pupil freedom and self-direction, whereas simpler, more concrete measures of pupil growth tend not to relate, or even in some cases to relate negatively. In contrast, but relatively consistently, the simpler measures of pupil growth tend to be related to classroom behavior dimensions representing more structure and more control on the teacher's part.

These relations do not hold in Soar's analysis of later Follow-Through data (Soar, 1972). Nonetheless the early patterns are of interest because they parallel those reported for the 1970-71 analyses: on the PSI, a test heavily loaded with informational items, children in more-directive models score a little higher than those in the less-directive group. On the Stanford-Binet, by contrast, the very modest (and insignificant) differences between model-groups favor the less-directive models.

[21] Had the Fort Walton Beach Weikart site been included in 1969-70 analyses the pattern across the two years would have been more similar.

[22] See Featherstone, 1973 for a list of variables on regression tables.

[23] Between-model differences are small for the other seven models, so that if Weikart were excluded, no one of the remaining models would produce optimum IQ gains for all types of children.

References

Bar-Yam, M. The Interaction of Instructional Strategies with Students' Characteristics. Harvard University, Center for Research and Development on Educational Differences, 1969.

Bissell, J.S. The Cognitive Effects of Pre-school Programs for Disadvantaged Children. Unpublished doctoral dissertation, Harvard University, 1970.

Featherstone, H. Cognitive Effects of Preschool Programs on Different Types of Children. Cambridge, Mass.: Huron Institute, 1973.

Hertzig, M. E., Birch, G., Momas, A. & Mendez, O.A. Class and Ethnic Differences in the Responsiveness of Preschool Children to Cognitive Demands. Monographs of the Society for Research in Child Development, 33(1), 1968.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Coehn, D., Gintis, H., Heyns, B., Michelson, S. Inequality. New York & London: Basic Books, 1972.

Labov, W. The Logic of Non-Standard English. In Georgetown Monograph Series on Language and Linguistics, Monograph No. 22, 1969.

Lesser, G. Matching Instruction to Student Characteristics. In G. Lesser (Ed.), Psychology and Educational Practice. Glenview, Illinois: Scott Foresman & Co., 1971.

Lukas, C.V.D., & Wohlleb, C. Implementation of Head Start Planned Variation, 1970-71. Cambridge, Mass.: Huron Institute, December, 1972.

Mayer, Rochelle. A Comparative Analysis of Preschool Curriculum Models in Anderson, R. & Shane, H., eds. As the Twig is Bent, Boston: Houghton Mifflin, 1971.

Monaghan, A. Teachers in the Central School. Harvard Graduate School of Education, Unpublished paper, 1971.

Reissman, F. The Culturally Deprived Child. New York: Harper & Row, 1962.

Smith, M.S. Some Short-term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation, 1970-71. Cambridge, Mass.: Huron Institute, January 1973.

Soar, R. S. Follow-Through Classroom Process Measurement and Pupil Growth (1970-71). Institute for Development of Human Resources, University of Florida, Gainesville, Florida, June, 1972.

Soar, R. S. A Measurement of Teacher Classroom Management, Paper presented at a Sumposium at the APA Meeting in Washington, D.C., September 1971.

Stanford Research Institute. Implementation of Planned Variation in Head Start: Preliminary Evaluations of Planned Variation in Head Start According to Follow-Through Approaches (1969-70). Interior Report on Contract HEW-OS-70-134. Menlo Park, California, March 1971.

Stearns, M.S. Report on Preschool Programs: The Effects of Preschool Programs on Disadvantaged Children and Their Families, Office of Child Development Report, draft, 1971.

Walker, D. R., Bane, M. J., & Bryk, T. The Quality of the Head Start Planned Variation Data. Cambridge, Mass.: Huron Institute, 1973.

Weikart, D. Relationship of Curriculum, Teaching, and Learning in Preschool Education. In J. C. Stanley (Ed.), Preschool Programs for the Disadvantaged: Five Experimental Approaches to Early Childhood Education. Baltimore and London: Johns Hopkins University Press, 1972.

Zigler, E., & Butterfield, E. Motivational Aspects of Changes in IQ Test Performance of Culturally-Deprived Nursery School Children, Child Development, 39(1), 1968.

Zimiles, H. Has Evaluation Failed Compensatory Education. In J. Hellmuth (Ed.), Disadvantaged Child, Vol. III, Brunner Mazel, 1970.

Weisberg, H.I. Short Term Cognitive Effects of Head Start Programs: Report on the third year of Planned Variation: 1971-72, Cambridge, Mass.: Huron Institute, 1974.